

Encyclopedia of  
**APPLIED  
PSYCHOLOGY**



Editor-in-Chief

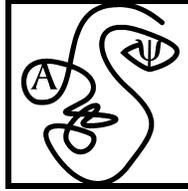
**CHARLES SPIELBERGER**



Editor-in-Chief  
**CHARLES D. SPIELBERGER**

University of South Florida  
Tampa, Florida, USA

Charles D. Spielberger is Distinguished Research Professor of Psychology and Director of the Center for Research in Behavioral Medicine and Health Psychology at the University of South Florida. Author, co-author, or editor of more than 350 professional publications, Spielberger's current research focuses on anxiety, anger, curiosity, depression, job stress, and lifestyle factors that contribute to the etiology and progression of cancer and cardiovascular disorders. His *State-Trait Anxiety Inventory* (1970, 1983), with translations in 66 languages and dialects, has become a standard international measure of anxiety. He has served as President of the International Association of Applied Psychology, the International Council of Psychologists, and the International Stress Management Association and was the 100th President of the American Psychological Association.



## Editorial Advisors

**John W. Adair**

*University of Manitoba  
Winnipeg, Manitoba, Canada*

**Michael Harris Bond**

*Chinese University of Hong Kong  
Hong Kong, China*

**Pieter J. D. Drenth**

*Vrije Universiteit Amsterdam  
Amsterdam, The Netherlands*

**Edwin A. Fleishman**

*George Mason University  
Potomac, Maryland, USA*

**Raymond D. Fowler**

*American Psychological Association  
Washington, D.C., USA*

**Michael Frese**

*Justus-Liebig University of Giessen  
Giessen, Germany*

**Machiko Fukuhara**

*Tokiwa University  
Mito, Japan*

**James Georgas**

*University of Athens  
Athens, Greece*

**Frank Heller**

*Tavistock Institute of Human Relations  
London, United Kingdom*

**Stevan Hobfoll**

*Kent State University  
Kent, Ohio, USA*

**Anna Leonova**

*Moscow State University  
Moscow, Russian Federation*

**Claude Levy-Leboyer**

*Université René Descartes  
Paris, France*

**Joe Matarazzo**

*Oregon Health Sciences University  
Portland, Oregon, USA*

**Charles Spielberger**

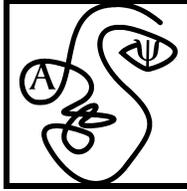
*University of South Florida  
Tampa, Florida, USA*

**Tuomo Tikkanen**

*Finnish Psychological Association  
Helsinki, Finland*

**Bernhard Wilpert**

*Technische Universität Berlin  
Berlin, Germany*



## Section Editors

**Neil Anderson**

*Universiteit van Amsterdam  
Amsterdam, The Netherlands*

**Andrew Baum**

*University of Pittsburgh  
Pittsburgh, Pennsylvania, USA*

**John W. Berry**

*Queen's University  
Kingston, Ontario, Canada*

**Mirilia Bonnes**

*Università degli Studi di Roma "La Sapienza"  
Rome, Italy*

**Heliodoro Carpintero**

*Universidad Complutense de Madrid  
Madrid, Spain*

**Giuseppe Carrus**

*Università degli Studi di Roma "La Sapienza"  
Rome, Italy*

**Patrizia Catellani**

*Catholic University of Milan  
Milan, Italy*

**Florence Denmark**

*Pace University  
New York, New York, USA*

**Rocio Fernández-Ballesteros Garcia**

*Autónoma Universidad de Madrid  
Madrid, Spain*

**Jane Goodman-Delahunty**

*University of New South Wales  
Sydney, New South Wales, Australia*

**Yuri Hanin**

*Research Institute for Olympic Sports  
Jyväskylä, Finland*

**Ann Higgins-D'Alessandro**

*Fordham University  
Bronx, New York, USA*

**Robert Hogan**

*Hogan Assessment Systems  
Tulsa, Oklahoma, USA*

**Barry Kantowitz**

*University of Michigan  
Ann Arbor, Michigan, USA*

**Uichol Kim**

*Chung-Ang University  
Seoul, Korea*

**Howard M. Knoff**

*U. S. Department of Health and Human Services  
Little Rock, Arkansas, USA*

**Frederick T. L. Leong**

*University of Tennessee  
Knoxville, Tennessee, USA*

**Peter A. Lichtenberg**

*Wayne State University  
Detroit, Michigan, USA*

**Ingrid Lunt**

*University of London  
London, United Kingdom*

**Roy Malpass**

*University of Texas  
El Paso, Texas, USA*

**Rick R. McCown**

*Duquesne University  
Pittsburgh, Pennsylvania, USA*

**José M. Prieto**

*Universidad Complutense de Madrid  
Madrid, Spain*

**Donald K. Routh**

*University of Miami  
Miami, Florida, USA*

**Mark L. Savickas**

*Northeastern Ohio Universities College  
of Medicine  
Rootstown, Ohio, USA*

**Margaret Semrud-Clikeman**

*University of Texas  
Austin, Texas, USA*

**Harry C. Triandis**

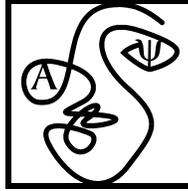
*University of Illinois, Urbana-Champaign  
Champaign, Illinois, USA*

**W. Fred van Raaij**

*Universiteit van Tilburg  
Tilburg, The Netherlands*

**Sheldon Zedeck**

*University of California, Berkeley  
Berkeley, California, USA*



## Preface

During the past 50 years, there has been remarkable progress in the articulation of psychological theory, increasingly sophisticated reports of empirical research, and a dramatic expansion in professional applications of psychology in all areas of modern life. Applied psychology requires both a sound scientific foundation and attention to the practical problems that are encountered in applying psychology to an infinite variety of life settings. Growing public recognition of the potential contributions of psychology to solving life's problems has also fostered rapid growth in the number of psychologists who are engaged in research and in professional practice throughout the world.

The need for an authoritative publication that reviews and evaluates the numerous applications of professional psychology has been recognized for a number of years, and has been of special interest to the members of the International Association of Applied Psychology (IAAP), the world's oldest and largest membership organization of individual psychologists. The interests and concerns of IAAP members, who reside in more than 80 countries, have been strongly noted by former IAAP presidents Edwin A. Fleishman, Claude Levy-Leboyer, and Harry C. Triandis.

Significant contributions to international applications of psychology were explicitly recognized and discussed in the context of their diverse societal settings by Bernhard Wilpert in his Presidential address at the IAAP's 24th International Congress of Applied Psychology in 1998. It was my honor and pleasure to succeed Professor Wilpert as President of the IAAP at this Congress, which was held in San Francisco, California, and to begin participating in discussions to identify the most important contributions to applied psychology that are reported in this encyclopedia.

The development and current status of applied psychology can be best understood by examining the historical evolution of the field, which is cogently described by Helio Carpintero, the official Historian of the IAAP. In his overview article on the history of applied psychology, Carpintero highlights the benefits resulting from psychologists joining together to form professional associations that have facilitated the development and integration of psychological theory and research and that have stimulated and guided more effective applications of psychology in professional work.

Founded in 1920 as the *Association Internationale des Conférences de Psychotechnique* (International Association for Psychotechnology), the IAAP adopted its present name in 1955. Edouard Claparède convened and chaired the first meeting of the association, held in Geneva, Switzerland in 1920, and served as its President for more than 20 years. Meetings were held in a number of European cities, including Barcelona, Milan, Paris, Moscow, and Prague. An excellent history of the evolution and development of applied psychology and the IAAP can be found in the address given by former President Edwin A. Fleishman on receiving the 1999 American Psychological Association's Award for Distinguished Contributions to the International Advancement of Psychology (*American Psychologist*, November, 1999, pp. 1008–1016).

For recommending that the IAAP take an active role in sponsoring and developing the *Encyclopedia of Applied Psychology*, we are indebted to George Zimmar, an active IAAP member and former senior sponsoring editor at Academic Press, now Elsevier. The first step in developing the encyclopedia was the appointment of an Advisory Board composed of distinguished colleagues representing all areas of

applied psychology. In addition to recommending the major areas and the specific topics to be covered in the encyclopedia, the Advisory Board identified highly qualified colleagues who would be invited to serve as Section Editors for the selected areas.

The responsibilities of the Section Editors included writing a major article to provide an overview and analysis of their respective areas and working with the Editor-in-Chief and the Advisory Board in identifying the specific topics to be included in each area. They also helped to determine the amount of space to be allocated for each topic and recommended colleagues with outstanding knowledge and expertise to serve as the authors for the individual articles. In addition, the Section Editors reviewed and edited the articles that presented information on the topics that were related to the subject matter that was included in their sections.

The articles that are included in the encyclopedia are presented in alphabetical order, as determined by the title of each article. Consequently, the underlying section organization will not be readily apparent. At the conclusion of each article, the reader is directed to additional articles that include related information. The authors of each article also provide a number of references that are recommended for further reading.

It should be noted, however, that it was not possible to cover all of the many areas of applied psychology. Several emerging areas are not included because we failed to recognize their importance at the time the topics were selected for the encyclopedia. It is also unfortunate that a few authors were unable to meet the deadlines for submitting the manuscripts for their articles.

The *Encyclopedia of Applied Psychology* is truly a product of the collaborative efforts of the members of the IAAP. We are especially indebted to former IAAP Presidents Edwin A. Fleishman, Harry C. Triandis, and Bernhard Wilpert, and to our current president, Michael Frese, for their leadership, encouragement, and tangible contributions to the encyclopedia. For their dedicated contributions to the encyclopedia, I would also like to thank the members of the Advisory Board and the Section Editors, whose names are listed in the front matter. Finally, on behalf of the IAAP, I thank the authors who contributed articles in their areas of expertise and the editorial and production staff of Elsevier for the tremendous amount of time and energy that was needed to bring this large project to fruition.

CHARLES D. SPIELBERGER, Ph.D., ABPP  
*President, IAAP, 1998–2002*



# Academic Failure, Prevention of

Mary Ann Rafoth

Indiana University of Pennsylvania, Indiana, Pennsylvania, USA

1. Defining Academic Failure
2. Causes of Academic Failure
3. Preventing Early School Failure
4. Preventing Failure in the Intermediate Grades and Middle School
5. Preventing Academic Failure in High School
6. Conclusion  
Further Reading

## GLOSSARY

**basic skills** The skills associated with functional literacy, including reading decoding and comprehension skills, math calculation and problem solving, and writing.

**cooperative learning** An instructional approach that uses peer groups to facilitate and reinforce learning.

**curriculum-based assessment** An assessment approach that uses students' progress in their actual curriculum as a measurement point.

**emergent literacy** The recognition of environmental print that leads to gradual understanding of the role of letters and words in written language.

**learning disability** A discrepancy between a student's measured ability and actual achievement caused by a dysfunction in a basic neuropsychological process.

**learning strategy** A voluntary activity students use to facilitate remembering, learning, or problem solving.

**mastery learning** An instructional approach that allows students to repeat instruction and assessment until competency is achieved.

**peer tutoring** An instructional approach that uses peers to reinforce classroom instruction by providing individual instruction to students.

**phonological awareness** Children's awareness of the way in which words sound, particularly rhyme and alliteration and the ability to blend sounds, recognize onset rime (initial sounds), and identify sound units.

**retention** The practice of requiring a student to repeat a grade or requiring a child of appropriate chronological age to delay entry to kindergarten or first grade.

**school readiness** The set of skills children need to learn successfully in school.

**social promotion** The practice of promoting a student to the next grade even when skills or knowledge sets to be mastered at that level have not been mastered.

**team teaching** An instructional approach that allows teachers, often with different areas of expertise, to share instructional duties.

Prevention of academic failure is a serious challenge because children who fail academically experience significant social and economic challenges throughout their lives. Causes of academic failure include familial, socioeconomic, and cultural issues that lead to a lack of readiness for school, academic, instructional, and motivational problems as well as physiological, cognitive, and neurological barriers to learning. Attempts to help students who are experiencing academic failure fall into three categories: prevention, intervention, and remediation. Preventive approaches

aim to stop academic failure before it occurs. Early intervention programs aim to catch children during key developmental periods and facilitate development and readiness skills. Remediation programs are usually applied when students have demonstrated significant skill deficits and are experiencing significant academic failure. Special education programs often take this form, as do other kinds of academic accommodations for students identified with special needs.

## **1. DEFINING ACADEMIC FAILURE**

A lot depends on children's success in school—their self-esteem, their sense of identity, their future employability. Preventing academic failure means that we, as a society, are much more likely to produce individuals who feel confident about their ability to contribute to the common good, whose literacy skills are competent, and who are able to hold jobs successfully. Thus, prevention of academic failure should be a primary concern for any society. But exactly what is meant by academic failure? What does the term connote? Generations of schoolchildren since the 1920s, when the system of grade progression began, have equated academic failure with retention in grade. School failure meant literally failing to progress onto the next grade, with the assumption that the skills and knowledge taught in that grade had not been mastered. To have flunked multiple grades quickly led to quitting school altogether—the ultimate academic failure.

More recently, academic failure has come to mean a failure to acquire the basic skills of literacy. Students who were unable to read at a functional level, to communicate effectively through writing, and to complete basic math calculations were seen as representing a failure of the academic system even though they might hold high school diplomas. The practice of moving students on from one grade to the next even though they might not have mastered basic competencies associated with lower grade levels is often referred to as social promotion. This type of academic failure led to calls for an increased emphasis on basic skills, that is, the “three R's”—reading, (w)riting, and (a)rithmetic—in public education. Partly in reaction to emphasis on basic skills, a third interpretation of academic failure has also emerged. In this view, academic failure occurs not only when students fail to master basic skills but also when they emerge from school without the ability to think critically, problem solve,

learn independently, and work collaboratively with others—a skill set deemed necessary for success in a digital age. This underachievement symbolizes a significant loss of intellectual capital for a culture. Finally, statistics show that students who do not complete high school are much more likely to need welfare support, have difficulties with the law and police, and struggle economically and socially throughout their lives. Thus, academic failure ultimately means both the failure to acquire the skill sets expected to be learned and the failure to acquire official documentation of achievement by the school system.

## **2. CAUSES OF ACADEMIC FAILURE**

Students struggle academically for many reasons, including familial, socioeconomic, and cultural issues that lead to a lack of readiness for school, academic, instructional, and motivational problems as well as physiological, cognitive, and neurological barriers to learning. Early school failure often occurs because children enter the structured school environment not ready to learn.

### **2.1. School Readiness**

School readiness refers to the idea that children need a certain set of skills to learn and work successfully in school. Often this term refers to whether or not children have reached the necessary emotional, behavioral, and cognitive maturity to start school in addition to how well they would adapt to the classroom environment. To create some consensus about when a child should begin school, states designate a specific cutoff date. If a child reaches a certain age by the cutoff (usually 5 years for kindergarten and 6 years for first grade), the child may begin school. However, cutoff dates are arbitrary and vary considerably across nations, and age is not the best determinant or most accurate measure of whether or not a child is ready to begin school. Research has suggested that we must look at all aspects of children's lives—their cognitive, social, emotional, and motor development—to get an accurate idea of their readiness to enter school. Most important, children's readiness for school is affected by their early home, parental, and preschool experiences.

Stated in its simplest form, school readiness means that a child is ready to enter a social environment that is focused primarily on education. The following list of

behaviors and characteristics are often associated with school readiness:

- Ability to follow structured daily routines
- Ability to dress independently
- Ability to work independently with supervision
- Ability to listen and pay attention to what someone else is saying
- Ability to get along with and cooperate with other children
- Ability to play with other children
- Ability to follow simple rules
- Ability to work with puzzles, scissors, coloring, paints, and the like
- Ability to write own name or to acquire the skill with instruction
- Ability to count or acquire skills with instruction
- Ability to recite the alphabet
- Ability to identify both shapes and colors
- Ability to identify sound units in words and to recognize rhyme

Family environment is very important in shaping children's early development. Some family factors that can influence school readiness include low family economic risk (poor readiness for school is associated with poverty), stable family structure (children from stable two-parent homes tend to have stronger school readiness than do children from one-parent homes and from homes where caregivers change frequently), and enriched home environment (children from homes where parents talk with their children, engage them in conversation, read to them, and engage in forms of discipline such as "time-out" that encourage self-discipline have stronger readiness skills).

Children's readiness to read, in particular, has gained greater attention from educators recently as the developmental precursors to reading have become more evident. During the preschool years, children develop emerging literacy skills—preacademic skills that allow children to develop a disposition to read, write, and compute. Children are ready to read when they have developed an ear for the way in which words sound and can identify rhyme and alliteration, blend sounds, recognize onset rime (initial sounds), and identify sound units in words. Together, these skills are called phonological awareness and usually emerge in children between 2 and 6 years of age. Children with good phonological awareness skills usually learn to read quickly. Children who are poor readers have weak phonological skills, and children who do not learn to read fail in school. Another important readiness skill that helps children

to learn to read is called print awareness. Print awareness means that children are capable of the following:

- Knowing the difference between pictures and print
- Recognizing environmental print (e.g., stop signs, McDonald's, Kmart)
- Understanding that print can appear alone or with pictures
- Recognizing that print occurs in different media (e.g., pencil, crayon, ink)
- Recognizing that print occurs on different surfaces (e.g., paper, computer screen, billboard)
- Understanding that words are read right to left
- Understanding that the lines of text are read from top to bottom
- Understanding the function of white space between words
- Understanding that the print corresponds to speech word for word
- Knowing the difference between letters and words

Children also need to learn book-handling skills such as orienting a book correctly and recognizing the beginning and the end of a book. Children who begin school without these basic readiness skills are at risk for school failure. The use of screening assessments during preschool and kindergarten to identify students who may be at risk for academic failure, particularly in the area of phonemic awareness, has been shown to be a sound method of predicting which children will have difficulty in learning to read. Most likely to be retained in kindergarten are children who are chronologically young for their grade, developmentally delayed, and/or living in poverty.

## 2.2. Academic, Instructional, and Motivational Reasons

Children who do not master basic reading skills, specifically the ability to automatically decode new words and build a sight word vocabulary that leads to fluency, experience academic failure. By third grade, learning to read has become reading to learn. In other words, in third grade the curriculum becomes focused much less on teaching students to acquire the basic tools of literacy (reading, writing, and computing) and much more on using those tools to learn content, express ideas, and solve problems. At this point, students are likely to be given content textbooks in science and social studies and to read nonfiction for the purpose of gaining new information. Thus, the inability to read effectively and to learn

to study independently often leads to failure at the elementary and middle school levels and also creates profound motivation problems at the high school level that contribute to the ultimate school failure—dropping out.

The inability to master key concepts in pivotal classes such as algebra, now typically taken at the middle or junior high school level, often limits students' ability to proceed in coursework. Students may fail to understand algebraic concepts due to their developmental level. (Many students are still thinking in concrete terms in middle school and have not yet moved into a stage of cognitive thinking allowing them to understand formal logic and manipulate symbols—a developmental source of failure.) In addition, some students might not have automatized basic arithmetic skills, particularly computing with fractions—an academic or instructional failure. Some students may have become turned off to math and accepted self-images that permit poor math skills—a motivational failure. Finally, many students will fail algebra for all of these reasons, and the impact will often be that they will finish school in a nonacademic or basic track or might even drop out.

Thus, academic and instructional reasons for school failure include the effectiveness of the instruction a student has received and the quality of remediation strategies or programs available. The following is a typical example that illustrates academic and instructional reasons for school failure. A teacher reports that a student is having difficulty in getting beyond the primer level in reading and is being considered for retention. The child was assessed as having average intelligence. No behavioral or attention problems were noted. Closer inspections of the student's reading skills indicated that she had poor phonological skills and was not profiting from the type of classroom reading instruction she was receiving that depended heavily on auditory phonics instruction stressing "sounding out words" and matching sound-symbol connections. Appropriate interventions included using techniques to build up a sight word vocabulary through repetition and distributed learning and introducing the student to a visual decoding system to provide her with a method for reading unknown words by analyzing the words and breaking them down into more familiar visual units.

### **2.3. Physiological, Neurological, and Cognitive Reasons**

Imagine a child spending most of the year in kindergarten with an undetected hearing loss that has made it

very difficult for her to benefit from instruction. Imagine another child in first grade struggling to learn because her vision impairment has not been caught or corrected. Similarly, students suffering from a variety of conditions and illnesses, such as childhood diabetes, asthma and allergy-related problems, and sickle cell anemia, may have difficulty in maintaining energy and attention in school due to chronic fatigue and the impact of medications. Children may also suffer from orthopedic or motor impairments that make it difficult for them to explore their environment, interact with others, and/or master tasks that demand motor skills.

Students who suffer from various kinds of neurological disorders or learning disabilities may also have cognitive learning problems that make it difficult for their brains to process information, interpret sounds and symbols efficiently in reading, calculate and understand number concepts, and/or write effectively. Other children may have cognitive deficits, such as mental retardation, that limit their ability to absorb and apply regular classroom instruction. Children with attention deficit disorders have difficulty in directing and maintaining their attention, may exhibit impulsive behavior, and have trouble in interacting independently in typical classroom environments without support. Specialized and/or special education interventions are designed to provide individualized strategies and approaches for students who have physiological-based learning problems interfering with their ability to learn.

## **3. PREVENTING EARLY SCHOOL FAILURE**

### **3.1. Early Intervention Programs**

Programmatic interventions may include developing screening programs to identify children at risk for school failure and to ensure early access to readiness programs already available in the school or community such as Head Start. Many states are now developing guidelines for children age 6 years or under based on the National Association for the Education of Young Children's (NAEYC) list of developmentally appropriate practices. The major challenge facing early intervention programs is to provide developmentally and individually appropriate learning environments for all children. Essential ingredients to successful preschool experiences include small group and individualized teacher-directed activities as well as child-initiated

activities. Quality programs recognize the importance of play and view teachers as facilitators of learning.

### 3.2. Preventing School Failure in the Elementary Grades

Full day kindergarten (as opposed to half day) programs provide more time for field trips, activity centers, projects, and free play. At-risk students who attend rigorous yet nurturing full day programs have a greater chance of experiencing academic success. Full day kindergarten programs help increase academic achievement as well as decrease the number of children retained in the early elementary grades. Research shows that full day kindergarten programs for children who come from disadvantaged backgrounds lead to stronger achievement in basic skill areas and generally better preparation for first grade.

Table I summarizes the research on full-day kindergarten. Any decisions about whether or not to schedule

full- or half-day programs should recognize that what a child is doing during the kindergarten day is more important than the length of the school day.

The instructional technology that enables classroom teachers to meet the needs of students of different skill levels is already available, but in many cases teachers do not have access to that technology. Reading interventions that provide intensive, early, and individualized help that targets a child's specific weaknesses (e.g., Success for All, Reading Recovery, Direct Instruction) have been shown to be effective in reducing early reading failure. Instructional approaches such as mastery learning, adaptive education, team teaching, cooperative learning, peer tutoring, and curriculum-based assessment are methods that have been shown to produce academic gains in students of all achievement levels in the elementary grades. Recently, technology has offered greater individualization of instruction and increased flexibility in allowing students to progress at their own pace and to respond to instruction.

TABLE I  
Research Summary: Full-Day Versus Half-Day Kindergarten Programming

<i>Research on effects of full-day programs on achievement</i>	<i>Research on effects of full-day programs with educationally disadvantaged and low-SES children</i>	<i>Research on other effects of full-day versus half-day programs</i>
Full-day programs increase long-term achievement.	Full-day programs allow for greater opportunity for assessment and screening.	Full-day programs have been shown to raise children's self-esteem, independence, and creativity. There is no evidence that the curriculum is more individualized or innovative.
Full-day programs produce higher reading scores in second and third grades.	Full-day programs result in higher academic and social gains.	Parental involvement is diminished in full-day programs.
Full-day programs result in fewer grade retentions, higher report card marks, and higher scores on standardized tests.	Full-day programs allow for more informal and formal instructional time on an individual basis.	Full-day programs are not found to cause excessive fatigue or stress in children.
Full-day programs provide more time for individualized instruction, academics, and the reinforcement of children's positive self-images.	Full-day programs provide access to nutritional lunch as well as breakfast.	Full-day programs allow for an unhurried relaxed school day with more variety of experiences.
Full-day programs are found to have no effect on achievement.	Title I children have significantly higher achievement in full-day programs.	Children in full-day programs have been shown to be less likely to be dependent, shy, and withdrawn. Children in half-day programs spend more time in large groups than in free play.

Note. SES, socioeconomic status.

#### **4. PREVENTING FAILURE IN THE INTERMEDIATE GRADES AND MIDDLE SCHOOL**

Remedial programs, such as the Title I or Chapter 1 programs, have also been used to remediate early skill deficits in reading and math. However, developing intervention programs, such as after-school tutoring and summer school courses, might not be sufficient to make up serious deficits in short amounts of time and cannot take the place of preventive systemic approaches.

The use of learning strategies instruction has been shown to be very effective in improving study skills and performance in middle school students. Because unsuccessful middle school students often lack basic strategic learning skills, intervention programs should also target these areas. Similarly, approaches that use learning, problem-solving, and memory strategies are the most effective interventions in terms of producing actual gains in student achievement in the classroom.

#### **5. PREVENTING ACADEMIC FAILURE IN HIGH SCHOOL**

At the secondary level, development of reentry programs for dropouts and alternative education programs, such as those that combine teaching skills with job training, are essential to prevent further academic failure. Research on academic failure at the secondary level has generally examined the relationships between grade retention and attendance, suspension, and self-concept, with an emphasis on the correlation between retention and dropout rates. Academic failure at the high school level is related to attendance and suspension rates. In general, students who are failing do not attend school on a regular basis. In addition, students who have been retained prior to the secondary level are less likely to attend school on a regular basis in junior and senior high school. Furthermore, regardless of the grade in which retention occurs, secondary students who have been retained often exhibit low self-esteem.

Many studies have reported that students who drop out are five times more likely to have repeated a grade than are students who eventually graduate. Being retained twice virtually guarantees that a student will drop out of school, and grade retention alone has been identified as the single most powerful predictor of dropping out. The dropout rate of overage students is appreciably higher than the dropout rate of regularly

promoted students when reading achievement scores are equivalent for the two groups. Even in high-socioeconomic school districts, where students are less likely to leave school, a significant increase in dropout rates has been found for retained students.

Successful programs at the high school level often have two characteristics: (a) one or more individuals who develop relationships with students individually and monitor their progress carefully and (b) some mechanism to allow students who have failed courses and lost credits to regain these credits in quicker than normal time, allowing for graduation at the expected time. Simply put, successful programs must address the motivational issues that have developed by adolescence and the lack of academic achievement identity typically present in students who drop out of school. School-to-work programs that combine vocational counseling with on-the-job experience are successful ways in which to increase a sense of academic competence while connecting to students' current self-concepts and needs.

#### **6. CONCLUSION**

Attempts to help students who are experiencing academic failure fall into three categories: prevention, intervention, and remediation. Preventive approaches aim to stop academic failure before it occurs. Early intervention programs from birth to 5 years of age, for example, aim to catch children during key developmental periods and facilitate development and readiness skills. Intervention programs, such as Robert Slavin's Success for All program, aim to intervene as soon as students begin to show signs of slipping behind their peers. Intervention plans may also be designed under Section 504 of the Americans with Disabilities Act, which mandates accommodations in the instructional environment for students who have physical or neurological problems that may interfere with their ability to learn or succeed in a typical classroom. Remediation programs are usually applied when students have demonstrated significant skill deficits and are experiencing significant academic failure. Special education programs often take this form, as do other kinds of academic accommodations for students identified with special needs.

Of course, early identification and prevention of academic problems is always preferable to later intervention and remediation. Thus, systemic solutions that target early reading deficits, independent learning skills, and motivational problems from a developmental

perspective are essential to the prevention of academic failure. Working to change school practices will require sharing the research with educators, conducting evaluations on the outcomes of alternative interventions at the local level, and lobbying at the state level to promote changes in policy and to advocate for alternative service delivery systems that more effectively meet the needs of students experiencing school failure. Successful programs to boost student achievement, however, must attack underachievement in three key areas. These key areas—early reading intervention, acquisition of strategic learning and study skills, and motivation to achieve—are highly related to school failure.

First, acquisition of basic reading skills must be addressed. If students underachieve in the primary grades, it is most often because they have failed to learn to read. Kindergarten screenings should include an assessment of phonological awareness. Children identified with weak skills should be targeted for intervention through phonological awareness training in kindergarten. Prekindergarten programs for high-risk students are recommended. Students should be tracked using curriculum-based assessments of oral reading in the primary grades. Any student who falls behind the average rate of acquisition for his or her class should receive an individualized analysis of reading skill and additional after-school intervention based on that analysis to allow the student to “catch up” to classmates. This early, intensive, and individualized intervention allows for all students to enter the intermediate grades as able readers. Some students with special needs might not progress at the same rate as their classmates, but they too will benefit from early reading interventions.

Second, students must acquire independent learning and study skills during the intermediate and middle school years if they are to maximize achievement and be competitive in the job market of tomorrow. Many students underachieve in middle school because they lack the organizational and learning strategies to master the demands of the upper grades. Embedded approaches to strategy instruction facilitate generalization and encourage students to use all of their mental tools. Assessment of students’ study skills and metacognitive development (i.e., the degree to which they are aware of and control their own cognitive processes) leads directly to specific interventions.

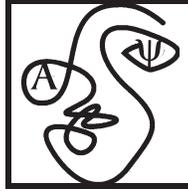
Third, students in high school often underachieve because they lack the motivation to excel academically. They often have failed to incorporate pictures of themselves as successful students into their self-concepts. Through a variety of approaches, including staff in-service, a study skills coach approach to peer tutoring, and an individualized profile of each student’s study style and vocational options, increased academic competence and a value for academic work can be built.

### See Also the Following Articles

Academic Interventions ■ Advertising and Culture ■ Educational Achievement and Culture ■ Educational and Child Assessment ■ Effective Classroom Instruction ■ Learning ■ Learning Styles and Approaches to Studying ■ Mathematics, Teaching of ■ Reading, Teaching of ■ Teaching Effectiveness ■ Transfer of Learning ■ Writing, Teaching of

### Further Reading

- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn*. Washington, DC: National Academy of the Sciences Press.
- Elicker, J. (2000). *Full day kindergarten: Exploring the research*. Bloomington, IN: Phi Delta Kappa International.
- Gredler, G. (1992). *School readiness: Assessment and educational issues*. Brandon, VT: CPPC Publishing.
- Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology, 37*, 243–272.
- Minke, K. M., & Bear, G. G. (Eds.). (2000). *Preventing school problems, promoting school success: Strategies and programs that work*. Silver Spring, MD: National Association of School Psychologists.
- National Institute of Child Health and Development. (2000). *Report of the National Reading Panel: Report of the subgroups*. Bethesda, MD: Author.
- Slavin, R. E., Karweit, N. L., & Madden, N. A. (1989). *Effective programs for students at risk*. Boston: Allyn & Bacon.
- Smith, M. L., & Shepard, L. A. (Eds.). (1989). *Flunking grades: Research and policies on retention*. New York: Falmer.
- Springfield, S., & Land, D. (Eds.). (2002). *Educating at-risk students*. Chicago: National Survey of Student Engagement. (Distributed by University of Chicago Press).



# Academic Interventions

Natalie Rathvon

Center City Consortium Schools, Washington, D.C., USA

1. Overview
  2. Types of Academic Problems
  3. Intervention Delivery Systems
  4. Guidelines for Selecting, Implementing, and Evaluating Academic Interventions
  5. Academic Intervention Targets
  6. Summary and Future Directions
- Further Reading

This article reviews the field of academic interventions. It describes types and targets of academic interventions; intervention delivery systems; guidelines relating to the selection, implementation, and evaluation of academic interventions; and selected evidence-based strategies.

## 1. OVERVIEW

Today's educators are encountering increasing numbers and proportions of students who have trouble achieving at grade-level expectations. Whether these students are termed difficult to teach, learning disabled, or at risk, they fail to respond to traditional instructional methods. The need for academic interventions is underscored by reports from the National Assessment of Educational Progress (NAEP) indicating that fewer than one-third of American fourth- and eighth-grade students are performing at proficient levels in reading, mathematics, and writing. Recent federal legislation, including the 1997 reauthorization of the Individuals with Disabilities Education Act, which promotes the use of interventions prior to referral for special education services, and the reauthorization of the Elementary and Secondary Education Act in the No Child Left Behind Act of 2001, with its mandates for empirically validated practices, is also spurring interest in strategies that can enhance teachers' capacity to meet students' needs and students' capacity to respond to instruction.

## GLOSSARY

- academic enablers* Nonacademic skills, behaviors, and attitudes that permit students to benefit from instruction.
- academic engaged time* The proportion of instructional time in which students are actively engaged in learning.
- academic intervention* A modification in the classroom environment and/or instructional practices designed to produce positive changes in academic performance.
- intervention assistance programs (IAPs)* Programs based on a consultation model of service delivery and designed to help teachers provide interventions in the regular classroom for difficult-to-teach students; they are also called prereferral intervention programs.
- intervention scripts* Detailed guidelines to assist intervention agents in delivering an intervention as intended.
- social validity* The degree to which an intervention is judged as acceptable and important by intervention participants and stakeholders, including the social significance of the goals, the social acceptability of the procedures, and the social importance of the outcomes.
- treatment integrity* The degree to which an intervention is implemented as planned.

## 2. TYPES OF ACADEMIC PROBLEMS

Academic problems may be characterized as skill deficits, fluency deficits, performance deficits, or some combination of these. Skill deficits refer to deficiencies reflecting inadequate mastery of previously taught academic skills. Fluency deficits refer to deficiencies in the rate at which skills are performed accurately. Students with performance deficits possess adequate skills and fluency but do not produce work of satisfactory quantity, quality, or both. Many of the interventions presented in [Table I](#) are designed to enhance skill acquisition and fluency by increasing opportunities to respond to academic material. Others target performance problems by using self-monitoring or contingency-based procedures, especially group-oriented contingencies that capitalize on peer influence to encourage academic productivity and motivation.

## 3. INTERVENTION DELIVERY SYSTEMS

Academic interventions can be implemented through a variety of delivery systems, including (a) case-centered teacher consultation, (b) small-group or classroom-centered teacher consultation, (c) staff development programs, and (d) intervention assistance programs (IAPs). For academic interventions with a home component, intervention services can be delivered through case-centered parent consultation, parent training programs, or parent participation in IAPs.

### 3.1. Intervention Assistance Programs

IAPs are based on a consultation model of service delivery and are designed to increase the success of difficult-to-teach students in the regular classroom by providing consultative assistance to teachers. Since the 1990s, IAPs have become widespread, with the majority of states now requiring or recommending interventions prior to special education referral. Several IAP approaches based on collaborative consultative models of service delivery have been developed to meet the needs of difficult-to-teach students in the regular classroom. These models fall into two general categories depending on whether special education personnel are involved. Key factors in successful implementation and maintenance of IAPs include administrative support, provision of high-quality

interventions, and support of teachers during the intervention process. Although an increasing body of evidence supports the efficacy of IAPs in reducing referrals to special education and improving teachers' attitudes toward diverse learners, relatively few studies have documented that IAPs produce measurable gains in student performance, and many of the studies reporting academic improvement suffer from methodological problems.

## 4. GUIDELINES FOR SELECTING, IMPLEMENTING, AND EVALUATING ACADEMIC INTERVENTIONS

In designing interventions for students with academic problems, intervention effectiveness can be enhanced by following nine guidelines that are presented in [Table II](#) and serve as the foundation for the discussion that follows. The guidelines reflect the importance of balancing treatment efficacy with usability considerations to accommodate the realities of today's classrooms.

### 4.1. Selecting Academic Interventions

Despite the growing database of evidence-based interventions, studies indicate that teachers and IAPs continue to rely on interventions characterized by familiarity or ease of implementation rather than on those with documented effectiveness. The importance of considering efficacy in intervention selection has been underscored by the Task Force on Evidence-Based Interventions in School Psychology, jointly sponsored by the Division of School Psychology of the American Psychological Association and the Society for the Study of School Psychology and endorsed by the National Association of School Psychologists. Founded in 1998 as an effort to bridge the often-cited gap between research and practice, the Task Force has developed a framework with specific efficacy criteria for evaluating empirically supported intervention and prevention programs described in the literature.

Consultants should also give priority to proactive interventions that help teachers to create learning environments that prevent academic problems from occurring by promoting on-task behavior and productivity rather than using reactive strategies that are applied after problems have already developed. Academic achievement is significantly related to the amount of time allotted for instruction and to academic engagement

**TABLE 1**  
**Descriptions of Selected Academic Interventions<sup>a</sup>**

<i>Intervention</i>	<i>Treatment components</i>	<i>Description</i>	<i>Target(s)</i>	<i>Primary intervention agent(s)</i>
<b>Interventions to improve academic productivity</b>				
Self-management	Goal setting Self-monitoring Self-evaluation Self-reinforcement	Students set goals for work completion and/or accuracy and monitor their own progress	On-task behavior Classwork completion and/or accuracy Homework completion and/or accuracy Academic achievement	Students
Classwide peer tutoring	Increased opportunities to respond Timing Self-management Peer-mediated instruction and feedback Error correction	Students practice oral reading, math facts, or other basic academic skills in pairs	On-task behavior Classwork completion and/or accuracy Academic achievement	Peers
<b>Interventions to improve reading performance</b>				
Listening previewing	Modeling Guided practice Performance feedback Error correction	Students read a passage aloud after listening to a proficient reader read the same passage	Reading accuracy and fluency	Teacher, peers, or parents
Collaborative strategic reading	Strategic instruction Modeling Guided practice and feedback Peer-mediated instruction and feedback	Students learn four basic reading strategies and then work in pairs or groups to implement the strategies	Reading comprehension and vocabulary Study skills	Teacher and peers
<b>Interventions to improve mathematics performance</b>				
Cover-copy-compare	Increased opportunities to respond Error correction Timing Self-instruction, evaluation, and reinforcement	Students respond to questions, compare their answers with instructional prompts, and correct their own errors	Academic productivity, accuracy, and fluency	Teacher

*Continues*

Continued

<i>Intervention</i>	<i>Treatment components</i>	<i>Description</i>	<i>Target(s)</i>	<i>Primary intervention agent(s)</i>
Reciprocal peer tutoring	Increased opportunities to respond Error correction Timing Self-management Peer-mediated instruction and feedback Home-based reinforcement (optional)	Tutoring pairs select their own math goals and practice computational skills in pairs	Mathematics productivity, accuracy, and fluency	Peers
<b>Interventions to improve written language performance</b>				
Self-regulated strategy development in writing	Explicit teaching Modeling Guided feedback Goal setting Self-monitoring, self-instruction, and self-reinforcement Collaborative practice	In a series of lessons, students learn writing skills and strategies as well as procedures for regulating strategy use and the writing process	Writing mechanics Written expression skills Attitudes toward writing	Teacher Peers
Self-recording of writing productivity	Goal setting Self-monitoring Self-evaluation Individual or group reinforcement (optional)	Students evaluate their progress in writing by counting and recording the numbers of words they write	Writing fluency Writing productivity Attitudes toward writing	Students
<b>Interventions to promote academic enablers</b>				
Public posting	Performance Feedback Teacher praise Individual or group contingencies (optional)	Student grades or work completion rates are displayed individually or by teams	On-task behavior Classwork completion and/or accuracy Homework completion and/or accuracy	Teacher
School-home notes	Performance feedback Teacher and parent praise Home-based contingencies (optional)	Students share ratings of academic productivity, and/or achievement with parents and may receive social or material rewards for improvement	Classroom behavior Classwork and/or homework accuracy and/or productivity Attitudes toward school	Teacher Parent

<sup>a</sup>Adapted from *Evidence-Based School Interventions* by N. Rathvon (in press) New York: Guilford Press. Adapted with permission.

**TABLE II**  
**Best Practices in Selecting, Implementing, and Evaluating Academic Interventions<sup>a</sup>**

- 
- A. Select research-based strategies with a documented history of success with the target student population, target setting, and target skill or subskill.
  - B. Give priority to strategies that are proactive in nature and have the potential of enhancing the achievement of groups of students.
  - C. Select interventions that are minimally intrusive, fit into the ecology of the classroom setting, and capitalize on the human and material resources already present in school ecosystems.
  - D. Involve key stakeholders, including teachers, parents, and students themselves, if developmentally appropriate, in the intervention process.
  - E. Be sensitive to cultural diversity issues in the consultation and intervention processes.
  - F. Use a variety of training formats to expand teachers' knowledge of interventions and the intervention process.
  - G. Build in treatment integrity checks to ensure accurate implementation.
  - H. Provide classroom-based assistance to teachers in the form of modeling, guided practice, and feedback during the implementation and evaluation phases.
  - I. Design evaluation strategies for assessing changes in student performance that are reliable, valid, and practical for regular classroom teachers.
- 

<sup>a</sup>Adapted from *Effective School Interventions: Strategies for Enhancing Academic Achievement and Social Competence* by N. Rathvon, 1999, New York: Guilford Press. Copyright 1999 by Guilford Press. Adapted with permission.

rates, that is, the proportion of instructional time in which students are actively engaged in learning as demonstrated by behaviors such as paying attention, working on assignments, and participating in class discussions. Although all students profit from proactive strategies that increase instructional time and academic engagement, such interventions are especially important for diverse learners, who are more likely to need additional practice on academic tasks to keep up with their grade peers.

Academic interventions should be minimally intrusive so that they can be implemented in regular classroom settings without unduly disrupting instructional and management routines. Interventions with low ecological validity (i.e., strategies that require major alterations in classroom procedures or cumbersome reinforcement delivery systems) are unlikely to become integrated into regular education routines. Priority should also be given

to strategies that benefit more than one student. Traditional case-centered intervention approaches directed at a single low-performing student are of limited utility in enhancing teachers' overall instructional effectiveness and can be very time-consuming. Group-focused strategies are both time- and labor-efficient, are more acceptable to teachers and students alike, and foster a positive classroom climate by incorporating peer influence.

Moreover, intervention targets and procedures should be reviewed in terms of their social validity. Although many interventions target on-task behavior, increasing on-task behavior is less socially significant as an intervention goal than is enhancing rates of academic responding because increasing on-task behavior does not necessarily result in higher student achievement. Although there is little systematic research on the benefits of involving parents and students in designing interventions, input should be obtained from all stakeholders on the acceptability of proposed intervention goals and procedures.

In designing interventions, consultants must be sensitive not only to individual differences but also to differences in cultural values and norms. For example, interventions that provide concrete operant reinforcers for academic performance might be considered unacceptable by individuals from certain cultures. Unfortunately, little research to date has examined the relative efficacy and acceptability of various academic interventions with culturally and linguistically diverse learners. However, by maintaining an awareness of their own ethnocentrism and encouraging an open dialogue throughout the intervention process, consultants will be better prepared to adjust intervention services to address the needs of stakeholders from nonmainstream groups.

## 4.2. Implementing Academic Interventions

Regardless of the quality of the intervention design and the documented efficacy of the intervention components, no strategy will be effective in improving student achievement unless the teacher implements it accurately and consistently, that is, with treatment integrity. In the absence of treatment integrity measures, there is no way to determine whether changes in academic performance are due to the effects of the intervention or to factors that are unrelated to intervention components. Among the factors influencing treatment integrity are (a) intervention complexity, (b) time and

material resources required for implementation, (c) the number of intervention agents, (d) efficacy (actual and as perceived by the intervention agents and stakeholders), and (e) the motivation of the intervention agents and stakeholders. Strategies for enhancing treatment integrity include (a) delivering interventions by means of a videotape or an audiotape, (b) documenting consultation contacts, (c) using an intervention manual or script, (d) having a written intervention plan, and

(e) providing direct feedback to intervention agents during implementation. **Table III** presents an intervention script for listening previewing, a strategy targeting reading achievement. Treatment integrity can be assessed by direct observation, videotaping or audiotaping intervention sessions, or teacher-, parent-, or student-completed fidelity checklists.

Consultants seeking to help teachers implement classroom strategies must be prepared to provide

**TABLE III**  
**Listening Previewing<sup>a</sup>**

---

#### Overview

*Listening previewing*, also termed *passage previewing*, has been shown to improve oral reading performance for a variety of student populations. In this intervention, the teacher reads the assigned passage aloud while students follow along silently prior to independent reading. Higher reading rates have been obtained with bilingual learners and students with learning disabilities, as well as with students in general education. Listening reviewing has also been shown to be superior in improving reading performance to previewing by the traditional silent reading method. Variations include using peers, parents, or tutors as passage previewers.

#### Purpose

To enhance reading accuracy and fluency by providing an opportunity for students to hear what they will read prior to reading it for themselves.

#### Materials

1. Stopwatch or watch with second hand (for assessment component)
2. Sheets of paper or copies of students' reading passages for marking oral reading errors (optional)
3. Notebooks or folders with loose-leaf paper for graphing words read correctly per minute (WCPM) rates, one per student (optional)

#### Observation (Select one or both)

##### Option 1

1. Administer curriculum-based measurement (CBM) oral reading probes to the students in a reading group or a group of selected students by having each student read a 2-minute timed sample from the previous day's reading passage.
2. Calculate each student's WCPM rate by subtracting the number of errors from the total number of words read and then dividing that figure by two (for number of minutes).
3. Record and graph (or have students graph) WCPM rates.

##### Option 2

1. Calculate scores on reading skill sheets, quizzes, or end-of-unit tests for all of the students in a reading group or for selected students for 5 to 10 days or for several weeks.

#### Procedure

1. Divide the reading lesson into passages according to the number of students in the reading group.
  2. Tell the students to follow along as you read the first passage aloud.
  3. After completing the first listening preview, have the first student read that passage aloud.
  4. Repeat this process until you have previewed the entire reading lesson and each student in the reading group has read aloud. If desired, record the number of errors (mispronunciations, substitutions, and omissions) on your copy of the passage or on a separate sheet of paper.
  5. Calculate and graph WCPM rates for students in the reading group as often as possible (at least once a week for several weeks) or teach students to chart their own progress on line graphs in their work folders.
- 

*Continues*

Continued

---

**Evaluation (Select one or both)**

*Option 1*

1. Compare WCPM rates for the target students before and after implementation.

*Option 2*

1. Compare scores on reading skill sheets, quizzes, or end-of-unit tests for the target students before and after implementation.

**Variations**

1. Have parent volunteers, aides, or peer tutors conduct the listening previews.
2. Divide the class into pairs, with a higher performing student paired with a lower performing student. Have the higher performing students read first to serve as models for the lower performing students.
3. Prerecord (or have aides, parent volunteers, or high-performing students prerecord) selected reading passages for independent listening previews at listening stations in the classroom.
4. To enhance vocabulary and comprehension, have previewers discuss the meaning of important words in the text with the students prior to having the students read.

**Notes**

Including a self-monitoring component by having students graph their WCPM performance increases motivation and enhances the effectiveness of this intervention.

**Sources**

- Rose, T. L. (1984). The effects of two prepractice procedures on oral reading. *Journal of Learning Disabilities, 17*, 544–548.
- Rose, T. L., & Sherry, L. (1984). Relative effects of two previewing procedures on LD adolescents' oral reading performance. *Learning Disability Quarterly, 7*, 39–44.
- Rousseau, J. K., Tam, B. K., & Ramnarain, R. (1993). Increasing the reading proficiency of language-minority students with speech and language impairments. *Education and Treatment of Children, 16*, 254–271.

---

<sup>4</sup>Adapted from *Effective School Interventions: Strategies for Enhancing Academic Achievement and Social Competence* (pp. 168–170) by N. Rathvon, 1999, New York: Guilford Press. Copyright 1999 by Guilford Press. Adapted with permission.

support with many aspects of the intervention process. Although teachers must implement the strategy, studies suggest that treatment integrity and, ultimately, the success of the intervention are related to the degree to which classroom-based assistance is provided to teachers during implementation. For example, IAPs in which a designated case manager works collaboratively with the referring teacher throughout the implementation process are more likely to report measurable gains in student performance than are IAPs lacking such support systems.

### **4.3. Evaluating Academic Interventions**

Systematically evaluating performance change not only provides information that is useful in monitoring and increasing intervention effectiveness but also contributes to teachers' maintenance of interventions by demonstrating that positive change is occurring.

Although researchers have developed several measures for assessing teachers' perceptions of changes in students' academic performance, it is important to assess actual student outcomes and not merely teachers' or parents' perceptions of improvement. Similarly, researchers have often evaluated the effects of academic interventions in terms of task completion rates without regard for accuracy or the absolute level of achievement. Ultimately, the effectiveness of academic interventions should be evaluated in terms of meaningful changes in students' academic achievement relative to grade-level expectations.

#### **4.3.1. Curriculum-Based Assessment**

In recent years, intervention-oriented researchers have developed alternative assessment methodologies to traditional norm-referenced tests with the goal of identifying students in need of supplementary academic services and documenting the effectiveness of school-based interventions. One of these methods, curriculum-based assessment (CBA), refers to a set of procedures

that link assessment directly to instruction and evaluate progress using measures taken from the students' own curricula. Among the many different CBA models, the most fully developed is curriculum-based measurement (CBM), which has become the standard for assessing changes in student performance subsequent to interventions, especially in reading. Developed by Deno, Mirkin, and colleagues at the University of Minnesota Institute for Learning Disabilities, CBM is a generic measurement system that uses brief, fluency-based measures of basic skills in reading, mathematics, spelling, and written expression. CBM is ideally suited to monitoring the progress of students receiving academic interventions because measures are brief (1–3 min), can be administered frequently, and are based on students' own instructional materials. Procedures for conducting CBMs in reading, mathematics, spelling, and written expression can be found in Rathvon's 1999 book.

## 5. ACADEMIC INTERVENTION TARGETS

Academic interventions can be categorized according to targets as follows: (a) interventions designed to enhance academic productivity, including classwork, independent seatwork, and homework; (b) interventions targeting achievement in specific academic subjects; and (c) interventions targeting what DiPerna and Elliott termed academic enablers, that is, non-academic skills, behaviors, and attitudes that contribute to academic competence. Table I describes empirically validated academic interventions from each of the three categories. The categorization is necessarily somewhat arbitrary because all of the interventions include procedures that facilitate productivity and academic enabling behaviors; however, the interventions in the academic enablers category include the largest number of behavioral and attitudinal components.

### 5.1. Interventions Targeting Academic Productivity

#### 5.1.1. Self-Management

Self-management techniques involve teaching students to engage in some form of behavior, such as self-observation or self-recording, in an effort to alter a target behavior. Self-management interventions fall into one of two categories: (a) contingency-based strategies with self-reinforcement for the performance of

specified tasks or (b) cognitively based strategies that use self-instruction to address academic deficits. Self-management interventions are especially appropriate for targeting academic problems because they not only enhance students' sense of responsibility for their own behavior but also increase the likelihood that students will be able to generalize their new competencies to other situations. Many academic interventions, including most of those listed in Table I, include at least one self-management component.

#### 5.1.2. Classwide Peer Tutoring

Increased academic responding is associated with higher levels of on-task behavior and achievement. In classwide peer tutoring, peers supervise academic responding so that every student can engage in direct skill practice during instructional periods, leaving teachers free to supervise the tutoring process. Moreover, because peer tutors are provided with the correct answers for tutoring tasks, the strategy permits immediate error correction. Of the several variations of this strategy, the best known is the Classwide Peer Tutoring (CWPT) program developed by Greenwood and colleagues at the University of Kansas to improve the achievement of entire classrooms of low-socioeconomic status urban students. CWPT has been successful in improving academic skills and productivity in a variety of domains, including oral reading, spelling, and mathematics computation, and with both regular education and special needs students.

### 5.2. Interventions Targeting Academic Achievement

This section discusses some of the best-known and most widely validated interventions in three academic areas: reading, mathematics, and written language.

#### 5.2.1. Interventions to Improve Reading Performance

Reading problems are the most frequent cause of referrals to school psychologists and IAPs. Three sets of skills are required for proficient reading: (a) decoding (i.e., the process leading to word recognition), (b) comprehension (i.e., the ability to derive meaning from text), and (c) fluency (i.e., the ability to read quickly and accurately). Although reading interventions can be categorized according to their primary subskill target, interventions focusing on one subskill

have the potential to improve other competencies due to the interrelated nature of the reading process. The two interventions described in this section primarily target fluency and comprehension.

**5.2.1.1. Listening Previewing** Listening previewing, one of the best-known stand-alone reading interventions, promotes fluency by providing students with an effective reading model prior to the students reading aloud. Combining listening previewing with discussion of key words in the selection to be read is associated with enhanced outcomes, especially in terms of comprehension. Listening previewing ranks high in both efficacy and usability and has been implemented successfully with regular education students, bilingual students, and students with learning and behavior disorders. Additional information about listening previewing is provided in [Table III](#).

**5.2.1.2. Collaborative Strategic Reading** Collaborative strategic reading (CSR) combines instruction in comprehension strategies and study skills with collaborative peer practice. Students learn four strategies through direct instruction and teacher modeling: (a) Preview (i.e., previewing and predicting), (b) Click and Clunk (i.e., monitoring for understanding and vocabulary knowledge), (c) Get the Gist (i.e., understanding the main idea), and (d) Wrap-Up (i.e., self-questioning for understanding). After students have mastered the strategies, they implement them within cooperative groups. CSR has been successful in improving reading proficiency in regular education, multilevel, inclusive, and special education settings. Originally designed for use with expository text in content area textbooks, it can also be applied to narrative material.

### **5.2.2. Interventions to Improve Mathematics Performance**

The 2003 NAEP report on mathematics revealed serious deficiencies in math achievement in the general student population. Although the percentage of fourth graders performing at or above the proficient level (29%) was higher in 2003 than in all previous assessment years since 1990, sizable numbers of students failed to reach even the basic level of proficiency (23%). The situation was even more dismal for eighth graders, with only 23% performing well enough to be classified as proficient. Although students with difficulties in learning mathematics constitute a very heterogeneous group, they generally exhibit deficits in one or more of three areas:

(a) computational skills, including the basic operations of addition, subtraction, multiplication, and division; (b) computational fluency, that is, speed and automaticity with math facts; and (c) mathematics applications, including areas such as money, measurement, time, and word problems. Targets of mathematics interventions can be characterized as (a) foundational arithmetic skills (e.g., number knowledge, basic understanding of mathematical operations), (b) acquisition and automatization of basic computational skills, and (c) problem-solving skills. Recent meta-analyses of mathematics interventions with low-achieving students and/or students with disabilities have reported the greatest efficacy for interventions targeting basic skills and the least efficacy for interventions focusing on higher order mathematics skills. Promising intervention components include frequent feedback to teachers and parents regarding student performance, explicit instruction in math concepts and procedures, and peer-assisted learning. The two interventions described in this section use highly structured, multicomponent approaches to enhance basic skills acquisition and fluency.

**5.2.2.1. Cover–Copy–Compare** The cover–copy–compare (CCC) strategy, originally developed by Skinner and colleagues at Mississippi State University, is a self-management intervention that can be used to enhance accuracy and fluency in a variety of academic subjects. Students look at an academic stimulus (e.g., a multiplication problem for CCC mathematics), cover it, copy it, and evaluate their response by comparing it to the original stimulus. CCC combines several empirically based intervention components, including self-instruction, increased opportunities to respond to academic material, and immediate corrective feedback.

**5.2.2.2. Reciprocal Peer Tutoring** Reciprocal peer tutoring (RPT) in mathematics, developed by Fantuzzo and associates at the University of Pennsylvania, combines self-management techniques and group contingencies within a peer tutoring format. Although both CWPT and RPT involve peer-mediated instruction, RPT includes self-management and subgroup contingencies, with teams of students selecting and working to obtain their own rewards. RPT has been demonstrated to enhance not only math performance but also students' perceptions of their own scholastic competence. Including a home-based reinforcement component enhances positive outcomes.

### **5.2.3. Interventions to Improve Written Language Performance**

Writing is a crucial skill for school success because it is a fundamental way in which to communicate ideas and demonstrate knowledge in the content areas. Unfortunately, writing problems not only are characteristic of most students with learning disabilities but also are prevalent in the general student population. According to the 2002 NAEP report on writing, only 26% of 4th graders scored at the proficient level, with the percentage dropping to 22% for 12th graders. The pervasiveness of writing problems suggests that poor writing achievement is related less to internal student disabilities than to inadequate writing instruction. During the past decade or so, research on the cognitive processes underlying writing has led to a shift in writing instruction from an emphasis on product (e.g., grammar, mechanics, content) to an emphasis on the processes used to generate written productions (e.g., brainstorming, writing multiple drafts, developing a sense of audience, incorporating feedback from others). As a result, writing interventions increasingly focus on student performance of various aspects of the writing process, including planning, sentence generation, and revising. The interventions in this section target several writing process components, including fluency and compositional elements such as planning and editing.

**5.2.3.1. Self-Regulated Strategy Development in Writing** Self-regulated strategy development (SRSD) is an instructional approach that combines explicit teaching and modeling of compositional strategies with a set of self-regulation procedures. In a series of studies, Graham, Harris, and colleagues at the University of Maryland have documented that SRSD improves the quantity and quality of narrative and expository writing for students with writing disabilities as well as for normally developing writers. In addition to enhancing writing performance, SRSD is associated with improved feelings of self-efficacy and more positive attitudes toward writing for students. Most applications of SRSD include paired and/or small-group learning activities at specific steps in the training process to provide additional opportunities for feedback and collaboration without direct teacher supervision.

**5.2.3.2. Self-Recording of Writing Productivity** This intervention uses CBM-type methodology in the context of self-monitoring procedures to increase writing fluency. Self-recording word counts during free writing periods provides students with opportunities for

positively evaluating their own writing fluency and yields useful progress monitoring data by showing teachers when changes in instruction are followed by increases in writing performance rates. Setting writing production goals and self-recording word counts produce increases in the number of words written and improvements in free writing expressiveness without a deterioration in writing mechanics, compared with untimed writing periods. Some variations include individual and group reinforcements for achieving preset goals.

### **5.3. Interventions to Enhance Academic Enablers**

According to DiPerna and Elliott, academic success requires more than skill in performing assigned tasks. That is, although classroom instruction focuses on the acquisition of concepts, knowledge, and skills in academic subjects, students must become active participants in their educational experiences to benefit from that instruction. The two interventions in this category have been widely used as stand-alone strategies or in combination with other intervention components to facilitate achievement and productivity across a range of academic skill domains.

#### **5.3.1. Public Posting**

Public posting involves displaying some kind of classroom record (e.g., a chart) that documents student achievement or productivity. Whereas traditional public posting strategies often record incidents of negative behavior (e.g., the names of disruptive or unproductive students), public posting as an academic intervention is a positive strategy that displays student progress in achieving specified academic goals. Originally designed to encourage improvement in individual student performance, small-group and classwide versions that target group achievement and capitalize on positive peer influence have also been developed. Adding individual or group contingencies can enhance outcomes but does not appear to be critical to intervention effectiveness.

#### **5.3.2. School–Home Notes**

School–home notes encourage parental involvement in children’s classroom performance, permit a broader range of reinforcers than are generally available to teachers, and have demonstrated efficacy for both

academic problems and behavior problems. School-home communications can be arrayed along a continuum of parental involvement, ranging from notes that merely provide information to notes that ask parents to deliver predetermined consequences contingent on the reported student performance. Although strategies that include home consequences can have powerful effects on student performance, establishing and maintaining an effective school-home communication system can be difficult for even a single student, much less for groups or entire classrooms of unproductive students. Not surprisingly, the majority of published school-home note interventions have targeted individual students or small groups of students, usually in special education settings.

## 6. SUMMARY AND FUTURE DIRECTIONS

Given the growing diversity and needs of the student population, interest in the development and empirical validation of academic interventions is likely to increase. Although the knowledge base of effective school-based interventions has increased dramatically during the past 15 years, determining which intervention components and which intervention parameters (e.g., intensity, duration, delivery system) are maximally effective for which types of students will be a continuing challenge for researchers and practitioners alike. Additional studies are especially needed to identify strategies that are high in both efficacy and usability for second-language learners and secondary school students with special needs.

### See Also the Following Articles

Academic Failure, Prevention of ■ Advertising and Culture ■ Behavioral Assessment in Schools ■ Educational Achievement and Culture ■ Educational and Child Assessment ■ Effective Classroom Instruction ■ Learning

■ Learning Styles and Approaches to Studying ■ Mathematics, Teaching of ■ Reading Interventions ■ Reading, Teaching of ■ Teaching Effectiveness ■ Transfer of Learning ■ Writing, Teaching of

## Further Reading

- Daly, E. J., & McCurdy, M. (Eds.). (2002). Developments in academic assessment and intervention [Special issue]. *School Psychology Review*, 31(4).
- DiPerna, J. C., & Elliott, S. N. (Eds.). (2002). Promoting academic enablers to improve student performance: Considerations for research and practice [Miniseries]. *School Psychology Review*, 31(3).
- Elliott, S. N., Busse, R. T., & Shapiro, E. S. (1999). Intervention techniques for academic performance problems. In C. R. Reynolds, & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed., pp. 664–685). New York: John Wiley.
- Gutkin, T. B. (Ed.). (2002). Evidence-based interventions in school psychology: The state of the art and future directions [special issue]. *School Psychology Quarterly*, 17(4).
- Rathvon, N. (1996). *The unmotivated child: Strategies for helping your underachiever become a successful student*. New York: Simon and Schuster.
- Rathvon, N. (1999). *Effective school interventions: Strategies for enhancing academic achievement and social competence*. New York: Guilford Press.
- Shinn, M. R., Walker, H. M., & Stoner, G. (Eds.). (2002). *Interventions for academic and behavior problems II: Preventive and remedial approaches*. Bethesda, MD: National Association of School Psychologists.
- Swanson, H. L. (2000). What instruction works for students with learning disabilities? Summarizing the results from a meta-analysis of intervention studies. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 1–30). Mahwah, NJ: Lawrence Erlbaum.
- Upah, K. R. F., & Tilly, W. D. (2002). Best practices in designing, implementing, and evaluating quality interventions. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 483–502). Bethesda, MD: National Association of School Psychologists.



# Accidents in Transportation

Sidney Dekker

Linköping Institute of Technology, Linköping, Sweden

1. Why Is It Important to Understand Accidents?
  2. Applied Psychology and Accidents
  3. Why Do Accidents Happen?
  4. Regulation of Transportation Systems and Incident Reporting
  5. Interventions, New Technology, and Future Developments
- Further Reading

Transportation accidents carry great personal, organizational, societal, and economic costs. Applied psychology can derive research inspiration from, and can contribute much to the understanding and eventual prevention of, accidents in all kinds of transportation modes. The applied psychological questions surrounding accidents can vary widely, from issues of perception in an individual driver, to trade-offs made in complex organizations under pressures of scarcity and competition, to biases and heuristics in our understanding of culpability, cause, and control. Further progress on safety in transportation systems hinges in part on contributions from applied psychology.

## GLOSSARY

**accident** Generally unforeseen occurrence with damaging consequences for life, limb, or property.

**automation** Replacement of human work by machines or computers in many transportation modes; it often creates new human work in form of new knowledge and skill demands and has been linked to the emergence of new kinds of accidents.

**chain of events** Model of accident causation that captures the multiple subsequent failures necessary to cause an accident (i.e., domino effect); it has recently been challenged by systemic and control models.

**regulation** Activity of defining rules to govern (transportation) practice and the enforcement of those rules; this often requires large organizations that are nearly always government sponsored and run.

**systemic and control models** Models that see accidents as a result of loss of control or as a gradual erosion of safety constraints on system design and operation.

## 1. WHY IS IT IMPORTANT TO UNDERSTAND ACCIDENTS?

Transportation accidents carry tremendous social, environmental, and economic costs. Traffic accidents are among the leading causes of death for people under 34 years of age. In 2002 alone, 42,815 people died in highway accidents in the United States; that equates to more than 117 fatalities per day or 1.51 fatalities per 100 million vehicle miles traveled. The number of injuries hovers around 3 million per year. The economic costs of these accidents are enormous—more than \$230 billion per year for the United States alone. Survivors of transportation accidents often require treatment and rehabilitation and also suffer psychological consequences that may carry over into work and social settings.

The nature and frequency of accidents differ widely among the various transportation modes. For example, the year 1999 witnessed 51 major accidents in commercial aviation and 2768 train-related accidents that were responsible for 932 casualties. At the same time, safety of the various transportation modes varies with geographical location. For example, in terms of casualties, passenger shipping may be less safe than driving in parts of Southeast Asia, and flying may be more dangerous in certain parts of Africa. Indeed, it would seem that the phrase “richer is safer” is true except when considering road traffic accidents in heavily populated, affluent parts of world. All the same, a transportation accident is one of the most common risks to which any member of society is exposed.

## **2. APPLIED PSYCHOLOGY AND ACCIDENTS**

Transportation accidents are interesting from an applied psychology perspective because they are, most basically, about human behavior in applied settings. In fact, accidents are typically about human cognition as much as they are about technical problems and issues. The applied psychological questions surrounding accidents can vary widely, from issues of sensation and perception in an individual driver to shortcuts made by organizational members under pressures of scarcity and competition. Applied psychology as a field of scientific inquiry can use its insights for, and derive much empirical inspiration from, transportation accidents. Indeed, applied psychology in the broadest, most inclusive sense is relevant to transportation accidents.

In studying whether a car driver could have seen a pedestrian, applied psychology may be called on to assess the effect of atmospheric modulation (e.g., rain, mist, dust) on stimulus perceptibility (e.g., the pedestrian). Models exist to calculate so-called atmospheric modulation transfer that can help applied psychologists to reconstruct the existing visibility during a particular occurrence. Additional questions may arise when substance abuse is suspected on the part of the driver; in such a case, applied psychology relies more on models of psychopathology. Much research leverage exists in gauging the (sub)cultural acceptability of consuming alcohol before engaging in transportation activities. Other questions with relevance to applied psychology arise when notable mismatches occur in people’s risk perceptions. For example, people might

not want to use the railroads due to one highly publicized recent fatal accident. If they prefer to drive their cars, they actually expose themselves to the risk of an accident (perhaps even a fatal accident) several orders of magnitude larger than that which would exist while riding the train. Such individual risk assessments are deeply confounded, mixing personal histories, notions of control and destiny, and various kinds of biases and rationalizations (e.g., the availability heuristic) into a rich trove for applied psychology.

Perhaps even more complex are questions surrounding the trade-offs made by professionals who work in transportation systems, leading to issues such as “safety cultures” in transportation systems. Does the commander of the aircraft continue with the approach in bad weather or not? Does the master of the ship sail in the storm or not? Such trade-offs are more than the risk assessments of individual decision makers; they must also be understood as expressions of the preferences and priorities of entire sociotechnical systems that operate in environments of scarcity (e.g., only so many passengers, only so much money to be had) and competition (e.g., somebody else is always ready to take over one’s routes). In efforts to investigate these kinds of trade-offs and other applied problems, psychology has seen a methodological shift accelerate over the past decade or so, with an increasing emphasis on fieldwork and the study of applied settings. It is believed that to understand decision making in actual (transportation) work, where real decisions have real outcomes for real people, psychological researchers must get out of the laboratory and investigate applied settings directly. Issues of confounding factors are dealt with through intensive analysis and interpretation of research results, leading to findings (e.g., on naturalistic decision making) that are both internally valid and exportable to other applied settings.

There is also a role for applied psychology in understanding issues of culpability and control. To what extent, and why, do we judge participants in transportation (e.g., drivers, pilots) to be culpable for the accidents that they “cause”? Such questions are deeply complex and touch on much of what people hope and believe about the world in which they live. The questions are interconnected with yet another set of biases, including the “hindsight bias” that describes how knowledge of outcome profoundly alters one’s perception of the behavior and intentions that led up to that outcome. The hindsight bias also allows one to convert a complex tangled history leading up to an accident into a simple series of binary decisions, where

participants had clear choices to do the right thing or the wrong thing. These questions also inevitably coincide with assumptions about causation, whether justified or not. These tend to be quite problematic and affect applied psychological reasoning around accidents and human intention and behavior.

### 3. WHY DO ACCIDENTS HAPPEN?

One of the more compelling questions for those involved in transportation is why accidents happen. Indeed, a pressing issue after an accident is often to resolve the question, “What was the cause?” There is, of course, no one or unequivocal answer, and not just because the various modes of transportation may differ widely in their respective etiologies of failure. Models of accident causation develop continually, reflecting not only new insights or access to accident data but also the general scientific spirit of the times. In transportation, the so-called “chain of events” model is popular. In this model, one failure somewhere in the system can be seen to lead to (or trigger) the next, and so on, until this cascade of individually insignificant faults pushes the entire system over the edge of breakdown. For example, a piece of debris on the runway, left there by a preceding aircraft, manages to puncture the tire of a subsequent aircraft. Tire fragments slam into the wing, puncturing the fuel tank and triggering a fire that, not much later, brings down the entire aircraft. The chain of events model is also credited with distributing or refocusing the search for accident causes away from frontline operators (i.e., away from “operator error”), instead identifying higher up supervisory and organizational shortcomings that may have contributed.

However, recent insight into the sociotechnical nature of large transportation accidents, such as disasters involving the Space Shuttles *Challenger* (in 1986) and *Columbia* (in 2003), has exposed the limits of the chain of events model. Most critically, the model presupposes (or even requires) failures so as to cause a failure. This contradicts characterizations of organizational and operational practice preceding accidents as normal everyday routine. Accidents can still happen even if everybody follows the rules and there are no “failures” or “shortcomings” as seen from the perspective of those running and regulating the transportation system. As Perrow suggested in 1984, accidents in these systems may be quite “normal” given their structural properties of interactive complexity and tight coupling. Nothing extraordinary (i.e., no obvious breakdowns or failures) as seen

from inside the system is necessary to produce an accident. For example, aviation is a tightly coupled system, as are railroads. This means that there is little slack to recover if things do start to go wrong. The workings of a railroad system, however, are generally more linear and transparent than those of aviation, with the former allowing for better insight into how to manage the system.

These insights have put pressure on the old label “human error” as an explanation for accidents. Indeed, human error today can no longer be legitimately seen as the cause of accidents; rather, it is seen as an effect, as a symptom, or as a sign of trouble deeper inside the system. Human error is no longer an explanation; instead, it demands one. Human error, to the extent that it exists as a separable category of human performance (which many researchers doubt or deny), is systematically connected to features of people’s tools and tasks. This is consistent with the ecological commitment in much of applied psychology, where an analysis of agent–environment mutuality, or person–context couplings, is critical to producing useful insights into the success and failure of transportation (and other) systems.

The latest accident models deemphasize “cause” altogether, noting that it is a deeply Newtonian concept that might not carry over well into an understanding of why complex transportation systems fail, with the *Columbia* accident being a case in point. These models see transportation accidents rather as control problems; as the gradual erosion and eventual loss of control over a safety-critical process, where safety constraints on design or operation are violated. Such models can be reconciled better with increasing evidence that large transportation accidents are nearly invariably preceded by some kind of slow but gradual drift (e.g., away from procedures, away from design specifications) that is hard to notice or characterize as deviant when seen from the inside, that is, from the perspective of participants themselves. The consistent, if slight, speeding that most male drivers exhibit (e.g., nearly always 5 miles per hour over the speed limit) is another example of this. The deviance is normalized; the nonroutine becomes routine, and there is consistent noncompliance. Production pressures (e.g., pressure to be on time, pressure to maximize capacity use) form the major engine behind such drift toward failure in nearly all transportation systems because virtually no transportation system is immune to the pressures of competition and resource scarcity. Recent research debates are not necessarily about the existence or reality of such pressures; rather, they discuss to what extent these pressures consciously affect people’s

trade-offs and to what extent people deliberately gamble (and lose if they have an accident). Much research points to the prerational insidious nature of production pressures on people's trade-offs rather than to conscious immoral or risky calculation, although male drivers under 25 years of age might be an exception.

Acknowledging that such sociotechnical complexity underlies most transportation accidents means that safety and risk are really social constructs rather than objective engineering measurements. Risk is constructed at the intersection of social forces and technical knowledge (or the lack thereof). Slogans such as "safety comes first" are mere posturing. They are disconnected from reality where risk is negotiated as subjective social activity and where it will be different in different cultures, different among organizations within a single culture, and even different among subcultures within one organization.

#### **4. REGULATION OF TRANSPORTATION SYSTEMS AND INCIDENT REPORTING**

Most, if not all, transportation systems are regulated to some extent. Regulation means that rules are created to govern practice and that the rules are enforced to generate compliance. This often requires large organizations—nearly always public or government agencies—tasked with issuing rules, overseeing practice, and certifying and checking systems, operators, and organizations. The extent of regulation of various transportation modes varies greatly from country to country. In some countries, regulators are criticized for having too close a relationship with the transportation industries they are supposed to regulate and for having a dual and supposedly incompatible mandate (i.e., promoting the use and growth of the transportation system as well as overseeing and monitoring its compliance).

The success of safety regulation depends on the safety level that the transportation system has already achieved. In relatively unsafe transportation modes (e.g., private flying), regulation can pay great and rather immediate dividends. It can help to standardize practice; it can issue, broadcast, and enforce rules and remind operators of them; and it can help to build a corpus of cases on routes to accidents that can be shared in the community. There is often leverage in changing designs (e.g., operator interfaces) to make them more error

resistant and error tolerant. In safer systems (e.g., charter flights), such error-resistant designs have typically evolved further. Accidents are often preceded by so-called "dress rehearsals"—sequences of events similar to real accidents but without fatal outcomes. Learning from those dress rehearsals is encouraged, particularly through incident-reporting systems. However, in ultra-safe transportation systems (e.g., European railroads), overregulation becomes a problem. More rules and more monitoring are no longer accompanied by safety gains; they serve only to increase system complexity and potentially decrease transparency as well as rule compliance. Incident-reporting systems in ultra-safe transportation systems may be of limited value. The typical accident there emerges from routine, everyday banal factors that combine and align in ways that are hard to foresee. Incident reports would neither notice the relevance of those factors nor be able to subsequently project their interplay. This, in fact, is a problem allied with all incident-reporting systems. Reporting is not the same as analysis, which in turn is not the same as learning from potential failure. Just having an incident reporting system in place guarantees little in terms of progress on safety.

#### **5. INTERVENTIONS, NEW TECHNOLOGY, AND FUTURE DEVELOPMENTS**

Prevention and improvement strategies vary considerably from one transportation mode to another because many of the factors that affect accident likelihood are mode specific. In general, prevention strategies seek either to minimize the likelihood of a particular kind of harmful event (e.g., a crash) or to minimize its impact (in terms of property damage, injury, or environmental pollution). New technology often plays a dominant role here. Automatic seat belt systems, side impact protection systems, collision detection and avoidance systems, anti-lock breaking systems, double-hull vessels, automated cocoons to keep an aircraft within its prescribed envelope—all of these systems intervene, prevent, and protect in one sense or another. Even though enormous progress is made through such interventions, technology is sometimes a mixed blessing. Airbags may protect mostly drivers who do not wear seat belts (forcing a large majority of other people, who do wear seat belts, to pay a premium on their cars), and automation in virtually all transportation

modes has been associated with the emergence of new human-machine coordination problems (e.g., mode errors, “automation surprises,” people getting lost in display page architectures). Indeed, automation has been associated with the emergence of new types of accidents in virtually all transportation modes, where it has introduced new capabilities and new complexities. One typical signature of accidents in automated transportation systems is the “going sour” scenario, where a small trigger event (i.e., an unusual occurrence or a small failure in a system somewhere), itself innocuous, leads to a series of misassessments and miscommunications between humans and computers. The system is managed into greater hazard due to coordination breakdowns among the various players (both humans and machines). The end result is often an automation surprise, where humans discover that what they thought they had instructed the automation to do was not quite what the automation was doing. More research, much of it in applied psychology, will be necessary to use new technology and other

interventions to their full potential in helping to reduce transportation accidents.

### See Also the Following Articles

Aviation ■ Driving Safety ■ Models for Transportation  
■ Traffic Safety Assessment ■ Transportation Systems, Overview

### Further Reading

- Amalberti, R. (2002). The paradoxes of almost totally safe transportation systems. *Safety Science*, 37, 109–126.
- Billings, C. E. (1996). *Aviation automation: The search for a human centered approach*. Mahwah, NJ: Lawrence Erlbaum.
- Dekker, S. W. A. (2002). *The field guide to human error investigations*. Bedford, UK: Cranfield University Press.
- Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. New York: Basic Books.
- Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press.



# Acculturation

J. W. Berry

Queen's University, Kingston, Ontario, Canada

1. Acculturation Concept
  2. Acculturation Contexts
  3. Acculturation Strategies
  4. Acculturative Stress
  5. Adaptation
  6. Applications
- Further Reading

## GLOSSARY

**acculturation** A process of cultural and psychological change resulting from contact between cultural groups and their individual members.

**acculturation strategies** Variations in the way acculturating groups and individuals attempt to manage the process of acculturation (see Assimilation, Integration, Marginalization, Separation).

**acculturative stress** A negative psychological reaction to the experiences of acculturation, often characterized by anxiety, depression and a variety of psychosomatic problems.

**adaptation** The long-term outcome of acculturation, including psychological adaptation (feelings of personal well-being and self-esteem) (and sociocultural adaptation (competence in dealing with life in the new society).

**assimilation** The acculturation strategy in which people do not wish to maintain their heritage culture, and seek to participate in the larger society.

**cultural distance** The degree of cultural dissimilarity between two groups, measured by ethnographic indicators, or by an individual's perception of such difference.

**cultural identity** How individuals think and feel about themselves in relation to their two cultural groups in contact.

**integration** The acculturation strategy in which people wish to maintain their heritage culture, and seek to participate in the larger society.

**larger society** An evolving social and political framework representing a civic consensus about how to live together in a culturally plural society.

**marginalization** The acculturation strategy in which people do not maintain their heritage culture, and also do not participate in the larger society.

**multicultural ideology** A positive orientation of individuals to cultural diversity in plural societies involving the acceptance of ethnocultural groups and their participation in the larger society.

**plural society** A society in which a number of ethnocultural groups live together within a shared civic political and economic framework.

**psychological acculturation** A process in which psychological features of a person change as a result of their contact with other cultural groups.

**separation** The acculturation strategy in which people maintain their heritage culture, and seek to avoid participation in the larger society.

Acculturation is the process of cultural and psychological change that takes place as a result of contact between cultural groups and their individual members. Such contact and change occur during colonization, military invasion, migration, and sojourning (e.g., tourism, international study, overseas posting). It continues after initial contact in culturally plural societies, where ethnocultural communities maintain features of their heritage cultures. Adaptation to living in culture contact

settings takes place over time. Occasionally it is stressful, but often it results in some form of accommodation. Acculturation and adaptation are now reasonably well understood, permitting the development of policies and programs to promote successful outcomes for all parties.

## 1. ACCULTURATION CONCEPT

The initial interest in acculturation grew out of a concern for the effects of European domination of colonial and indigenous peoples. Later, it focused on how immigrants (both voluntary and involuntary) changed following their entry and settlement into receiving societies. More recently, much of the work has been involved with how ethnocultural groups relate to each other, and to change, as a result of their attempts to live together in culturally plural societies. Nowadays, all three foci are important as globalization results in ever larger trading and political relations. Indigenous national populations experience neocolonization; new waves of immigrants, sojourners, and refugees flow from these economic and political changes; and large ethnocultural populations become established in most countries.

Early views about the nature of acculturation are a useful foundation for contemporary discussion. Two formulations in particular have been widely quoted. The first, from Redfield and colleagues in a 1936 article, is as follows:

Acculturation comprehends those phenomena which result when groups of individuals having different cultures come into continuous first-hand contact, with subsequent changes in the original culture patterns of either or both groups. . . . Under this definition, acculturation is to be distinguished from culture change, of which it is but one aspect, and assimilation, which is at times a phase of acculturation.

In another formulation, the Social Science Research Council in 1954 defined acculturation as

culture change that is initiated by the conjunction of two or more autonomous cultural systems. Acculturative change may be the consequence of direct cultural transmission; it may be derived from non-cultural causes, such as ecological or demographic modification induced by an impinging culture; it may be delayed, as with internal adjustments following upon the acceptance of alien traits or patterns; or it may be a reactive adaptation of traditional modes of life.

In the first formulation, acculturation is seen as one aspect of the broader concept of culture change (that

which results from intercultural contact), is considered to generate change in “either or both groups,” and is distinguished from assimilation (which may be “at times a phase”). These are important distinctions for psychological work and are pursued later in this article. In the second definition, a few extra features are added, including change that is indirect (not cultural but rather “ecological”), is delayed (internal adjustments, presumably of both a cultural and a psychological character, take time), and can be “reactive” (i.e., rejecting the cultural influence and changing toward a more “traditional” way of life rather than inevitably toward greater similarity with the dominant culture).

In 1967, Graves introduced the concept of psychological acculturation, which refers to changes in an individual who is a participant in a culture contact situation, being influenced both by the external culture and by the changing culture of which the individual is a member. There are two reasons for keeping these two levels distinct. The first is that in cross-cultural psychology, we view individual human behavior as interacting with the cultural context within which it occurs; hence, separate conceptions and measurements are required at the two levels. The second is that not every individual enters into, and participates in, a culture in the same way, nor does every individual change in the same way; there are vast individual differences in psychological acculturation, even among individuals who live in the same acculturative arena.

A framework that outlines and links cultural and psychological acculturation and identifies the two (or more) groups in contact is presented in Fig. 1. This framework serves as a map of those phenomena that the author believes need to be conceptualized and measured during acculturation research. At the cultural level (on the left of the figure), we need to understand key features of the two original cultural groups (A and B) prior to their major contact, the nature of their contact relationships, and the resulting dynamic cultural changes in both groups, and in the emergent ethnocultural groups, during the process of acculturation. The gathering of this information requires extensive ethnographic, community-level work. These changes can be minor or substantial and can range from being easily accomplished to being a source of major cultural disruption. At the individual level (on the right of the figure), we need to consider the psychological changes that individuals in all groups undergo and their eventual adaptation to their new situations. Identifying these changes requires sampling a population and studying individuals who are variably

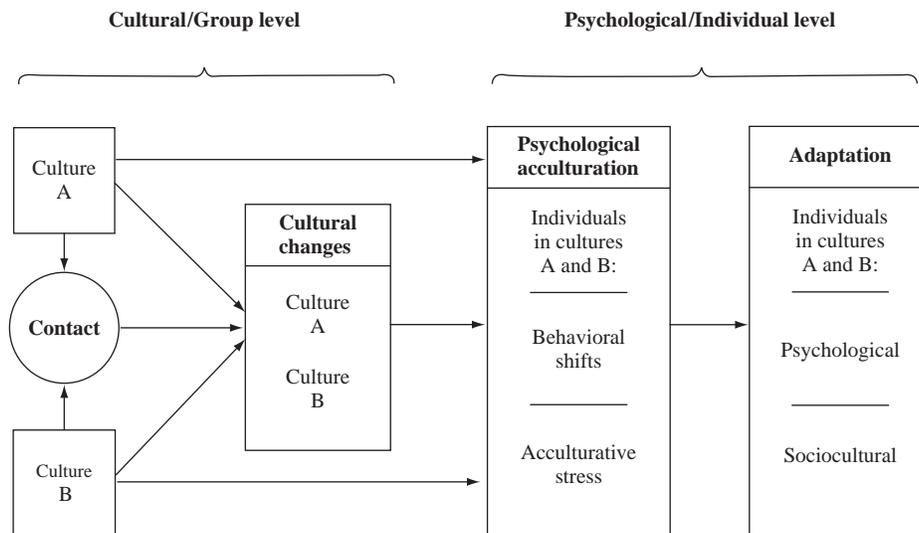


FIGURE 1 A general framework for understanding acculturation.

involved in the process of acculturation. These changes can be a set of rather easily accomplished behavioral shifts (e.g., in ways of speaking, dressing, or eating; in one’s cultural identity), or they can be more problematic, producing acculturative stress as manifested by uncertainty, anxiety, and depression. Adaptations can be primarily internal or psychological (e.g., sense of well-being or self-esteem) or sociocultural, linking the individual to others in the new society as manifested, for example, in competence in the activities of daily intercultural living.

## 2. ACCULTURATION CONTEXTS

As for all cross-cultural psychology, it is imperative that we base our work on acculturation by examining its cultural contexts. We need to understand, in ethnographic terms, both cultures that are in contact if we are to understand the individuals that are in contact.

In Fig. 1, we saw that there are five aspects of cultural contexts: the two original cultures (A and B), the two changing ethnocultural groups (A and B), and the nature of their contact and interactions.

Taking the immigration process as an example, we may refer to the society of origin (A), the society of settlement (B), and their respective changing cultural features following contact (A1 and B1). A complete understanding of acculturation would need to start with a fairly comprehensive examination of the societal contexts. In the society of origin, the cultural characteristics

that accompany individuals into the acculturation process need description, in part to understand (literally) where the person is coming from and in part to establish cultural features for comparison with the society of settlement as a basis for estimating an important factor to be discussed later: cultural distance. The combination of political, economic, and demographic conditions being faced by individuals in their society of origin also needs to be studied as a basis for understanding the degree of voluntariness in the migration motivation of acculturating individuals. Arguments by Richmond in 1993 suggested that migrants can be arranged on a continuum between reactive and proactive, with the former being motivated by factors that are constraining or exclusionary (and generally negative in character), and the latter being motivated by factors that are facilitating or enabling (and generally positive in character). These contrasting factors were also referred to as push/pull factors in the earlier literature on migration motivation.

In the society of settlement, a number of factors have importance. First, there are the general orientations that a society and its citizens have toward immigration and pluralism. Some societies have been built by immigration over the centuries, and this process may be a continuing one, guided by a deliberate immigration policy. The important issue to understand for the process of acculturation is both the historical and attitudinal situations faced by migrants in the society of settlement. Some societies are accepting of cultural pluralism resulting from immigration, taking steps to

support the continuation of cultural diversity as a shared communal resource. This position represents a positive multicultural ideology and corresponds to the integration strategy. Other societies seek to eliminate diversity through policies and programs of assimilation, whereas still others attempt to segregate or marginalize diverse populations in their societies. Murphy argued in 1965 that societies that are supportive of cultural pluralism (i.e., with a positive multicultural ideology) provide a more positive settlement context for two reasons. First, they are less likely to enforce cultural change (assimilation) or exclusion (segregation and marginalization) on immigrants. Second, they are more likely to provide social support both from the institutions of the larger society (e.g., culturally sensitive health care, multicultural curricula in schools) and from the continuing and evolving ethnocultural communities that usually make up pluralistic societies. However, even where pluralism is accepted, there are well-known variations in the relative acceptance of specific cultural, “racial,” and religious groups. Those groups that are less well accepted experience hostility, rejection, and discrimination, one factor that is predictive of poor long-term adaptation.

### 3. ACCULTURATION STRATEGIES

Not all groups and individuals undergo acculturation in the same way; there are large variations in how people seek to engage the process. These variations have been termed acculturation strategies. Which strategies are used depends on a variety of antecedent factors (both cultural and psychological), and there are variable consequences (again both cultural and psychological) of these different strategies. These strategies consist of two (usually related) components: attitudes and behaviors (i.e., the preferences and actual outcomes) that are exhibited in day-to-day intercultural encounters.

The centrality of the concept of acculturation strategies can be illustrated by reference to each of the components included in Fig. 1. At the cultural level, the two groups in contact (whether dominant or nondominant) usually have some notion about what they are attempting to do (e.g., colonial policies, motivations for migration), or of what is being done to them, during the contact. Similarly, the elements of culture that will change depend on the group’s acculturation strategies. At the individual level, both the behavior changes and acculturative stress phenomena are now known to be a function, at least to some extent, of

what people try to do during their acculturation, and the longer term outcomes (both psychological and sociocultural adaptations) often correspond to the strategic goals set by the groups of which they are members.

Four acculturation strategies have been derived from two basic issues facing all acculturating peoples. These issues are based on the distinction between orientations toward one’s own group and those toward other groups. This distinction is rendered as a relative preference for maintaining one’s heritage culture and identity and as a relative preference for having contact with and participating in the larger society along with other ethnocultural groups. This formulation is presented in Fig. 2.

These two issues can be responded to on attitudinal dimensions, represented by bipolar arrows. For purposes of presentation, generally positive or negative orientations to these issues intersect to define four acculturation strategies. These strategies carry different names, depending on which ethnocultural group (the dominant or nondominant one) is being considered. From the point of view of nondominant groups (on the left of Fig. 2), when individuals do not wish to maintain their cultural identity and instead seek daily interaction with other cultures, the assimilation strategy is defined. In contrast, when individuals place a value on holding onto their original culture while wishing to avoid interaction with others, the separation alternative is defined. When individuals have an interest in maintaining their original culture while in daily interactions with other groups, integration is the option. In this case, they maintain some degree of cultural integrity while seeking, as members of an ethnocultural group, to participate as an integral part of the larger social network. Finally, when individuals have little possibility of, or interest in, cultural maintenance (often for reasons of enforced cultural loss) and little interest in having relations with others (often for reasons of exclusion or discrimination), marginalization is defined.

This presentation was based on the assumption that nondominant groups and their individual members have the freedom to choose how they want to acculturate. This, of course, is not always the case. When the dominant group enforces certain forms of acculturation or constrains the choices of nondominant groups or individuals, other terms need to be used.

Integration can be “freely” chosen and successfully pursued by nondominant groups only when the dominant society is open and inclusive in its orientation

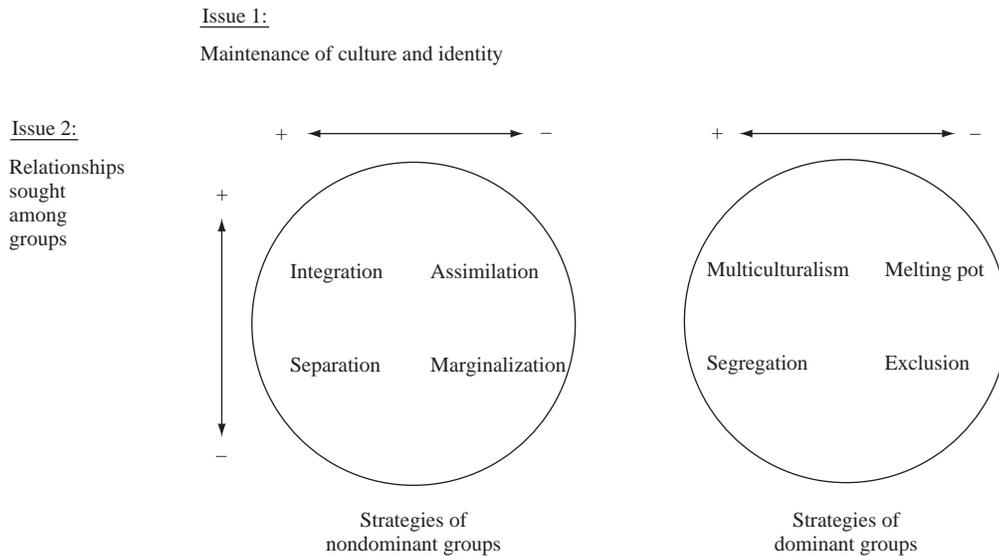


FIGURE 2 Four acculturation strategies based on two issues in ethnocultural groups and the larger society.

toward cultural diversity. Thus, a mutual accommodation is required for integration to be attained and involves the acceptance by both groups of the right of all groups to live as culturally different peoples. This strategy requires nondominant groups to adopt the basic values of the larger society, whereas the dominant group must be prepared to adapt national institutions (e.g., education, health, labor) to better meet the needs of all groups now living together in the plural society.

These two basic issues were initially approached from the point of view of the nondominant ethnocultural groups. However, the original anthropological definition clearly established that both groups in contact would become acculturated. Hence, a third dimension was added: the powerful role played by the dominant group in influencing the way in which mutual acculturation would take place. The addition of this third dimension produces the right side of Fig. 2. Assimilation, when sought by the dominant group, is termed the “melting pot”; however, when it is demanded by the dominant group, it is called the “pressure cooker.” When separation is forced by the dominant group, it is “segregation.” Marginalization, when imposed by the dominant group, is termed “exclusion.” Finally, integration, when diversity is a feature of the society as a whole (including all of the various ethnocultural groups), is called “multiculturalism.” With the use of this framework, comparisons can be made between individuals and their groups and between nondominant peoples and the larger society within which they

are acculturating. The ideologies and policies of the dominant group constitute an important element of ethnic relations research, whereas the preferences of nondominant peoples are a core feature in acculturation research. Inconsistencies and conflicts among these various acculturation preferences are sources of difficulty for acculturating individuals. In general, when acculturation experiences cause problems for acculturating individuals, we observe the phenomenon of acculturative stress.

#### 4. ACCULTURATIVE STRESS

Three ways in which to conceptualize outcomes of acculturation have been proposed in the literature. In the first (behavioral shifts), we observe those changes in an individual’s behavioral repertoire that take place rather easily and are usually nonproblematic. This process encompasses three subprocesses: culture shedding; culture learning; and culture conflict. The first two involve the selective, accidental, or deliberate loss of behaviors and their replacement by behaviors that allow the individual a better “fit” with the society of settlement. Most often, this process has been termed “adjustment” because virtually all of the adaptive changes take place in the acculturating individual, with few changes occurring among members of the larger society. These adjustments are typically made with minimal difficulty, in keeping with the appraisal

of the acculturation experiences as nonproblematic. However, some degree of conflict may occur, and this is usually resolved by the acculturating person yielding to the behavioral norms of the dominant group. In this case, assimilation is the most likely outcome.

When greater levels of conflict are experienced and the experiences are judged to be problematic but controllable and surmountable, the second approach (acculturative stress) is the appropriate conceptualization. In this case, individuals understand that they are facing problems resulting from intercultural contact that cannot be dealt with easily or quickly by simply adjusting or assimilating to them. Drawing on the broader stress and adaptation paradigms, this approach advocates the study of the process of how individuals deal with acculturative problems on first encountering them and over time. In this sense, acculturative stress is a stress reaction in response to life events that are rooted in the experience of acculturation.

A third approach (psychopathology) has had long use in clinical psychology and psychiatry. In this view, acculturation is nearly always seen as problematic; individuals usually require assistance to deal with virtually insurmountable stressors in their lives. However, contemporary evidence shows that most people deal with stressors and reestablish their lives rather well, with health, psychological, and social outcomes that approximate those of individuals in the larger society.

Instead of using the term “culture shock” to encompass these three approaches, the author prefers to use the term “acculturative stress” for two reasons. First, the notion of shock carries only negative connotations, whereas stress can vary from positive (eustress) to negative (dis-stress) in valence. Because acculturation has both positive (e.g., new opportunities) and negative (e.g., discrimination) aspects, the stress conceptualization better matches the range of affect experienced during acculturation. Moreover, shock has no cultural or psychological theory or research context associated with it, whereas stress (as noted previously) has a place in a well-developed theoretical matrix (i.e., stress–coping–adaptation). Second, the phenomena of interest have their life in the intersection of two cultures; they are intercultural, rather than cultural, in their origin. The term “culture” implies that only one culture is involved, whereas the term “acculturation” draws our attention to the fact that two cultures are interacting with, and producing, the phenomena. Hence, for both reasons, the author prefers the notion of acculturative stress to that of culture shock.

Relating these three approaches to acculturation strategies, some consistent empirical findings allow the

following generalizations. For behavioral shifts, the fewest behavioral changes result from the separation strategy, whereas most result from the assimilation strategy. Integration involves the selective adoption of new behaviors from the larger society and retention of valued features of one’s heritage culture. Marginalization is often associated with major heritage culture loss and the appearance of a number of dysfunctional and deviant behaviors (e.g., delinquency, substance abuse, familial abuse). For acculturative stress, there is a clear picture that the pursuit of integration is least stressful (at least where it is accommodated by the larger society), whereas marginalization is the most stressful. In between are the assimilation and separation strategies, with sometimes one and sometimes the other being the less stressful. This pattern of findings holds for various indicators of mental health.

## 5. ADAPTATION

As a result of attempts to cope with these acculturation changes, some long-term adaptations may be achieved. As mentioned earlier, adaptation refers to the relatively stable changes that take place in an individual or group in response to external demands. Moreover, adaptation may or may not improve the fit between individuals and their environments. Thus, it is not a term that necessarily implies that individuals or groups change to become more like their environments (i.e., adjustment by way of assimilation), but it may involve resistance and attempts to change their environments or to move away from them altogether (i.e., by separation). In this use, adaptation is an outcome that may or may not be positive in valence (i.e., meaning only well adapted). This bipolar sense of the concept of adaptation is used in the framework in Fig. 1 where long-term adaptation to acculturation is highly variable, ranging from well adapted to poorly adapted and varying from a situation where individuals can manage their new lives very well to one where they are unable to carry on in the new society.

Adaptation is also multifaceted. The initial distinction between psychological and sociocultural adaptation was proposed and validated by Ward in 1996. Psychological adaptation largely involves one’s psychological and physical well-being, whereas sociocultural adaptation refers to how well an acculturating individual is able to manage daily life in the new cultural context. Although conceptually distinct, the two measures are empirically related to some extent (correlations between them are in the +.40 to +.50 range). However, they are

also empirically distinct in the sense that they usually have different time courses and different experiential predictors. Psychological problems often increase soon after contact, followed by a general (but variable) decrease over time. However, sociocultural adaptation typically has a linear improvement with time. Analyses of the factors affecting adaptation reveal a generally consistent pattern. Good psychological adaptation is predicted by personality variables, life change events, and social support, whereas good sociocultural adaptation is predicted by cultural knowledge, degree of contact, and positive intergroup attitudes.

Research relating adaptation to acculturation strategies allows for some further generalizations. For all three forms of adaptation, those who pursue and accomplish integration appear to be better adapted, whereas those who are marginalized are least well adapted. And again, the assimilation and separation strategies are associated with intermediate adaptation outcomes. Although there are occasional variations on this pattern, it is remarkably consistent and parallels the generalization regarding acculturative stress.

## 6. APPLICATIONS

There is now widespread evidence that most people who have experienced acculturation actually do survive. They are not destroyed or substantially diminished by it; rather, they find opportunities and achieve their goals, sometimes beyond their initial imaginings. The tendency to “pathologize” the acculturation process and outcomes may be due partly to the history of its study in psychiatry and in clinical psychology. Second, researchers often presume to know what acculturating individuals want and impose their own ideologies or personal views rather than informing themselves about culturally rooted individual preferences and differences. One key concept (but certainly not the only one) in understanding this variability, acculturation strategies, has been emphasized in this article.

The generalizations that have been made in this article on the basis of a wide range of empirical findings allow us to propose that public policies and programs that seek to reduce acculturative stress and to improve intercultural relationships should emphasize the integration approach to acculturation. This is equally true of national policies, institutional arrangements, and the goals of ethnocultural groups. It is also true of individuals in the larger society as well as members of nondominant acculturating groups.

In some countries, the integrationist perspective has become legislated in policies of multiculturalism that encourage and support the maintenance of valued features of all cultures while supporting full participation of all ethnocultural groups in the evolving institutions of the larger society. What seems certain is that cultural diversity and the resultant acculturation are here to stay in all countries. Finding a way in which to accommodate each other poses a challenge and an opportunity to social and cross-cultural psychologists everywhere. Diversity is a fact of contemporary life. Whether it is the “spice of life” or the main “irritant” is probably the central question that confronts us all—citizens and social scientists alike.

### See Also the Following Articles

Conformity across Cultures ■ Cross-Cultural Psychology, Overview ■ Cultural Complexity ■ Ideological Orientation and Values

### Further Reading

- Berry, J. W. (1976). *Human ecology and cognitive style: Comparative studies in cultural and psychological adaptation*. New York: Russell Sage/Halsted.
- Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. Padilla (Ed.), *Acculturation: Theory, models, and findings* (pp. 9–25). Boulder, CO: Westview.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology*, 46, 5–68.
- Berry, J. W., Kalin, R., & Taylor, D. (1977). *Multiculturalism and ethnic attitudes in Canada*. Ottawa, Ontario: Supply & Services.
- Berry, J. W., & Kim, U. (1988). Acculturation and mental health. In P. Dasen, J. W. Berry, & N. Sartorius (Eds.), *Health and cross-cultural psychology* (pp. 207–236). Newbury Park, CA: Sage.
- Berry, J. W., Kim, U., Minde, T., & Mok, D. (1987). Comparative studies of acculturative stress. *International Migration Review*, 21, 491–511.
- Berry, J. W., Kim, U., Power, S., Young, M., & Bujaki, M. (1989). Acculturation attitudes in plural societies. *Applied Psychology*, 38, 185–206.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications* (2nd ed.). New York: Cambridge University Press.
- Graves, T. (1967). Psychological acculturation in a tri-ethnic community. *South-Western Journal of Anthropology*, 23, 337–350.
- Murphy, H. B. M. (1965). Migration and the major mental disorders. In M. B. Kantor (Ed.), *Mobility and mental health* (pp. 221–249). Springfield, IL: Thomas.

- Ogbu, J. U. (1992). Understanding cultural diversity and learning. *Educational Researcher*, 21, 5–14.
- Redfield, R., Linton, R., & Herskovits, M. (1936). Memorandum on the study of acculturation. *American Anthropologist*, 38, 149–152.
- Richmond, A. (1993). Reactive migration: Sociological perspectives on refugee movements. *Journal of Refugee Studies*, 6, 7–24.
- Social Science Research Council. (1954). Acculturation: An exploratory formulation. *American Anthropologist*, 56, 973–1002.
- Ward, C. (1996). Acculturation. In D. Landis, & R. Bhagat (Eds.), *Handbook of intercultural training* (2nd ed., pp. 124–147). Newbury Park, CA: Sage.
- Ward, C., Bochner, S., & Funham, A. (2001). *The psychology of culture shock*. London: Routledge.



# Achievement Motivation in Academics

Dale H. Schunk

University of North Carolina, Greensboro, North Carolina, USA

1. Introduction
2. Background
3. Current Perspectives
4. Promoting Achievement Motivation  
Further Reading

Achievement motivation is the desire to excel at effortful activities. Achievement motivation originally was thought of as extending across many areas, but current views conceive of it as more specific to situations. Personal, social, instructional, familial, and cultural factors affect achievement motivation, and parents and educators can help students to improve their achievement motivation.

## GLOSSARY

**achievement motivation** The desire to excel at effortful activities.

**attributions** Perceived causes of outcomes.

**classroom climate** The social, psychological, and emotional characteristics of the classroom.

**expectancy-value theory** Theory predicting that achievement behavior depends on expectations about the outcomes of actions and how much those outcomes are valued.

**goal** What one is trying to accomplish.

**outcome expectations** Beliefs about the expected outcomes of one's actions.

**peer networks** Large groups of peers with whom students associate.

**self-efficacy** Beliefs about one's capabilities for learning or performing given actions.

**Thematic Apperception Test (TAT)** Projective personality measure in which people view ambiguous pictures and make up stories or answer questions.

**value** The importance of an outcome.

## 1. INTRODUCTION

At the end of her ninth-grade geometry class, Mrs. Lollar passed out the "Problem of the Week" for students to take home and solve. Two students in the class, Ashley and Marella, each have a B+ average. Because the problem of the week is worth extra credit, doing well on it could raise each of their averages to an A. That evening at home, Ashley looked at the problem for a while, spent a little time on it, but then quit without solving it. Marella studied the problem and began to work. When her parents called her to dinner, she was reluctant to come because she was working on it. After dinner, she worked on it some more and finally exclaimed, "Aha—I get it!" The next day in class, Marella was one of only three students to have solved the geometric proof. Ashley got some steps in the proof correct, but her solution was far from complete.

Achievement motivation is the desire to excel at effortful activities. In the opening scenario, it seems that Marella was more motivated to achieve in geometry than was Ashley. Marella displayed interest in solving the proof, persisted at it, was excited when she understood what to do, and solved it correctly.

Achievement motivation has a long history in psychology and education, and for good reason. Motivation to achieve is necessary for all but the simplest tasks. Achievement motivation helps students to learn in school, fuels creative activities, and helps individuals and societies to attain goals.

Despite the intuitive importance of achievement motivation, researchers disagree on its critical components. The next two sections examine the background of achievement motivation and some current perspectives.

## 2. BACKGROUND

It is difficult to pinpoint the historical onset of the study of achievement motivation because the human desire to achieve has been of interest for ages. The scientific study of achievement motivation received impetus from work by Murray, who included it as one of several human needs that contribute to personality development. Murray also devised the Thematic Apperception Test (TAT) to study personality processes. The TAT is a projective technique in which people view a series of ambiguous pictures (i.e., inkblots) and make up a story or answer questions for each picture.

McClelland and colleagues adapted the TAT to study achievement motivation. People were shown pictures of individuals in ambiguous situations (e.g., a student at a desk holding a pencil and looking into the air) and were asked questions such as the following: “What is happening?,” “What led up to this situation?,” “What is wanted?,” and “What will happen?” Responses were scored and categorized according to strength of achievement motivation. Unfortunately the TAT has some measurement problems, and TAT achievement motivation scores often did not relate well to other measures of achievement. Over the years, researchers have devised other methods for assessing achievement motivation.

Important early work on achievement motivation was done by Atkinson and colleagues. Atkinson drew on work by Lewin and others on the level of aspiration or the goal that people set in a task. Lewin’s research showed that successes raised and failures lowered the level of aspiration, that people felt more successful when they met the goals they had set for themselves than when

they attained objective standards, and that the level of aspiration reflected individual and group differences.

Atkinson’s expectancy–value theory of achievement motivation states that behavior depends on how much people value a particular goal and their expectancy of attaining that goal as a result of performing in a given fashion. Atkinson postulated that achievement behavior involved a conflict between a motive to approach a task (hope for success) and a motive to avoid it (fear of failure). These motives conflict because any achievement task carries with it the possibilities of success and failure. The achievement motivation that results in any situation depends on people’s expectancies of success and failure and their incentive values of success and failure.

The historical research focused on achievement motivation globally, that is, motivation across many situations. But research and everyday observations show that people rarely are motivated to achieve at high levels in every situation. Rather, students typically have greater achievement motivation in some content areas than in others. In the opening scenario, although Marella’s achievement motivation was higher than Ashley’s in geometry, Ashley may strive to excel more than Marella in history. Given that the achievement motive differs depending on the domain (and even depending on the task within the domain), the validity of general achievement motivation is questionable.

## 3. CURRENT PERSPECTIVES

People who study achievement motivation today believe that it is more situationally specific and is affected by many factors: personal, social, instructional, familial, and cultural (Table I). To understand why students differ in achievement motivation, one must examine the roles of these factors in their lives. Although researchers agree that achievement motivation is complex, they disagree on which factors are the most important.

Personal factors reside within students. Some key personal factors are goals, expectations, values, and attributions. Goals are what one is trying to accomplish. Goals usually are cast in terms of products—Marella and Ashley’s goal was to work the geometry proof—but they also can represent processes. Thus, an academic goal may be to improve one’s skill in comprehending scientific texts. Goals contribute to achievement motivation because people pursuing goals persist and expend effort to succeed. As they work on the tasks, they evaluate their goal processes, and the belief that they are making progress sustains motivation.

TABLE I  
Influences on Achievement Motivation

<i>Factors</i>	<i>Sample influences</i>
Personal	Goals, outcome expectations, self-efficacy, values, attributions
Social	Peers, peer models, peer networks
Instructional	Planning, decision making, instructional grouping, interactions, activities, classroom climate
Familial	Socioeconomic status, capital, guidance
Cultural	Educational emphasis, social orientation

But goals actually are less important than their properties: specificity, proximity, and difficulty. Goals may be specific (e.g., read 10 pages) or general (e.g., read some pages), proximal (e.g., read 10 pages tonight) or distant (e.g., read 10 pages by next week), and difficult (e.g., read 400 pages) or easier (e.g., read 50 pages). From a motivational perspective, goals that are specific, proximal, and moderately difficult produce higher achievement motivation than do goals that are general, distant, and either too difficult or too easy. Thus, motivation is not aided when goals denote general outcomes (because nearly any action will satisfy them), are temporally distant (it is easy to put off until tomorrow what does not need to be done today), and too difficult or too easy (people are not motivated to attempt the impossible and may procrastinate completing easy tasks).

Expectations can be of two types. Outcome expectations refer to the expected outcomes of one's actions (e.g., "If I study hard, I should make a high grade on the test," "No matter how hard I study, I probably will fail the test"). Outcome expectations motivate students because the belief that a certain action will lead to a given outcome should lead students to pursue or avoid that action, assuming that they value that outcome or want to avoid it.

A second type of expectation involves beliefs about one's capabilities to learn or perform given actions. Bandura and other researchers termed this "self-efficacy." Self-efficacious students believe that they can study diligently, whereas those lacking efficacy might believe that they cannot. Marella's self-efficacy in geometry likely was higher than Ashley's. Self-efficacy is important because students who believe that they can learn or perform desired actions are more likely to choose to engage in them, expend effort, and persist.

Self-efficacy and outcome expectations often are related, but they need not be. Students who believe that they are capable typically expect to perform well and receive high grades and other rewards. However,

students may believe that diligent studying will lead to good grades (i.e., positive outcome expectation) but also may doubt their ability to study diligently (i.e., low self-efficacy), in which case they might feel demoralized. Positive outcome expectations and strong self-efficacy for learning and performing capably produce high achievement motivation.

The values and/or importance that students ascribe to learning and achievement are central components of motivation. Those who do not value what they are learning are not motivated to improve or perform well. Research shows that value relates positively to persistence, choice, and performance. Students who value learning choose challenging activities, persist at them, and perform well. Thus, students who value history are apt to study diligently for tests, set goals for their learning, monitor their learning progress, use effective learning strategies, and not be daunted by obstacles. Marella likely valued geometry more than did Ashley.

Attributions are perceived causes of outcomes or the factors that people believe are responsible for their successes and failures. In achievement situations, learners may ask themselves questions such as the following: "Why did I get an A on my biology test?" and "Why can't I learn French?"

Weiner and colleagues formulated an attribution theory of achievement motivation that contends that each attribution can be classified along three dimensions: locus, stability, and controllability. The locus dimension refers to whether the attribution is internal or external to the person. The stability dimension denotes how much the attribution varies over time. The controllability dimension involves the extent that the attribution is under the individual's control.

Weiner found that common attributions in achievement settings are ability, luck, task difficulty, and effort; however, there are many others (e.g., fatigue, illness, personal dislike, time available, attitude). Ability is internal, stable, and uncontrollable (although one's ability

can improve over time); luck is external, unstable, and uncontrollable; task difficulty is external, stable (assuming that the task does not change), and uncontrollable; and effort is internal, unstable (although a general effort factor also seems to exist), and controllable.

Students who attribute academic success to high ability and effort are more likely to expect future success than are those who attribute it to task ease and good luck. Students who attribute failure to low ability are less likely to expect future success than are those who attribute it to low effort. The latter finding is especially critical because a low expectation of future success stifles motivation. From the perspective of achievement motivation, it is far better to stress effort than ability because the former has stronger motivational effects. Of course, as skills develop, students do become more able, so the attribution of success to ability is credible. Teachers often have difficulty in motivating students who believe that they lack the ability to succeed and that no amount of effort will help them to be successful.

Social factors are those that are inherent in interactions with others. Peers are a key social group for students. Peer models are especially influential, especially those who are similar to observers in important ways, because they convey that tasks are valuable.

Peer group goals are highly valued by students. Thus, students may want to be liked and approved, develop social and/or intimate relationships, learn to cooperate, win favors, and be sensitive to the needs of others. Students' perceptions of their capabilities also are affected by peers through social comparisons. Observing similar peers succeeding can raise observers' self-efficacy.

Peer networks are large groups of peers with whom students associate. Students in networks tend to be highly similar and, thus, are key models. Students' motivational engagement in school across the school year is predicted by their network membership at the start of the year. Those in more academically inclined networks demonstrate higher academic motivation than do those in low-motivation networks.

Although students may select their peer networks, parents can play a key role by "launching" their children onto particular trajectories. For example, parents who want their children to be academically oriented are likely to involve them in activities and groups that stress academics. Peers with whom children associate reinforce the emphasis on academics.

Instructional factors include teacher planning and decision making, grouping for instruction, teacher-student interactions (including feedback to students), activities, and classroom climate. Teachers can enhance student

motivation by planning interesting activities that maximize student involvement in lessons. Teachers who plan only lectures are less apt to promote student motivation. Teachers also can promote motivation by basing instructional decisions not only on how well students are learning but also on how much the material appeals to students.

Three types of grouping structures are competitive, cooperative, and individualist. Competitive situations are those in which the goals of individuals are linked negatively such that if one attains his or her goal, the chance that others will attain their goals diminishes. Cooperative structures are those in which the goals of the group members are linked positively such that one can attain his or her goal only if others attain their goals. In individualist situations, there is no link among the goals of individuals such that one's goal attainment has no effect on the goal attainment of others. From a motivational perspective, competitive situations highlight differences among students, and lower achievers may become discouraged if they believe that they have no chance to earn rewards. Cooperative situations are better if all students contribute to the project. If only one or two students do most of the work, there is apt to be resentment. In individualist situations, achievement motivation can be developed and sustained when learners focus on their progress or on how much better they are performing now compared with earlier.

Performance feedback provides information on accuracy of work and may include corrective information (e.g., "The first part is correct, but you need to bring down the next number"). Motivational feedback can provide information on progress and competence (e.g., "You've gotten much better at this"), link student performance with one or more attributions (e.g., "You've been working hard"), and inform students about how well they are applying a strategy and how strategy use is improving their work (e.g., "You got it correct because you followed the proper method"). Feedback motivates students when it informs them that they are making progress and becoming more competent.

Classroom climate refers to the atmosphere of the classroom—its social, psychological, and emotional characteristics. Climate is important for motivation because classroom interactions define the climate. Climate often is referred to in terms such as "warm," "cold," "permissive," "democratic," and "autocratic." Research shows that a democratic environment—one based on mutual respect and collaboration—fosters goal attainment by learners without their becoming frustrated or aggressive.

There are many familial factors, but a key one is socioeconomic status, whose link with students'

academic motivation is well established. Children from lower socioeconomic backgrounds typically display lower motivation and achievement, and are at greater risk for school failure and dropout, compared with children from wealthier families. But socioeconomic status is only a descriptive term and does not explain why effects occur. There also are many people who grew up in impoverished environments but became well educated and successful.

A major contributor is a family's capital or resources. Poor families have less financial resources to support their children's learning outside of school than do wealthier families. Socialization in lower class homes often does not match or prepare students for the middle-class orientation of schools and classrooms. Lower socioeconomic students might not understand the full benefits of schooling or comprehend that getting a good education will increase their chances of college acceptance, good jobs, and financial stability. Because of family financial strain, they also might not be able to resist the short-term benefits of working in favor of the long-term benefits of schooling. Such students might have few, if any, educated role models in their environments. Many enter school without the needed social, cognitive, and emotional prerequisites to learn successfully.

Families continue to influence children's motivation throughout childhood and adolescence by steering children in given directions. Families that provide rich resources in the home and guide their children into activities that stress motivation and achievement are apt to develop children higher in achievement motivation.

Cultural factors also affect motivation. Cross-cultural research shows that there are differences in how much cultures emphasize education and motivation for learning. For example, children from Asian cultures often place greater emphasis on effort as a cause of success than do students in the United States. Some cultures are more socially oriented, whereas others place more emphasis on individual accomplishments. Although it is hard to make generalizations about cultures given that not all members of a culture act alike, it does seem that motivational differences stem in part from influences in cultural background.

#### **4. PROMOTING ACHIEVEMENT MOTIVATION**

How can parents and educators help to improve students' achievement motivation for academics? First,

they can help students to set challenging but attainable goals. Students must believe that the goals are attainable; they will not be motivated to attempt what they see as impossible. Teachers and parents might need to work with students to ensure that goals are realistic; if students are unaware of the demands of an assignment, they might set an unattainable goal. Challenging and attainable goals help to build self-efficacy as students perceive that they are making progress.

Second, it is important for parents and educators to stress the value of learning. Motivation is enhanced when learners understand how they can use what they are learning and how it will help them in the future. People are not motivated to engage in "busy work." Marella and Ashley's teacher would do well to point out applications of geometry in daily life.

Third, parents and educators should build students' perceptions of their competence or self-efficacy. Self-efficacy is increased when learners believe that they are developing skills, making progress toward their goals, and performing better. Self-efficacy develops through actual performance accomplishments, vicarious (modeled) experiences, and social persuasion. Teachers and parents should ensure that students work at tasks on which they can be successful, observe similar peers succeeding, and receive feedback indicating that they are capable (e.g., "I know you can do this").

Fourth, parents should get involved in children's learning. Parents can be involved in school by assisting in classrooms and at home. Parents serve as models for learning, and they can help to establish a productive study environment. They also can be taught tutoring skills. The importance of parents as motivational models cannot be overemphasized.

Fifth, teachers should use peers effectively to build motivation. Teachers who group students for collaborative project work must ensure that each student has responsibility for part of the task so that the bulk of the work is not done by one or two students. In using models for comparisons, it is imperative that students view the models as similar to themselves. Motivation of lower achieving students might not improve when these students are asked to observe higher achieving students perform.

Finally, parents and educators should use feedback to teach and motivate. Feedback typically informs students whether they are correct or incorrect and what to do to perform better. Feedback motivates when it informs students about their progress—how much better they are performing now than they were previously—and when it links their performances to effort, good strategy use, and enhanced ability.

Academic motivation for academics is important not only for schooling but also for the future. Educators want students to continue to be motivated to learn once they leave their classrooms. This is the essence of “lifelong learning” by which citizens move their societies forward.

### See Also the Following Articles

Educational Achievement and Culture ■ Learning ■ Learning Styles and Approaches to Studying ■ Motivation and Culture

### Further Reading

- Ames, C. (1984). Competitive, cooperative, and individualistic goal structures: A cognitive–motivational analysis. In R. Ames, & C. Ames (Eds.), *Research on motivation in education* (Vol. 1, pp. 177–207). New York: Academic Press.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*, 359–372.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399.
- McClelland, D., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The achievement motive*. New York: Appleton–Century–Crofts.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Steinberg, L., Brown, B. B., & Dornbusch, S. M. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do*. New York: Simon & Schuster.
- Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: Sage.



# Achievement Tests

Gregory J. Cizek

University of North Carolina, Chapel Hill, North Carolina, USA

1. Introduction
  2. Definitions and Examples
  3. Types of Achievement Tests
  4. Achievement Test Construction
  5. Evaluating Achievement Tests
- Further Reading

## GLOSSARY

**ability test** A procedure designed to measure an examinee's potential for achievement, sometimes also referred to as an aptitude test.

**achievement test** A procedure designed to measure an examinee's attainment of knowledge or skills.

**battery** A collection of related subtests, the performance on which can be aggregated to form a composite or "total battery" score on the entire test.

**classroom assessments** Achievement tests that are developed, administered, and scored for classroom use.

**content standards** Statements that describe specific knowledge or skills over which examinees are expected to have mastery for a given age/grade and subject area.

**criterion-referenced test (CRT)** An instrument whose primary purpose is to gauge whether an examinee knows or can do specific things.

**high-stakes/low-stakes** Terms referring to the severity of the consequences associated with test performance.

**inference** A reasoned leap, supported by evidence, made whenever observed test scores are interpreted in terms of examinees' underlying standing on a construct or actual level of knowledge or skill.

**norm-referenced test (NRT)** An instrument whose primary purpose is to describe the relative standing of examinees at a particular age or grade level.

**performance standards** Prescribed levels of performance representing differential degrees of knowledge or skill with respect to a set of content standards.

**reliability** The dependability of test scores; because an assignment or test consists only of a sample of questions or tasks, and because both the students who respond and those who score the responses are susceptible to various unpredictabilities in their performance (called random errors), no score can be considered to be a perfectly dependable representation of a student's performance.

**standardized** A term used to describe any instrument that is developed, administered, and scored under controlled conditions.

**standards-referenced test (SRT)** An instrument whose primary purpose is to describe examinees' performance with respect to a set of content standards.

**validity** The degree to which the conclusions yielded by any sample of behavior (e.g., test, quiz, interview) are meaningful, accurate, and useful; in the context of achievement testing, validity is the degree to which students' test scores yield inferences about the students' level of knowledge, skill, or ability that are "correct" or "on target."

Achievement tests are used in diverse contexts to measure the degree to which examinees can demonstrate acquisition of knowledge or skills deemed to be important. The contexts range from teacher-made testing in elementary and secondary school settings to

high-stakes testing for college admission, licensure to practice a profession, or certification. The design of achievement tests varies depending on whether the inference intended to be drawn regarding examinees' performance is the absolute or relative level of mastery of specific knowledge and skills.

## 1. INTRODUCTION

Achievement testing is a general term used to describe any measurement process or instrument whose purpose is to estimate an examinee's degree of attainment of specified knowledge or skills. Beyond that central purpose, achievement tests differ according to their specific intended inference. Common inferences include either absolute level of performance on the specified content or relative standing vis-à-vis other examinees on the same content. Achievement tests may be group or individually administered. They may consist of differing formats, including multiple-choice items, essays, performance tasks, and portfolios.

Achievement tests are administered in diverse contexts. For example, they are used when the school-related skills of preschool pupils are measured to assess their readiness for kindergarten. During the K–12 school years, students typically take a variety of achievement tests, ranging from teacher-made informal assessments, to commercially prepared achievement batteries, to state-mandated high school graduation tests. Following formal schooling, achievement tests are administered to assess whether examinees have an acceptable level of knowledge or skill for safe and competent practice in a regulated profession for which licensure is required. In other situations, professional organizations establish certification procedures, often including an achievement test, to determine examinees' eligibility to attain a credential or certification of advanced status in a field. Even the ubiquitous requirements to obtain a driver's license involve an achievement testing component.

Although the purposes and contexts may vary, fairly uniform procedures are implemented for developing achievement tests and for evaluating their technical quality. Several sources exist for potential users of achievement tests to ascertain the quality of a particular test and its suitability for their purposes.

In any context where an achievement test is used, consequences for individual persons or groups may follow from test performance. In addition, the context and extent of achievement testing may have broad and

sometimes unforeseen consequences affecting, for example, the security of tests, the formats used for testing, and the relationship between testing and instruction.

## 2. DEFINITIONS AND EXAMPLES

Achievement testing refers to any procedure or instrument that is used to measure an examinee's attainment of knowledge or skills. Achievement testing can be done informally, as in when a teacher asks a student to perform a skill such as reading aloud or demonstrating correct laboratory technique. More formal, and perhaps more common, achievement tests are routinely administered in educational and occupational settings. Examples of formal achievement testing in education would include spelling tests, chemistry lab reports, end-of-unit tests, homework assignments, and so on.

More formal achievement testing is evident in large-scale, commercially available standardized instruments. The majority of these achievement tests would be referred to as standardized to the extent that the publishers of the instruments develop, administer, and score the tests under uniform, controlled conditions. It is important to note, however, that the term "standardized" is (a) unrelated to test format (although the multiple-choice format is often used for standardized tests, any format may be included) and (b) not synonymous with norm referenced (although sometimes the term "standardized" is used to indicate that a test has norms).

Examples of standardized achievement tests used in K–12 education would include the Iowa Tests of Basic Skills, the 10th edition of the Stanford Achievement Test, and the TerraNova. These tests ordinarily consist of several subtests, measuring achievement in specific narrow areas such as language arts, mathematics, science, and study skills. The composite index formed from these subtests (often referred to as a "complete battery score") provides a more global measure of academic achievement.

The preceding tests are also usually administered in a group setting, although individually administered achievement tests are also available and are designed for administration in a one-on-one setting with individual students, usually of very young age. Examples of individually administered achievement tests include the Woodcock–Johnson III Tests of Achievement, the third edition of the Developmental Indicators for the Assessment of Learning, and the Brigance Comprehensive Inventory of Basic Skills.

Following secondary schooling, achievement testing continues in colleges and universities, primarily in the

form of classroom achievement measures, but would also include standardized in-training examinations and board examinations for persons pursuing professional careers. Achievement testing has a long history in diverse occupational fields. Achievement tests are routinely administered to ascertain levels of knowledge or skills when screening or selecting applicants for positions in business and industry. These tests have traditionally been administered in paper-and-pencil format, although technology has enabled administration via computer or over the Internet to be secure, fast, and accessible. For example, one vendor of computerized achievement tests offers computerized “work sample” achievement tests to assist human resources personnel in selecting applicants for positions in legal offices, food service, information technology, accounting, medical offices, and others. Many state, federal, and private organizations also provide achievement tests for a variety of fields in which licensure or certification is required.

### 3. TYPES OF ACHIEVEMENT TESTS

In the previous section, it was noted that achievement tests could be categorized according to administration (group or individual) and scale (informal classroom tests or more formal commercially available tests). Another more important distinction focuses on the intended purpose, use, or inference that is to be made from the observed test score.

Less formal classroom achievement tests are usually developed by a teacher to align with an instructional unit, or they may be pre-prepared by publishers of classroom textbooks or related materials. The primary purposes of such tests are for educators’ use in refining instruction and assigning grades as well as for both teacher and pupil use in understanding and responding to individual students’ strengths and weaknesses.

More formal standardized achievement tests can also be categorized according to the inferences they yield. Three such types of tests—criterion-referenced tests (CRTs), standards-referenced tests (SRTs), and norm-referenced tests (NRTs)—are described in this section.

CRTs are designed to measure absolute achievement of fixed objectives comprising a domain of interest. The content of CRTs is narrow, highly specific, and tightly linked to the specific objectives. Importantly, a criterion for judging success on a CRT is specified a priori, and performance is usually reported in terms of pass/fail, number of objectives mastered, or similar terms. Thus, an examinee’s performance or score on a CRT is

interpreted with reference to the criterion. The written driver’s license test is a familiar example of a CRT.

SRTs are similar to CRTs in that they are designed to measure an examinee’s absolute level of achievement vis-à-vis fixed outcomes. These outcomes are narrowly defined and are referred to as content standards. Unlike CRTs, however, interpretation of examinees’ performance is referenced not to a single criterion but rather to descriptions of multiple levels of achievement called performance standards that illustrate what performance at the various levels means. Typical reporting methods for SRTs would consist of performance standard categories such as basic, proficient, and advanced or beginner, novice, intermediate, and expert levels. Familiar examples of SRTs include state-mandated testing for K–12 students in English language arts, mathematics, and so on to the extent that the tests are aligned with the state’s content standards in those subjects. At the national level, the National Assessment of Educational Progress (NAEP) is administered at regular intervals to samples of students across the United States.

NRTs are designed to measure achievement in a relative sense. Although NRTs are also constructed based on a fixed set of objectives, the domain covered by an NRT is usually broader than that covered by a CRT. The important distinction of NRTs is that examinee performance is reported with respect to the performance of one or more comparison groups of other examinees. These comparison groups are called norm groups. Tables showing the correspondence between a student’s performance and the norm group’s performance are called norms. Thus, an examinee’s performance or score on an NRT is interpreted with reference to the norms. Typical reporting methods for NRTs include  $z$  scores, percentile ranks, normal curve equivalent scores, grade- or age-equivalent scores, stanines, and other derived scale scores. Familiar examples of NRTs include the Iowa Tests of Basic Skills (ITBS), the Scholastic Assessment Test (SAT), and the Graduate Record Examinations (GRE).

Many publishers of large-scale achievement tests for school students also provide companion ability tests to be administered in conjunction with the achievement batteries. The tests are administered in tandem to derive ability/achievement comparisons that describe the extent to which a student is “underachieving” or “overachieving” in school given his or her measured potential. Examples of these test pairings include the Otis–Lennon School Abilities Test (administered with the Stanford Achievement Test) and the Cognitive Abilities Test (administered with the ITBS).

#### 4. ACHIEVEMENT TEST CONSTRUCTION

Rigorous achievement test development consists of numerous common steps. Achievement test construction differs slightly based on whether the focus of the assessment is classroom use or larger scale. **Table I** provides a sequence listing 18 steps that would be common to most achievement test development.

In both large and smaller contexts, the test maker would begin with specification of a clear purpose for the test or battery and a careful delineation of the domain to be sampled. Following this, the specific standards or objectives to be tested are developed. If it is a classroom achievement test, the objectives may be derived from a textbook, an instructional unit, a school district curriculum guide, content standards, or another source. Larger scale achievement tests (e.g., state mandated, standards referenced) would begin the test development process with reference to adopted state content standards. Standardized norm-referenced instruments would ordinarily be based on large-scale curriculum reviews, based on analysis of content standards adopted in various states, or

**TABLE I**  
**Common Steps in Achievement Test Development**

---

1. Establish need, purpose
2. Delineate domain to be tested
3. Develop specific objectives, content standards
4. Decide on item and test specifications, formats, length, costs
5. Develop items, tasks, scoring guides
6. Conduct item/task review (editorial, appropriateness, alignment, sensitivity)
7. Pilot/Field test items/tasks/scoring guides
8. Review item/task performance
9. Create item bank/pool
10. Assemble test form(s) according to specifications
11. Develop test administration guidelines, materials
12. Establish performance standards
13. Administer operational test forms
14. Score test
15. Evaluate preliminary item/task and examinee performance
16. Report scores to appropriate audiences
17. Evaluate test, document test cycle
18. Update item pool, revise development procedures, develop new items/tasks

---

promulgated by content area professional associations. Licensure, certification, or other credentialing tests would seek a foundation in job analysis or survey of practitioners in the particular occupation. Regardless of the context, these first steps involving grounding of the test in content standards, curriculum, or professional practice provide an important foundation for the validity of eventual test score interpretations.

Common next steps would include deciding on and developing appropriate items or tasks and related scoring guides to be field tested prior to actual administration of the test. At this stage, test developers pay particular attention to characteristics of items and tasks (e.g., clarity, discriminating power, amenability to dependable scoring) that will promote reliability of eventual scores obtained by examinees on the operational test.

Following item/task tryout in field testing, a database of acceptable items or tasks, called an item bank or item pool, would be created. From this pool, operational test forms would be drawn to match previously decided test specifications. Additional steps would be required, depending on whether the test is to be administered via paper-and-pencil format or computer. Ancillary materials, such as administrator guides and examinee information materials, would also be produced and distributed in advance of test administration. Following test administration, an evaluation of testing procedures and test item/task performance would be conducted. If obtaining scores on the current test form that were comparable to scores from a previous test administration is required, then statistical procedures for equating the two test forms would take place. Once quality assurance procedures have ensured accuracy of test results, scores for examinees would be reported to individual test takers and other groups as appropriate. Finally, documentation of the entire process would be gathered and refinements would be made prior to cycling back through the steps to develop subsequent test forms (Steps 5–18).

#### 5. EVALUATING ACHIEVEMENT TESTS

In some contexts, a specific achievement test may be required for use (e.g., state-mandated SRTs). However, in many other contexts, potential users of an achievement test may have a large number of options from which to choose. In such cases, users should be aware of the aids that exist to assist them in making informed choices.

One source of information about achievement tests is the various test publishers. Many publishers have online information available to help users gain a better understanding of the purposes, audiences, and uses of their products. Often, online information is somewhat limited and rather nontechnical. However, in addition to providing online information, many publishers will provide samples of test materials and technical documentation on request to potential users. Frequently, publishers will provide one set of these packets of information, called specimen sets, at no charge for evaluation purposes.

When evaluating an achievement test, it is important to examine many aspects. A number of authorities have provided advice on how to conduct such a review. For example, one textbook for school counselors by Whiston contains a section titled "Selection of an Assessment Instrument" that consists of several pages of advice and a user-friendly checklist. The single authoritative source for such information would likely be the Standards for Educational and Psychological Testing, jointly sponsored by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education.

Finally, a particularly useful framework for evaluating achievement tests was developed by Rudner in 1994. Table II provides a modified version of key points identified by Rudner that should be addressed when choosing an achievement test.

It is likely that some potential users will not have the time or technical expertise necessary to fully evaluate an achievement test independently. A rich source of information exists for such users in the form of published reviews of tests. Two compilations of test reviews are noteworthy: *Mental Measurements Yearbook (MMY)* and *Tests in Print*. These references are available in nearly all academic libraries. In the case of MMY, the editors of these volumes routinely gather test materials and forward those materials to two independent reviewers. The reviewers provide brief (two- to four-page) summaries of the purpose, technical qualities, and administration notes for the test. Along with the summaries, each entry in MMY contains the publication date for the test, information on how to contact the publisher, and cost information for purchasing the test. In these volumes, users can compare several options for an intended use in a relatively short time.

TABLE II  
Evaluation Criteria for Achievement Tests

- 
1. Is the purpose of the test clearly stated? What achievement construct is claimed? Is the construct or intended content domain clearly delineated?
  2. What are the intended uses of the test? What are the intended audiences for the test results?
  3. For what ages, grade levels, or subject areas is the test intended?
  4. Are the test materials (e.g., booklets, answer sheets) clear, engaging, and appropriate for the age/grade level of the examinees?
  5. What are the costs of the test materials, scoring, training personnel, and time required to administer the test?
  6. Are the procedures for administering the test clear? Is the information provided sufficiently detailed so as to provide consistent administrations across users and contexts?
  7. What are the qualifications of those who participated in the development of the test? What qualifications are required for test administrators?
  8. How were samples selected for developing, pilot testing, norming, estimating reliability, screening out potentially biased items, and gathering validity evidence for the test? Were the samples relevant? Were they representative? Were they collected recently?
  9. Does the test yield scores of acceptable reliability? Were appropriate methods used to compute reliability estimates? If decisions are to be made based in part on test performance, what is the evidence regarding decision consistency?
  10. Is there adequate validity evidence to support the intended inferences, uses, interpretations, or meanings that will be made from test scores? Is there evidence to support the use of the test with various groups (including non-native speakers of English, special needs learners, students in need of testing accommodations, etc.)?
  11. Is the scoring system likely to produce accurate scores? If hand scoring is involved, are scoring keys easy to use and is conversion of raw scores to derived scores facilitated by tables, look-up charts, and the like? If machine scoring is used, are answer documents easy for examinees to use? Are timelines and costs for scoring reasonable?
  12. Do score reports provide a clear, detailed, and comprehensible summary of performance and diagnostic information? Are users appropriately cautioned about likely misinterpretations?
- 

Source. Adapted from Rudner (1994).

A fee-based search capability for locating test reviews is available at the MMY Web site ([www.unl.edu](http://www.unl.edu)).

### See Also the Following Articles

Educational Achievement and Culture

### Further Reading

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Brigance, A. H., & Glascoe, F. P. (1999). *Brigance Comprehensive Inventory of Basic Skills* (rev. ed.). North Billerica, MA: Curriculum Associates.
- Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 1–32). San Diego: Academic Press.
- Cizek, G. J. (2003). *Detecting and preventing classroom cheating: Promoting integrity in schools*. Thousand Oaks, CA: Corwin.
- CTB/McGraw-Hill. (1997). *TerraNova*. Monterey, CA: Author.
- Gronlund, N. E. (1993). *How to make achievement tests and assessments*. Boston: Allyn & Bacon.
- Harcourt Educational Measurement. (2002). *Stanford Achievement Test* (10th ed.). San Antonio, TX: Author.
- Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Tests of Basic Skills*. Itasca, IL: Riverside.
- Mardell-Czudnowski, C., & Goldenberg, D. S. (1998). *Developmental indicators for the assessment of learning* (3rd ed.). Circle Pines, MN: American Guidance Services.
- Rudner, L. (1994, April). *Questions to ask when evaluating tests* (ERIC/AE Digest, EDO-TM-94-06). Washington, DC: ERIC Clearinghouse on Assessment and Evaluation.
- Whiston, S. C. (2000). *Principles and applications of assessment in counseling*. Belmont, CA: Wadsworth.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.



# Advertising and Culture

Sharon Shavitt and Jing Zhang

University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

1. Advertising and Advertising Appeals
2. Approaches to Culture in the Context of Advertising and Consumer Behavior
3. Conclusions  
Further Reading

## GLOSSARY

**advertising appeal** The central theme or idea behind an advertising message.

**collectivism** A cultural value system that emphasizes in-group goals such as family integrity, harmonious relationships, and the well-being of the in-group.

**horizontal-vertical dimension** A cultural distinction within the individualism-collectivism framework; members of horizontal societies value equality and view the self as having the same status as others in society, whereas members of vertical societies view the self as differing from others along a hierarchy, accepting inequality, and believing that rank has its privileges.

**independent self-construal** A view of the self that focuses on one's own psychological traits, feelings, and actions.

**individualism** A cultural value system that emphasizes being independent and pursuing individual goals rather than in-group goals.

**interdependent self-construal** A view of the self that focuses on one's social roles or memberships.

Culture is a crucial concept to the understanding of advertising and consumer behavior because it is the

lens through which people view advertising messages and products. This article reviews and discusses a series of studies on the relations between culture and advertising. The content is organized around different levels of analysis of culture, for example, national level versus individual level. The individualism-collectivism cultural construct and independent-interdependent self-construal construct are given special attention because extensive research has demonstrated the persuasion implications of these variables for advertising content, attitude toward advertisements, and cognitive processing of advertisements.

## 1. ADVERTISING AND ADVERTISING APPEALS

Advertising has been defined as the nonpersonal communication of information, usually paid for and usually persuasive in nature, about products (goods or services) or ideas by identified sponsors through various media. The central theme or idea behind an advertising message is called the advertising appeal.

## 2. APPROACHES TO CULTURE IN THE CONTEXT OF ADVERTISING AND CONSUMER BEHAVIOR

Intensified global marketing and advertising activities have given rise in the advertising industry to the

advertising standardization versus specification debate. The standardization proponents believe that people around the world have the same needs and desires and that they would react to advertising messages in the same or a similar way. The specification proponents argue that cultural differences among nations make advertising one of the most difficult marketing elements to standardize.

Culture consists of shared elements that provide the standards for perceiving, believing, evaluating, communicating, and acting among those who share a language, a historical period, and a geographical location. As a psychological construct, culture can be studied at multiple levels—across nations, across individuals within nations, and even within individuals through the priming of cultural values. Regardless of how culture is studied, cultural distinctions have been demonstrated to have important implications for advertising content, advertising persuasion, and consumer response.

## 2.1. Cultural Differences at the National Level

### 2.1.1. Individualism and Collectivism

Individualism and collectivism comprise the main cultural distinction that has been explored in studying both advertising content and the persuasiveness of ad appeals. Individualism and collectivism have been conceptualized as two powerful cultural models that represent broad differences among nations. As Hofstede, Triandis, and others have proposed, members of collectivistic cultures endorse in-group goals, such as family integrity, harmonious relationships, and the well-being of the in-group, whereas members of individualistic cultures endorse being independent and pursuing individual goals instead of in-group goals. Extensive cross-cultural (cross-national) data have shown that North American and most European countries, such as the United States, Canada, Germany, and Denmark, are individualistic societies and that most East Asian and Latin American countries, such as China, Korea, Japan, and Mexico, are collectivistic societies.

Past research has indicated that more individualistic and less collectivistic advertising appeals are present in individualistic cultures than in collectivistic cultures. For example, Han and Shavitt's 1994 study found that Korean ads tended to employ appeals emphasizing in-group benefits, harmony, and family integrity, whereas U.S. ads tended to employ appeals emphasizing individual enjoyment, personal success, and independence.

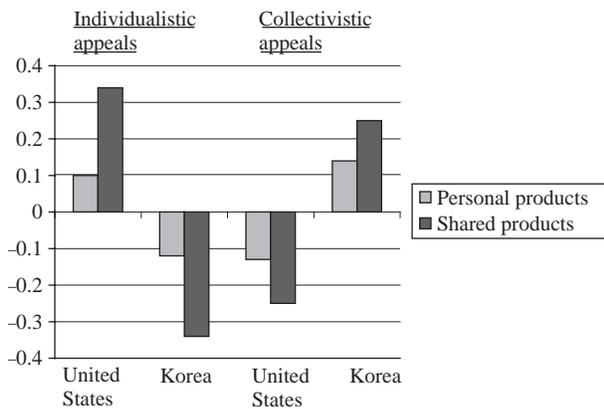
Similarly, in other research, more conformity themes (e.g., respect for collective values and beliefs) and fewer uniqueness themes (e.g., rebelling against collective values and beliefs) were found to be present in Korean ads compared with U.S. ads. In studying humorous appeals, ads from both Korea and Thailand were found to contain more group-oriented situations than those from Germany and the United States.

Existing research has also indicated that culturally matched ad appeals are more likely to be persuasive than are mismatched appeals. U.S. respondents found ads emphasizing individualistic benefits to be more persuasive, and found ads emphasizing family or in-group benefits to be less persuasive, than did Korean respondents. Also, Chinese participants responded more favorably to collective ad appeals (e.g., "share the moments of happiness") than to individualistic appeals (e.g., "the joy of self-expression"), whereas the reverse was true for U.S. participants. In addition, advertisements in Mexico that depict values that are consistent with the local cultural norms and roles (e.g., familial norms and roles in Mexican culture) elicited more favorable attitudes and purchase intentions than did ads that depict inconsistencies.

Product characteristics, such as whether a product is personal or shared (i.e., privately used vs publicly visible), have been found to moderate the effect of culture on the persuasiveness of ad appeals. Only for shared or socially visible products—those purchased or used with in-group members—did strong cultural differences emerge between Americans and Koreans in Han and Shavitt's 1994 study (Fig. 1) as well as between Americans and Chinese in other research.

### 2.1.2. Horizontal and Vertical Cultural Orientations

Within the individualism–collectivism framework, a distinction between horizontal and vertical societies has recently been introduced. Members of horizontal societies, such as Denmark and Sweden, value equality and view the self as having the same status as others in society. In contrast, members of vertical societies view the self as differing from others along a hierarchy, accepting inequality and believing that rank has its privileges. In comparison with the individualism–collectivism dimension, the horizontal–vertical dimension is less explored. However, recent studies have begun to examine the implications of the horizontal–vertical cultural dimension. In 2002, Nelson and Shavitt found that members of a relatively vertical society (the



**FIGURE 1** Persuasiveness of individualistic and collectivistic appeals in the United States and Korea. Bars represent mean standardized scores on an attitude index (the average of standardized scores across three evaluative measures). Adapted from Han and Shavitt (1994), *J. Exp. Soc. Psychol.*, 30, 326–350, Copyright 1994, with permission from Elsevier.

United States) are more likely to endorse achievement values than are members of a horizontal society (Denmark). Moreover, Shavitt and colleagues' 2002 research suggested that vertical collectivism in the United States is associated with a tendency to be persuaded by ad appeals to status and hierarchy.

In 2000, Gurhan-Canli and Maheswaran demonstrated that Japanese consumers tended to favor products from their own country over foreign products, whereas American consumers tended to favor high-quality products regardless of the countries of origin of these products. Mediation analyses indicated that the vertical dimension of individualism and collectivism accounted for the country-of-origin effects in Japan. This is consistent with the cultural orientation in Japan—vertical collectivism—in that Japanese tend to believe in and value the superiority of the in-group over others.

In terms of advertising content, studies examining humor executions in ads have indicated that relationships between the central characters in ads where humor was intended were more often unequal in high power distance cultures (or vertical cultures such as Korea and Thailand) than in low power distance cultures (or horizontal cultures such as Germany); in the latter, these relationships were more often equal. Such unequal relationships in the ads are believed to reflect the hierarchical interpersonal relationships that are more likely to exist in high power distance (or vertical) cultures.

## 2.2. Self-Construal: Cultural Differences at the Individual Level

In 1991, Markus and Kitayama proposed that individualistic and collectivistic distinctions affect how individuals define or construct themselves in relation to others. According to these authors, the cultural norms of individualism require defining oneself as an individual whose behavior is organized and made meaningful primarily by reference to one's own thoughts, feelings, and actions. This independent self-construal is associated with describing oneself with reference to individual psychological traits such as "I am easygoing." On the other hand, the cultural norms of collectivism require individuals to see themselves as part of harmonious relationships and to recognize that their own behavior should seek benefits to their in-groups. This interdependent self-construal is associated with describing oneself with reference to social roles or memberships such as "I am in the gymnastics club." Recent research has indicated that self-construals account for and moderate cultural-level persuasion effects.

### 2.2.1. Chronic Self-Construal

Cross-cultural experiments have demonstrated that consumers with a more interdependent self-construal responded favorably to ads with a connected theme (e.g., interdependence, togetherness), whereas consumers with a more independent self-construal responded favorably to ads with a separate theme (e.g., independence, autonomy).

Self-construals have also been found to affect regulatory focus. Regulatory focus is built on the general hedonic notion that people approach pleasure and avoid pain. In 2001, Aaker and Lee found that, for people with an independent self-construal, promotion-focused information (e.g., drinking juice promotes energy or enjoyment of life) led to more positive attitudes toward a juice product. In contrast, for people with an interdependent self-construal, prevention-focused information (e.g., drinking juice helps to prevent disease) led to more positive attitudes.

### 2.2.2. Salient Self-Construal

Although an individual's cultural orientation may strongly determine the self-construal that is chronically accessible, self-construals may shift in response to other factors such as situational accessibility, motives, and reference points. The effect of temporarily accessible

self-construals on persuasion parallels that of chronic self-construals. For instance, regardless of whether self-construal is measured with the Self-Construal Scale or is situationally activated, studies have shown a consistent effect of independent self-construal versus interdependent self-construal on preference for either promotion-focused information or prevention-focused information.

The persuasion implications of salient self-construals have been further evidenced among bicultural consumers in the United States, that is, individuals who have been equally influenced by East Asian and Western cultural orientations. For example, Lau-Gesk's 2003 study found that an independent self-construal was temporarily activated when bicultural individuals were exposed to an individually focused appeal, and this in turn guided them to respond favorably to the advertised product. On the other hand, an interdependent self-construal was activated when bicultural individuals were exposed to an interpersonally focused appeal, and this in turn guided them to respond favorably to the advertised product. A frame-switching mechanism has been used to explain these findings in that bicultural individuals appear to be capable of shifting between individualistic and collectivistic interpretive frames in response to persuasive cues.

### **2.3. Culture and Cognitive Processing of Advertising Appeals**

Just as psychological research on culture has focused extensively on cognitive process issues, recent research has demonstrated that a consumer's culture or cultural orientation influences the nature of information processing that accompanies a message.

Some studies along this line have focused on examining the effect of cultural orientation on persuasion by application of the Elaboration Likelihood Model/Heuristic-Systematic Model. Individualism and collectivism appear to be associated with different perceptions of cue diagnosticity in the processing of ads, and these in turn result in different patterns of attitudinal outcomes. Because of their individualistic tendencies, Americans are relatively uninterested in others' opinions; conversely, because of their collectivistic tendencies, people in Hong Kong tend to be more affected by the preferences of others. Consequently, among U.S. participants, product attribute information tends to be seen as more diagnostic than do consensus cues. In contrast, in 1997, Aaker and Maheswaran found that among Hong Kong participants, consensus cues are

treated as relatively diagnostic. Moreover, in 2000, Aaker and Sengupta's research demonstrated that individualistic versus collectivistic cultures are also associated with different strategies used in resolving incongruities between a source cue and product attribute information. In an individualistic cultural context, attribute information is perceived as more diagnostic than is a source cue and influences product evaluations. Hence, an attenuation strategy is used in information processing when incongruity between a source cue and attribute information occurs. In a collectivistic cultural context, where source information is relatively important and where there is less of a tendency to choose one type of information over another, the source cue and attribute information jointly influence evaluations. Hence, an additive strategy is applied when there is incongruity between a source cue and attribute information.

Culture also influences the role of affect in processing a message. Ego-focused (e.g., pride, happiness) versus other-focused (e.g., empathy, peacefulness) emotional appeals lead to more favorable attitudes among members of a collectivistic culture, whereas other-focused versus ego-focused emotional appeals lead to more favorable attitudes among members of an individualistic culture. This differs from the standard cultural congruity findings about persuasive effects. That is, relatively novel types of appeals, rather than culturally congruent appeals, elicit more elaboration. Because elaboration involves effortful thinking about the merits of a message as opposed to the use of simple cues, such elaborated thoughts tend to have a relatively large impact on evaluations.

### **2.4. Factors That Covary with Culture**

Race, gender, and economic factors also affect ad content, ad processing, and attitude formation, although these effects are beyond the scope of this article. Very often, these variables will covary with culture, making studies of race, gender, and economic factors in conjunction with cultural ones particularly important.

The standard cultural congruity findings, suggesting that culturally matched ad appeals are more prevalent or are more likely to be persuasive, were reviewed earlier. However, a growing number of studies have indicated that the situation in rapidly transitioning economies may be more complex. For example, Westernized appeals, such as appeals to technology, individuality/

independence, and youth/modernity, have been found to be rather salient in Chinese and Japanese ads and to have been used frequently in current Taiwanese advertising agencies. These cultural incongruity findings could be driven by government policies regarding internal development and modernization, public exposure to Western media, and demographic and geographic contact zones. In addition, consumers in developing countries tend to respond favorably to Western products. For example, in one study of Indian consumers, brands perceived as having nonlocal (Western) countries of origin were favored over brands perceived as local. Moreover, this effect was found to be stronger for consumers who have a greater admiration for the lifestyle in economically developed countries. These cultural incongruity findings are meaningful because, as discussed by Zhang and Shavitt in 2003, they suggest the important role that advertising plays in reshaping cultural values in developing countries that are experiencing rapid economic growth. Rather than reflecting existing cultural values, advertising content in those countries promotes new aspirational values such as individuality and modernity; hence, these new values become acceptable and desirable among consumers.

### 3. CONCLUSIONS

As the advertising industry has become increasingly globalized, there has been increased research attention given to the role of culture in advertising. Progress in understanding this role has been made on a number of fronts. This article has reviewed empirical research on the effects of a number of cultural variables on advertising-related processes. Distinctions between individualistic and collectivistic cultures have been demonstrated to have important implications for advertising content, advertising persuasion, and consumers' cognitive responses. Parallel effects of culture and self-construal (independent vs interdependent) have been shown as well. In addition, studies on the distinction between horizontal and vertical cultural orientations within individualism–collectivism have begun to show advertising implications.

### See Also the Following Articles

Industrial/Organizational Psychology across Cultures  
 ■ Perception and Culture ■ Values and Culture

### Further Reading

- Aaker, J. L., & Lee, A. Y. (2001). "I" seek pleasure and "we" avoid pains: The role of self-regulatory goals in information processing and persuasion. *Journal of Consumer Research*, 28, 33–49.
- Aaker, J. L., & Maheswaran, D. (1997). The effect of cultural orientation on persuasion. *Journal of Consumer Research*, 24, 315–328.
- Aaker, J. L., & Sengupta, J. (2000). Additivity versus attenuation: The role of culture in the resolution of information incongruity. *Journal of Consumer Psychology*, 9(2), 67–82.
- Gurhan-Canli, Z., & Maheswaran, D. (2000). Cultural variations in country of origin effects. *Journal of Marketing Research*, 37, 309–317.
- Han, S., & Shavitt, S. (1994). Persuasion and culture: Advertising appeals in individualistic and collectivistic societies. *Journal of Experimental Social Psychology*, 30, 326–350.
- Hofstede, G. (1980). *Culture's consequences*. Beverly Hills, CA: Sage.
- Lau-Gesk, L. G. (2003). Activating culture through persuasion appeals: An examination of the bicultural consumer. *Journal of Consumer Psychology*, 13, 301–315.
- Maheswaran, D., & Shavitt, S. (2000). Issues and new directions in global consumer psychology. *Journal of Consumer Psychology*, 9(2), 59–66.
- Markus, H., & Kitayama, S. (1991). Culture and the self: Implication for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253.
- Nelson, M. R., & Shavitt, S. (2002). Horizontal and vertical individualism and achievement values: A multimethod examination of Denmark and the United States. *Journal of Cross-Cultural Psychology*, 33, 439–458.
- Shavitt, S., Zhang, J., & Johnson, T. (2002). Horizontal and vertical orientations in cross-cultural consumer persuasion [abstract]. In S. Bronicarczy, & K. Nakamoto (Eds.), *Advances in consumer research* (Vol. 29, p. 47). Provo, UT: Association for Consumer Research.
- Triandis, H. C. (1995). *Individualism and collectivism*. Westview.
- Zhang, J., & Shavitt, S. (2003). Cultural values in advertisements to the Chinese X-Generation: Promoting modernity and individualism. *Journal of Advertising*, 32(1), 21–31.



# Advertising Psychology

Patrick T. Vargas and Sukki Yoon

University of Illinois at Urbana–Champaign, Urbana, Illinois, USA

1. Introduction
  2. The Psychological Study of Advertising
  3. Perception
  4. Memory and Learning
  5. Attitudes
  6. Persuasion
- Further Reading

## GLOSSARY

**advertising** A paid, mass-mediated attempt to inform and/or persuade.

**AIDA model** The acronym for Awareness, Interest, Desire, Action—a description of the stages by which advertising may have the desired effect on consumers.

**attitude** A psychological tendency to evaluate something (e.g., a person, an object, an idea) with some degree of favor or disfavor.

**balance theory** Theory that considers attitudinal relationships among a perceiver, an other, and a target; from the perceiver's perspective, balanced relationships are preferable to unbalanced relationships.

**belief** A piece of information about a person, an object, or an idea that an individual has in memory.

**cognitive consistency** A state when attitudes, beliefs, and behaviors are logically harmonious with one another.

**cognitive dissonance** An unpleasant state aroused by awareness of inconsistent cognitions.

**Elaboration Likelihood Model and Heuristic-Systematic Model** Dual-process theories of persuasion that posit both high-effort (central or systematic processing) and

low-effort (peripheral or heuristic processing) routes to persuasion.

**memory** The mental ability to store and recall (intentionally or unintentionally) past experiences.

**mere exposure effect** Generation of favorable attitudes toward a given stimulus due to repeated (subliminal) exposure to that stimulus.

**perception** Recognition and interpretation of sensory stimuli.

**persuasion** The processes by which an individual's attitudes are changed.

**priming** A technique for increasing the accessibility, and likelihood of use, of some concept.

**self-perception** A theory suggesting that individuals infer their attitudes based on past behavior, particularly when they are uncertain of their attitudes.

**subliminal advertising** An attempt to advertise without the awareness of the message recipient.

Principles of psychological science have been applied to advertising practice since both fields were relatively young. The study of information processing, attitudes, and persuasion creates a foundation for advertising psychology because each is an important determinant of achieving advertising's main functions: to inform, persuade, and influence.

## 1. INTRODUCTION

The *Advertisers Handbook*, published in 1910 by the International Textbook Company, offered a wide range of advice to advertising professionals. According to the

handbook, an advertisement should not be too general, a headline should not be silly or deceptive, and an ad should be arranged logically, be concise, contain a proper amount of matter for the commodity being sold, and so forth. A few pages of the handbook were even devoted to the use of psychology in advertising: "Study of the goods or service to be sold is highly important, but no more important than the study of that wonderful subject, the human mind. The advertiser will do well, in all his work, to give special attention to psychological principles."

The handbook's advice has been heeded. Advertisers have used both psychological theory and method to gain a better understanding of consumers. The study of information processing has been particularly influential as advertisers have sought to gain consumers' attention and ensure that consumers interpret and understand their messages and (hopefully) retain some of the information. The attitude construct has been particularly important in advertising psychology because advertisements are generally intended to create positive attitudes toward some objects and because attitudes are an important determinant of behavior. Thus, an understanding of attitudes is doubly important for advertisers given that attitudes are both desired outcomes and causes of desired behavior. Therefore, the psychological study of attitude change, or persuasion, is another crucial part of advertising psychology. This article reviews important aspects of psychology that also play an important role in advertising.

## 2. THE PSYCHOLOGICAL STUDY OF ADVERTISING

### 2.1. Historical Perspectives

The scientific study of psychology is roughly 125 years old, with the first laboratory of psychology being founded in 1879 by Wilhelm Wundt in Leipzig, Germany. Modern advertising is only slightly older than scientific psychology, having appeared in the earliest newspapers circa 1850, reaching a circulation of approximately 1 million readers. The two fields came together in the earliest application of psychological principles to advertising practice in America in 1896. At the University of Minnesota, Harlow Gale conducted systematic studies of the influence of ad placement within a page on attention, the impact of necessary versus superfluous words in headlines, and how various colors used in ads might influence readers' attention.

Gale's research did not receive wide attention, but several years later another researcher, Walter Dill Scott,

wrote a series of articles on psychological aspects of advertising. Scott's research on advertising psychology tended to focus on the concept of suggestion. He believed that advertising was primarily a persuasive tool, rather than an informational device, and that advertising had its effect on consumers in a nearly hypnotic manner. People were thought to be highly susceptible to suggestion, under a wide variety of conditions, so long as the suggestion was the only thought available to them. According to Scott, advertising was most effective when it presented consumers with a specific direct command. "Think Different," a suggestion used by Apple Computer, is a good example of a direct command, but it would not necessarily be an effective advertisement because it does not tell consumers how to think differently. A more effective advertisement might be "Use Apple Computers."

### 2.2. Contemporary Perspectives

It is interesting to note that the concept of suggestion went out of style for some time among scientific psychologists. It is somewhat unflattering to think of people as automatons who unthinkingly follow whatever instruction is given to them; however, contemporary scientific psychologists have once again embraced the notion that a fair proportion of human behavior is due to influences of which people are entirely unaware. One intriguing line of contemporary research has shown that simply asking people whether they intend to purchase a new car sometime during the next year dramatically increases the chances that they actually do purchase a new car during that year—even if they respond "no" to the initial inquiry. Remarkable as these ideas are, advertising has relied on a number of basic, tried-and-true psychological principles that are covered in the remainder of this article.

One very general model of advertising effects, the AIDA (Attention, Interest, Desire, Action) model, has roots in Scott's writings. In 1913, Scott proposed a model of advertising effects that had three stages: attention, comprehension, and understanding. Over the course of several decades, Scott's model evolved into the AIDA model, which is still in use. According to this model, advertising, or promotions more generally, must first garner the attention of consumers and help them to develop beliefs about the product or service. Second, advertising should create interest or positive feelings about the product/service. Third, advertising or promotions should instill in consumers a desire for the product/service, thereby helping them to form purchase intentions. Finally, consumers must be

convinced to take action, that is, to purchase the product/service. This article covers all of these aspects of advertising psychology and more. It is important to note, however, that the AIDA model represents a strong version of advertising effects.

### 2.3. Strong and Weak Models of Advertising Effects

Advertising can strongly persuade consumers into immediate buying, or it can have a more subtle effect by reinforcing people's existing propensities to buy certain brands. The psychological processes underlying these two mechanisms also differ. The strong model focuses on consumers' immediate psychological or behavioral reactions where explicit advocacy and rationales of advertising messages are vital. The weak model emphasizes brand awareness where advertising is viewed as a reminder of a brand or source of information. For example, according to the strong model, Mark may decide to go to the shopping mall immediately after viewing a television commercial that says, "40% off any purchase at Macy's." On the other hand, David may buy Coke instead of Pepsi because he has greater familiarity with the Coke brand name, although both brands were initially in his consideration set.

Whether the strong or weak theory provides a better explanation for advertising effects hinges on numerous factors. For example, strong effects are more likely to occur in "high-involvement" situations where advertising directly aims at changing individuals' attitudes. However, when consumers have a predetermined set of alternative brands or are in a "repeat buying" situation where purchase decisions tend to be habitual, weak reinforcement enhances long-term brand awareness, familiarity, salience, and brand associations. In any event, for advertising to have either strong or weak effects, it must first be perceived by prospective consumers.

## 3. PERCEPTION

The first issue is the process of getting the advertising "into" consumers. The sheer amount of stimuli around people is quite overwhelming. Consider all of the following stimuli to which people are constantly being exposed but are unconsciously "filtering out": the air rushing into their mouths or nostrils as they breathe, the sensation of tightness where their shoes are touching their feet, the color of the walls nearest

them, and so forth. When one considers all of these things, it is remarkable that any advertising makes it into people's consciousness at all. Nevertheless, if advertising is to reach consumers, they must first and foremost be exposed to it.

### 3.1. Exposure

Exposure is the first stage in perception, and it occurs when people achieve physical proximity to some stimulus. Whenever something activates their sensory receptors, it means they have been exposed to that something. A printed advertisement must activate rods and cones in people's eyes, and a radio advertisement must move the small bones that make up people's inner ears.

This activation of sensory receptors may occur at different threshold levels, with the lower, terminal, and difference thresholds being important to advertisers. The lower threshold is the minimum amount of stimulus that is necessary for people to be aware of a sensation. Images presented very briefly (i.e., a few milliseconds) fall below the lower threshold and are imperceptible to people. The terminal threshold is the point at which additional increases in stimulus intensity produce no awareness of the increase. The difference threshold is the smallest noticeable change in stimulus intensity. These threshold levels have aroused interest among advertisers due to their application to subliminal advertising.

During the late 1950s, Jim Vicary claimed to be able to influence people without them even knowing that they had been exposed to influence attempts. Vicary claimed to have increased soft drink and popcorn sales by subliminally presenting messages to movie viewers. Years later, Vicary confessed to fabricating his claims, but the notion that people could be subliminally influenced took hold somewhere in the popular imagination. Sporadically throughout the next several decades, the question of whether subliminal presentation of persuasive messages could influence consumer behavior was debated. It is now known that subliminal presentation of stimuli can influence attitudes and behavior, but not quite in the way that Vicary suggested.

In 1968, Robert Zajonc discovered that positive attitudes can be induced simply by repeatedly exposing people to a subliminally presented stimulus, a finding that he termed the "mere exposure" effect. Participants in these studies were repeatedly exposed to a set of unfamiliar stimuli (e.g., Chinese ideographs), although the exposures were so fast that participants were unaware of even having seen the stimuli. Yet, when asked to evaluate a variety of similar stimuli, some of

which they had been exposed to and others of which had not been presented to them, they evaluated the stimuli to which they had been previously exposed more positively than they evaluated similar stimuli that had not been presented to them subliminally. Of course, developing favorable attitudes toward something is a long way from actually getting people to buy something; however, it is now known that subliminal presentation of stimuli can also influence behavior.

Individuals can be “primed” by the subliminal (or even supraliminal under certain conditions) presentation of stimuli in essentially the same way that they prime a pump (i.e., by filling it with fluid so that it is ready for immediate use). Following the presentation of a primed concept, individuals are more likely to use that concept given an appropriate opportunity. For example, individuals primed with the concept of elderly were observed to walk slower than individuals primed with neutral concepts. However, priming is a relatively weak influence on behavior and simply could not make people get up in the middle of a movie to go fetch soft drinks and popcorn. There are far more effective ways in which to influence consumer behavior.

### 3.2. Attention

One of the enduring problems facing advertisers is attracting consumers’ attention. “Zipping and zapping” behavior is extremely common; television viewers record programs and zip (fast-forward) through the commercials or zap (eliminate commercials during recording or change the channel when they appear) them altogether. Drivers change the radio station when advertisements come on, they look away when billboards annoy them, and so forth. Attention is simply the conscious allocation of cognitive resources to some stimulus, but it is hard to grab.

A number of strategies have been used by advertisers to attract consumers’ attention. Theories of human motivation have been popular guidelines for advertisers. For example, Maslow’s hierarchy of needs posits five levels of motives that are common to all people: physiological (e.g., food, sex), safety (e.g., protection from the elements), belongingness (e.g., acceptance by others), esteem (e.g., admiration of others), and self-actualization (e.g., fulfillment of potential). Advertisers have tried to attract attention by positioning their products to fulfill one or more of these needs. Exercise equipment and healthy food appeal to physiological needs, bicycle helmets and smoke alarms appeal to safety needs, mouthwash and acne creams appeal to belongingness needs, designer clothing and luxury automobiles appeal

to esteem needs, and computer software and military service have even been made to appeal to self-actualization needs. One might imagine that nearly any product can be positioned to appeal to nearly any need—all in the name of attracting attention.

Properties of the stimulus can also be manipulated to attract attention. Size, color, intensity, contrast, position, movement, novelty, and so forth all can be used to attract attention.

### 3.3. Interpretation

Interpretation is the way in which meaning is assigned to stimuli. People exposed to exactly the same stimulus may interpret it very differently. For example, they may perceive a fuzzy kitten as an adorable target for cuddling or as a loathsome beast that causes sneezing and hives. Individual, situational, and stimulus characteristics all can influence the interpretation of stimuli.

At the individual level, consumers’ expectations, motives, and attitudes all can have a profound influence on the way in which they interpret advertisements. When clear cola was introduced, consumers expected it to taste different from the traditional caramel-colored colas; however, clear cola did not taste much different from brown colas, and many consumers were turned off by this. The violation of their expectations may have, at least in part, caused them to ultimately reject the clear cola. Motives and attitudes are particularly likely to have an influence on the interpretation of ambiguous stimuli. Advertisers who do not send clear messages may be opening themselves up to very different, and possibly unflattering, interpretations that reflect consumers’ own motives and attitudes.

At a situational level, the context in which an advertisement appears can have an influence on how it is interpreted. Several years ago, a prominent newspaper put out a Sunday supplement that featured a lengthy article on starvation and drought in Africa, followed immediately by a lengthy pictorial on the coming season’s high-fashion formal wear. The following week, there was a section in the same supplement that was devoted to outraged letters from readers and a formal apology by the newspaper. Rarely are pictorials on fashion the target of so much outrage, but the context made a huge difference.

At a stimulus level, interpretations vary a great deal as well. As noted previously, stimulus ambiguity opens doors for motives and attitudes to have their influence on interpretations. Other stimuli can have vastly different interpretations due to cultural differences.

Different colors and numbers have different culturally laden meanings in different countries. There used to be a brand of cold medicine called “666.” This brand likely did not fare well among conservative Christian consumers; but it might be prized in cultures where the number six has a positive meaning.

## 4. MEMORY AND LEARNING

Memory is complex; we can remember toys that we wanted as children, yet we sometimes cannot remember what we did the weekend before last. The simple change of a single word (e.g., how fast was the car going when it “bumped” [or “smashed”] into the pedestrian?) can substantially alter our memories. Yet it has been estimated that over the course of a lifetime, the average human stores approximately 500 times the amount of information that is in a full set of encyclopedias. Given these seemingly strange contradictions regarding memory, what hope can advertisers have of getting their products, brands, and ideas into consumers’ memories?

### 4.1. Encoding, Storage, and Retrieval

Memory involves three main processes: encoding (the process by which information is put into memory), storage (the process by which information is maintained in memory), and retrieval (the process by which information is recovered from memory).

Encoding may be visual, acoustic, or semantic. Visual encoding and acoustic encoding are self-explanatory; they are named for the sensory modality through which they operate. Semantic encoding refers to the general meaning of an event. For example, one might encode a television advertisement in terms of the visuals presented, the sounds that accompany it, or the general idea that there is a sale at the market.

Storage may be short term or long term. Short-term memory, or working memory, is of quite limited capacity and is used to hold information in consciousness for immediate use. Long-term memory is quite vast and can retain information for extremely long periods of time (e.g., some childhood memories last until death).

Retrieval also comes in different forms. Explicit memory is tapped by intentional recall or recognition of items or events. Implicit memory is the unintentional recollection and/or influence of prior experience on a current task. On implicit memory tests,

respondents are unaware that memory is being accessed. Implicit memory is assessed in a variety of ways such as word fragment completion (words seen previously are more likely to be completed than are words not seen previously) and time savings for tasks that have been done before. Advertisers may be particularly interested in explicit memory because the ability to intentionally recall information serves as a good measure of advertising effectiveness.

### 4.2. Short- and Long-Term Memory

Short-term memory is of relatively little consequence to advertisers because it does not, by definition, have any “staying power.” Short-term memory may be important for direct response advertising, where consumers are asked to call a phone number to order or request additional information. But there is little that an advertiser can do with short-term memory. The primary function of short-term memory is to allow people to perform mental work such as calculating a sum or remembering a telephone number until they can either dial it or write it down. The fairly recent trend of placing advertisements on shopping carts in supermarkets might be a good way for advertisers to keep their products in short-term memory while people are doing their grocery shopping. In general, however, for an advertisement to be effective, it must work its way into long-term memory.

One bright spot for advertisers hoping to etch their work in consumers’ long-term memory is the fact that most theorists believe that long-term memory has nearly unlimited capacity. So, there is always room for more information. But what is the best way in which to achieve storage in long-term memory?

Some information is stored in long-term memory accidentally (i.e., unintentionally), but most information that makes it to long-term memory is encoded semantically. According to one model of memory, information is better remembered when it is thoroughly processed. To the extent that people elaborate on, and think deeply about, a particular piece of information, they are more likely to be able to recall it later. This notion of cognitive elaboration is important in the understanding of persuasion as well. Clearly, advertisers would do well to generate ads that cause people to think deeply or discuss extensively. According to another model of memory, retrieval is enhanced to the extent that learning and retrieval occur under similar conditions. This bodes poorly for advertisers given that the encoding and storage of advertising rarely occur under the same

conditions as the actual buying behavior. Advertisements tend to be absorbed most everywhere except the marketplace.

### 4.3. Processing Pictures Versus Text

Pictures enhance advertising effectiveness in several ways: They help to get the audience's attention, they provide information about the brand and product use, and they help to create a unique brand image. Because people examine the visual elements before the verbal elements, pictures in advertising are a useful tool as attention grabbers. For example, extensive use of sex appeal and celebrity endorsement in advertising can be understood in this context.

Images and text are processed differently. Pictorial information seems to be processed more holistically, whereas verbal information is processed more sequentially. Information conveyed by pictures is recalled and recognized more easily than is textual information. Therefore, the fact that more than two-thirds of print advertisements today have pictures covering more than half of the available space is not surprising. So, why are pictures more memorable than words?

According to Paivio's dual-coding model, different codes exist in memory for verbal material and pictorial material. Pictures tend to be remembered over verbal information because they activate both verbal codes and pictorial codes spontaneously. This ease of cognitive access leads to the facilitation of memory. In addition, ads with concrete, easily identifiable pictures (e.g., celebrities) or realistic pictures are more memorable than ads with abstract or unidentifiable pictures. Furthermore, consumers exhibit more favorable attitudes when they see identifiable objects (e.g., an Adidas logo, a picture of Michael Jordan) than when they simply view text (e.g., the combination of alphabetical letters P-U-M-A).

Memory effects of text vary depending on how difficult it is for perceivers to comprehend the text. When the text is complex, people engage in more elaborative processing, resulting in a greater number of associations in memory as well as improved memory. Finally, pictures and text interact. Congruency between pictures and ad copy enhances consumers' recall.

### 4.4. Repetition and Learning

Advertising repetition is generally known to enhance memory by strengthening memory traces because it increases redundancy and provides more encoding opportunities to process the message, leading to higher

levels of brand name recall. Nevertheless, advertising repetition rarely works in such simple ways. For example, varying the interval between message repetitions affects memory of an advertising message. Although not without exceptions, memory for repeated material generally improves as the time between presentations of advertising material increases, particularly when there is a delay between the subsequent presentation of the stimulus and the memory test.

Another explanation of learning through repetition is derived from encoding variability theory, which predicts that presenting a series of ads containing slight variations of a theme enhances memory for the ad material. For example, according to this view, repeated presentation of a bottle of Absolut Vodka in different contexts helps consumers to retain its brand name in their memory.

Attention, recall, and brand awareness initially increase, then level off, and ultimately decline as the number of exposures increases. At least two explanations are available for this "wear-out" effect: inattention and active information processing. With increasing repetition, viewers no longer attend to the message, and this inattention causes forgetting. According to the active information processing perspective, an audience rehearses two kinds of thoughts: thoughts stimulated by the message reflecting message content (i.e., message-related thoughts) and other thoughts based on associations reflecting previous experiences (i.e., people's own thoughts). With the initial exposure to a message, people's thoughts tend to be message related, but at some level of repetition, people's thoughts stem mainly from associations that are only indirectly linked to the message. These thoughts are less positive toward the product than are message-related thoughts, primarily because the latter were selected to be highly positive.

### 4.5. Low- and High-Involvement Learning

Learning may occur either in a situation where consumers are highly motivated to process the advertising material or in a situation where consumers have little motivation to learn the material. For example, a woman reading an automobile magazine in a dentist's office is less involved than another woman reading the same magazine prior to purchasing a new car. Personal, product, and situational factors jointly affect the level of involvement. A different level of involvement is likely to follow, for example, depending on whether the individual perceives the advertised product as enhancing

self-image (i.e., personal factor) or entailing risk (i.e., product factor). And this whole perception of the product will again be influenced by whether the consumer views the product for personal use or views it as a gift (i.e., situational factor).

High involvement stimulates semantic processing, whereas low involvement is linked to sensory processing. Text-based information is better remembered when viewers are highly involved, whereas graphic-oriented information exerts a greater impact on viewers' memory when they have low involvement. The study of learning and memory is important for understanding how consumers obtain information about a product or service. Another primary goal of advertising is to persuade, so the next section considers the psychological study of attitudes and persuasion.

## 5. ATTITUDES

The attitude construct has been recognized as one of the most indispensable concepts in psychology. It is similarly crucial in the study of advertising. But what exactly is an attitude? For a psychologist or an advertiser, an attitude refers to an evaluation along a positive–negative continuum. For purposes of this article, an attitude is defined as a psychological tendency to evaluate an object with some degree of favor or disfavor. The target of an attitude is called the attitude object. When an individual says, “This pie is lovely” or “That car is no good at all,” that person is expressing his or her attitudes. When the individual says, “This pie is fattening” or “That car has poor acceleration,” the person is expressing his or her beliefs. Attitudes are different from beliefs, although the latter help to make up the former. Beliefs are units of information that an individual has. Beliefs may be facts or opinion, and they may be positive, negative, or neutral with regard to the target.

### 5.1. The Structure of Attitudes

Attitudes are generally not thought of as monolithic constructs; they are made up of conceptually and empirically distinct components. At a very basic level of analysis, attitudes have three important components: affective, behavioral, and cognitive. Affect refers to feelings and emotional components of attitudes. Behavior, of course, refers to behavior that an individual takes with regard to a target. Cognition refers to the beliefs or thoughts that an individual has about a target.

Affective, behavioral, and cognitive processes help to form attitudes. The mere exposure effect suggests one way in which positive affect may arise. Classical conditioning and operant conditioning are two additional ways in which affective processes influence people's attitudes. The continuous pairing of some stimulus and a reward (or punishment) creates positive (or negative) affect. In advertising, brands of clothing are nearly always paired with attractive models. The models are intended to create affectively positive feelings, and advertisers hope that people will be conditioned to like their brands of clothing.

Behavior also contributes to the formation of attitudes in that sometimes people infer their attitudes on the basis of their previous behavior. Self-perception theory posits that people infer their attitudes on the basis of their past behavior, particularly when they believe that their behavior has been freely chosen. For example, if someone points out that Jane always wears green, she may infer that she has some affinity for green. But if Jane always wears green because her school has a strict dress code requiring her to wear green, she is unlikely to infer that she has a favorable attitude toward green.

Cognition is another important antecedent of people's attitudes. A cognitively based learning process occurs when people acquire information about attitude objects. People may gain information through direct experience such as when a free trial product is sent in the mail or when a free sample is offered in a store. Or, people may gain information indirectly, for example, when a television commercial shows them the benefits of owning a particular make and model of automobile. People's beliefs about attitude objects have been proposed as a central determinant of attitudes. Indirect learning, or observational learning, is an important tool for advertisers. Consider any advertisement that shows a model using a product to benefit in some way. It is hoped that viewers will develop favorable attitudes toward the product by learning how others have benefited from the product.

So, attitudes generally have affective, behavioral, and cognitive components. However, it is not necessary for all attitudes to have all three components. Some attitudes may be based primarily on affective factors (e.g., attitudes toward tequila), and others may be based primarily on cognitive factors (e.g., most people probably feel mildly positive about photosynthesis due to the important functions performed by the process, but they probably do not have strong emotions about it).

One very influential model of the structure of attitudes is Martin Fishbein's expectancy-value model.

Fishbein proposed that attitudes are a multiplicative function of two things: (a) the beliefs that an individual holds about a particular attitude object and (b) the evaluation of each belief. According to the expectancy-value model, beliefs are represented as the subjective probability that the object has a particular attribute. The model can be expressed as a mathematical function:

$$A_o = \sum_{i=1}^n b_i e_i,$$

where  $A_o$  is the attitude toward the object,  $b_i$  is a belief about the object, and  $e_i$  is the evaluation of that belief. According to Fishbein, people's attitudes are typically based on five to nine salient beliefs. So, if a researcher wanted to know someone's attitude toward a particular brand of clothing, the researcher might ask that person to estimate the likelihood that a particular brand has a variety of attributes (e.g., fashionable, durable, well priced) and how positive or negative each of those attributes is. The researcher could then compute an estimate of the person's attitude by multiplying the pairs of scores and then summing the products.

The expectancy-value model also implies that persuasion is largely a function of message content. That is, favorable attitudes can be produced by making people believe that an object is very likely to have some desirable trait, by making people believe that some trait is very favorable, or by both. For example, an advertiser might endeavor to make people believe that its automobile is very reliable (i.e., influence the subjective probability of beliefs) or to make people believe that its automobile's ability to take turns at very high speeds is highly desirable (i.e., influence the evaluation of a particular attribute).

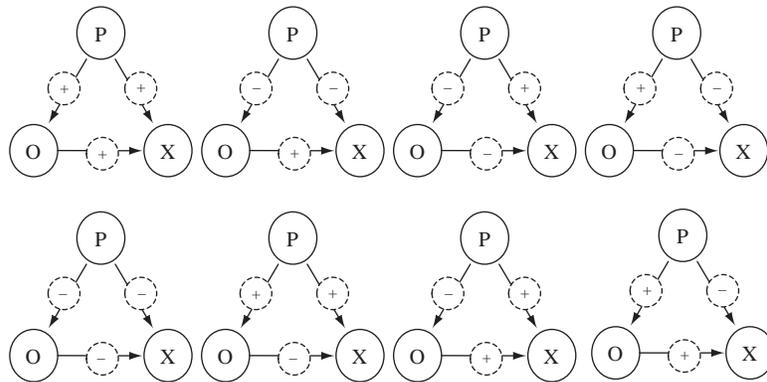
Although the expectancy-value model seems to be perfectly logical, it may seem surprising to suggest that all attitudes are based on a series of beliefs. Consider, for example, the mere exposure research discussed earlier. According to Zajonc, preferences need no inferences (i.e., people may like something without having any beliefs about it). Under some conditions, attitudes may be formed outside of people's conscious awareness, or attitudes may be directly retrieved from memory rather than "computed" based on a mental review of salient beliefs. However, it is generally accepted that highly elaborated attitudes are more influential than poorly elaborated attitudes. So far, the discussion of attitude structure has considered how different aspects of a single attitude relate to one another. Next, the

discussion considers how different attitudes relate to one another.

One of the most enduring psychological principles developed during the 20th century is the simple notion that people have a desire for cognitive consistency. Cognitive consistency is the simple notion that beliefs and actions should be logically harmonious. If an individual believes that cats make good pets but hates her pet cat, she has beliefs that are inconsistent; if an individual believes that cats make good pets and she loves her cat, she has beliefs that are consistent. For most people, cognitive inconsistency is unpleasant, so they take steps to achieve consistency.

One consistency theory with many advertising-related applications is Heider's balance theory. Balance theory was initially applied to cognitive consistency between dyads (two units) and among triads (three units), but because most research has examined triads, this article focuses on this arrangement. The triad arrangement pertains to the attitudinal relationships among a perceiver ( $p$ ), an other ( $o$ ), and an attitude object ( $x$ ). Consider the example where Cody has recently met an individual named Sam, and Cody likes Sam quite a bit. One afternoon, Cody learns that Sam loves to listen to country music. However, Cody cannot stand country music. How does the fact that Sam loves country music make Cody feel? Probably not very good; the triad of Cody, Sam, and country music is not balanced. However, if Cody loved country music, Cody liked Sam, and Sam loved country music, all would be simpatico. These ideas illustrate the basic tenets of Heider's balance theory. As can be seen in Fig. 1, there are eight possible sets of relationships among the triads: four balanced and four imbalanced. One simple way in which to identify whether a triad is balanced or not is to calculate the product of the three relationships. If the product is positive, the triad is balanced; if the product is negative, the triad is imbalanced.

The efficacy of well-liked, or celebrity, endorsers may be explained at least in part by evoking balance theory. The viewers of the advertisement are expected to have a favorable attitude toward the endorser (e.g., Britney Spears), and the endorser is clearly portrayed as having a positive attitude toward the advertised product (e.g., cola). To maintain balance, viewers also should adopt a positive attitude toward the cola. Alternatively, viewers could decide to dislike the cola and change their attitude toward Spears so as to maintain a balanced triad. Balance theory also helps to explain one way in which consumer trends migrate. People who become friends with one another often adopt attitudes similar to their friends'



**FIGURE 1** Eight possible triads proposed by Heider's balance theory. The triads in the top row are balanced, and the triads in the bottom row are imbalanced.

attitudes. A classic study by Theodore Newcomb illustrated this point with women who lived together at college; over time, the women's political attitudes became more and more similar.

Another theory that has roots in cognitive consistency, and has been very influential in advertising and consumer behavior, is cognitive dissonance theory. Leon Festinger proposed cognitive dissonance theory in 1957, and it spurred more research than perhaps any other social psychological theory. Cognitive dissonance has been defined as a feeling of discomfort that arises as a result of one's awareness of holding two or more inconsistent cognitions. Often, dissonance is aroused when one behaves in a manner that is inconsistent with his or her beliefs. For example, Greg may believe that Japanese cars are superior to cars made in America, but if he buys an American car, he will likely experience cognitive dissonance. Because cognitive dissonance is uncomfortable, people are motivated to reduce the feeling of dissonance by changing their behavior, trying to justify their behavior by changing their beliefs, or trying to justify their behavior by adding new beliefs. Having purchased an American car, Greg might try to reduce cognitive dissonance by investing in Japanese auto manufacturers, by changing his belief in the superiority of Japanese cars, or by adding a new belief to help regain consistency, for example, "My car may be American, but many of the engine parts are from Japan."

When people make large-scale purchases, they often experience what is known as postdecisional dissonance. Large expenditures may arouse dissonance because they are inconsistent with the need to save money or make other purchases. Furthermore, making a purchase decision necessarily means giving up some attractive features on the unchosen alternatives (e.g., buying a Sony means

not buying a Samsung). In a decision-making context, dissonance may be reduced by revoking the decision, by bolstering the attractiveness of the chosen alternative or undermining the attractiveness of the unchosen alternative, or by minimizing the differences between or among the alternatives. Another important role that advertising can play is in helping to reduce postdecisional dissonance. Advertising can help to reduce the feeling of discomfort that follows a major purchase by changing beliefs (e.g., "The new MP3 player has state-of-the-art technology") or by adding new beliefs (e.g., "The new MP3 player will make you the envy of your friends") that enable buyers to feel good about their recent major purchases.

## 5.2. Functions of Attitudes

Why do people have attitudes? Consider, for a moment, what life would be like without attitudes. For example, Tracy might sit down to dinner one day to find a plate of lima beans. She might eat the beans and discover that they taste horrible. The following week, Tracy is once again served lima beans and remembers them from the last time, but she has no attitude toward them. She eats them again and finds out, once again, that they taste quite horrible. For someone without attitudes, the discovery that lima beans are horrible could occur hundreds of separate times over a lifetime. At a most basic level, attitudes help people to navigate their world; they help them to know how to respond to things. Attitudes allow people to approach rewards and avoid punishments. Beyond this basic "object appraisal" function, attitudes have long been thought to serve a number of important functions.

More than 40 years ago, Daniel Katz described four functions of attitudes: ego-defensive, value-expressive,

knowledge, and utilitarian. Some attitudes have an ego-defensive function in that they help protect people from unflattering truths about themselves. People may bolster their own egos by holding negative attitudes about other groups (e.g., Hispanics, homosexuals). The value-expressive function occurs when attitudes allow people to express important values about themselves. For example, people may express a positive attitude toward recycling, suggesting that environmentalism is an important value. The knowledge function of attitudes allows people to better understand the world around them. For example, if a person dislikes politicians, it is easy to understand why politicians always seem to be giving themselves pay raises when the economy is particularly weak. Finally, attitudes may have a utilitarian function. Tracy's attitude toward lima beans helps her to know whether to approach them (because they are delicious) or avoid them (because they taste horrible). As suggested by all of these functions, attitudes also help to guide people's behavior.

### 5.3. Attitude–Behavior Relations

Indeed, the study of attitudes was at least partly initiated because attitudes seemed to be a logical predictor of behavior. As noted earlier in the article, behavior is an important form of attitudinal expression. For example, if Tara has a positive attitude toward a brand of pants, it follows that she will buy the pants; she will behave in a way that is consistent with her attitudes. Advertisers aim to create positive attitudes toward objects in the hope that consumers will purchase those objects. Unfortunately, the study of attitude–behavior relations has not been quite so simple. The following is only a very brief review of attitude–behavior relations.

A 1969 review of the literature on attitude–behavior relations found that attitudes and behaviors were modestly correlated with one another at best. This lack of attitude–behavior correspondence caused some researchers to suggest abandoning the attitude concept altogether. Fortunately, others rejected this suggestion and worked to better understand attitude–behavior relations. It is now known that attitudes reliably predict behavior under certain conditions. Attitudes and behaviors correlate when the attitude measures and behaviors correspond with regard to their level of specificity. If one wants to reliably predict a specific action, one should assess attitudes toward performing that action, with regard to a particular target, in a given context, and at a specific time. Or, one could enhance attitude–behavior correspondence by broadening the scope of

the behaviors. Knowing someone's attitude toward a particular attitude object (e.g., religion) is not necessarily going to predict whether that person attends church on a given Sunday, but it should reliably predict a variety of religious behaviors over the course of time (e.g., attending religious services over a period of weeks, having a religious text at home, wearing a symbol of religious faith).

Another way in which to enhance the attitude–behavior relationship is to provide people with direct experience with attitude objects. Indirect learning is not as powerful as direct learning, so an advertisement showing people enjoying a frosty beverage is not as persuasive as having people enjoy the frosty beverage in person. Behavioral prediction is enhanced when one also considers the influences of social norms and perceived control on behavior. Finally, attitudes are more predictive of behavior when attitudes are strong or readily accessible.

Attitude strength has been defined and measured in a variety of ways, but it is generally accepted that strong attitudes are resistant to change, persistent over time, and predictive of behavior. Clearly, these are the types of attitudes that advertisers want to inculcate. But given that there are multiple definitions of attitude strength, what is the best way in which to identify and produce strong attitudes? In considering various measures of attitude strength, it may be helpful to think in terms of whether the measures are operative or meta-attitudinal. An operative measure of attitude strength is one that reflects the operation of the attitude (e.g., the accessibility of the attitude can be measured by the speed with which people can provide evaluations of objects), whereas a meta-attitudinal measure of attitude strength is one that requires people to provide a subjective self-report of their own attitudes (e.g., the confidence with which an attitude is held cannot be directly measured but rather must be reported by an individual). One way in which attitudes can be strengthened is through cognitive elaboration. To the extent that people intentionally and carefully think about, and elaborate on, their attitudes, they are engaged in cognitive elaboration.

## 6. PERSUASION

The study of attitude change has existed since Aristotle's time. However, empirical research on persuasion and attitude change is a much more recent phenomenon. It was not until the mid-20th century that the study of attitude change developed into a thoroughly systematic process.

### 6.1. Message Learning Approach

Carl Hovland and colleagues, working at Yale University during the 1950s, sought to study persuasion by considering the question, “Who says what to whom with what effect?” That is, they were interested in studying the effects of different variables in the persuasion process. “Who” refers to the source of the persuasive communication, “what” refers to the message that is presented, and “to whom” refers to characteristics of the message recipient. This approach to the study of persuasion was an information-processing, or message-learning, paradigm. According to the message-learning paradigm, persuasive communications could have an effect only to the extent that they commanded attention and were comprehensible. Furthermore, message recipients had to yield to the persuasive communications and retain the information presented in the persuasive communications. If these conditions were met, beliefs, attitudes, and behaviors were liable to change. The following considers some of the critical variables studied by the Yale group.

Holding message, recipient, and channel characteristics constant, how do different characteristics of the source influence persuasion? Hovland and colleagues found that communicator credibility had an effect on persuasion, such that credible sources were more persuasive than noncredible sources. But what exactly is a credible source? Hovland and colleagues examined three characteristics of credible sources: expertise, trustworthiness, and the source’s intent to persuade. An expert source (e.g., a Nobel prize-winning scientist) is more persuasive than a nonexpert source (e.g., the director of a local YMCA). A trustworthy source (e.g., a news anchor) is more persuasive than a nontrustworthy source (e.g., a member of the Liars Club). A source that is known to have a persuasive intent (e.g., an advertisement) is often less persuasive than one delivering the same message but with no persuasive intent (e.g., a friend). Forewarning people of a communicator’s persuasive intent seems to instigate mental counterarguing in the audience. It is also known that physically attractive sources are generally more persuasive than unattractive ones and that similar communicators are usually more persuasive than dissimilar ones. Finally, powerful communicators (i.e., communicators who can administer punishments or rewards to the message recipients) tend to be more persuasive than powerless communicators.

Holding source, recipient, and channel characteristics constant, how do different characteristics of the message influence persuasion? The comprehensibility of a message is (obviously) an important determinant of persuasion; if

people cannot understand the message, it is unlikely that they will be persuaded by it. The number of arguments in a persuasive communication also matters; more arguments are generally better than fewer arguments. Of course, there is a limit to the number of arguments one can present before message recipients become annoyed and lose interest. Presenting a few strong convincing arguments is better than presenting dozens of weak specious arguments. Messages that arouse fear (e.g., “If you do not brush your teeth, you will end up ugly, toothless, and diseased”) can also be persuasive if certain conditions are met. An effective fear appeal must (a) convince the recipients that dire consequences are possible, (b) convince the recipients that the dire consequences will occur if instructions are not followed, and (c) provide strong assurance that the recommended course of action will prevent the dire consequences. One- and two-sided messages are differentially persuasive for different audiences; two-sided messages are generally more effective among knowledgeable audiences, whereas one-sided messages are more effective among less knowledgeable audiences. Two-sided messages can be more effective in general so long as the opposing arguments are effectively countered in the message.

Recipient characteristics also influence the efficacy of persuasive communications. An audience that is highly motivated and able to process information is more apt to be persuaded (provided that strong, rather than weak, arguments are presented) than is an audience that is distracted or apathetic. People who are of lower intelligence are generally easier to persuade than are people of high intelligence, people with moderate levels of self-esteem are generally easier to persuade than are people with either low or high self-esteem, and younger people are more susceptible to persuasive communications than are older people. Some personality traits also have important implications for persuasion. Self-monitoring is a characteristic that varies in the population. High self-monitors are particularly sensitive to situational cues and adjust their behavior accordingly, whereas low self-monitors are guided more by internal cues and tend to behave similarly across various situations. High self-monitors tend to be susceptible to persuasive communications that have image-based appeals, whereas low self-monitors tend to be more susceptible to value or quality-based appeals.

### 6.2. Dual-Process Theories

The message-learning approach to persuasion obtained a great deal of information about the influence of different

variables on persuasion. However, there were also a lot of apparently contradictory findings. One study might report that attractive sources had a large impact on persuasion, whereas another study would report no impact at all. During the late 1970s, a pair of integrative frameworks for understanding persuasion emerged. The Elaboration Likelihood Model (ELM) and the Heuristic-Systematic Model (HSM) are both dual-process models of persuasion, emphasizing that different variables can have different effects on persuasion. These models are fairly similar to one another and can be used to explain similar findings and make similar predictions, but the language used in describing the two models is different. The focus here is on the ELM because it has spurred the most research in advertising and marketing.

The ELM is based on the proposition that there is a continuum of elaboration likelihood. At one end of the continuum, the amount of cognitive effort that is used to scrutinize persuasive communications is negligible; information processing at this end of the continuum is minimal. This is known as the peripheral route to persuasion. At the other end of the continuum, there is a great deal of cognitive elaboration; people are highly motivated to carefully process information pertaining to the persuasive communications. This is known as the central route to persuasion. One might say that there are two routes to persuasion: central and peripheral. Which route is taken is dependent on message recipients' motivation and ability to process information. According to the ELM, people who are both motivated and able to process persuasive communications will take the central route, whereas people who are lacking either motivation or ability to process information will take the peripheral route.

The central and peripheral routes are metaphors for the amount of cognitive elaboration in which people engage when they are faced with persuasive communications. Central route processing is particularly likely to occur when the persuasive communications are personally relevant to the recipient. If an individual is in the market for a new automobile, automobile advertisements would likely be personally relevant and, therefore, more highly scrutinized. If a person happens to be a home theater enthusiast, he or she would be more motivated to attend to advertisements for televisions and stereo speakers. Of course, all of the motivation in the world cannot cause greater cognitive elaboration if the message recipients do not also have the ability to process the information carefully. Distractions and other

forms of cognitive business would prevent cognitive elaboration.

Persuasion through the central route occurs when the arguments presented are strong and compelling, and attitudes formed through the central route are persistent over time and resistant to change. Persuasion through the peripheral route occurs when there are compelling peripheral cues present (e.g., a long list of arguments, an attractive speaker, a credible source), and attitudes formed through the peripheral route are more temporary and subject to further change. The ELM and HSM can account for a wide variety of persuasion phenomena and have proven to be very robust and useful theories.

The psychology of advertising has come a long way. We now have a greater understanding than ever before of psychological processes crucial to advertising. From attracting attention to understanding how people are persuaded, the advancement of psychological science and that of advertising practice have much to learn from one another. This article has but scratched the surface of advertising psychology. New theories and applications are emerging at a rate that is unparalleled in the history of either advertising or psychology.

### See Also the Following Articles

Attitudes ■ Environmental Versus Individual Risk Taking: Perception, Decision, Behavior ■ Learning ■ Perception and Culture

### Further Reading

- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.
- O'Guinn, T. C., Allen, C. T., & Semenik, R. J. (2003). *Advertising and integrated brand promotion*. Mason, OH: Thomson South-Western.
- Packard, V. (1957). *The hidden persuaders*. New York: D. McKay.
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 4, pp. 323-390). New York: McGraw-Hill.
- Scott, L. M., & Rajeev, B. (Eds.). (2003). *Persuasive imagery: A consumer response perspective*. Mahwah, NJ: Lawrence Erlbaum.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.



# Affirmative Action

David A. Kravitz

George Mason University, Fairfax, Virginia, USA

1. What Is Affirmative Action?
  2. Economic Effects of Affirmative Action
  3. Affirmative Action and Stigmatization
  4. Attitudes toward Affirmative Action
  5. Conclusions
- Further Reading

## GLOSSARY

**affirmative action** Any measure, beyond a simple termination of discriminatory practice, adopted to correct for past or present discrimination or to prevent discrimination from recurring in the future.

**affirmative action plan (AAP)** A set of specific and results-oriented procedures designed to eliminate discrimination and enhance opportunities for members of an underrepresented group.

**affirmative action plan (AAP) strength** The weight given by an AAP to demographic status.

**discounting principle** In affirmative action, the tendency to derogate the competence of target group individuals who have been selected in the context of an affirmative action plan.

**preferential treatment** The provision of positive weight to minority status in an affirmative action plan; in weak preferential treatment (WPT), a member of the target group is preferred only if his or her qualifications are equivalent to those of a competing majority individual, whereas in strong preferential treatment (SPT), a member of the target group is preferred even if his or her qualifications are inferior to those of a competing majority individual.

**principled conservatism theory** The theory that opposition to affirmative action is not due to racial prejudice but rather is due to the inconsistency between affirmative action and conservative principles.

**racioethnicity** A term that refers to the combination of race and ethnicity that is used to distinguish among groups in affirmative action regulations.

**racism** The belief that a particular race is superior or inferior to other races.

**social dominance orientation (SDO)** A person's general approval of either dominance or egalitarian relationships among groups; individuals who are high in social dominance orientation support group-based inequality.

**stigmatization** In the context of affirmative action, the derogation of the competence or qualifications of an individual who is believed to have been selected on the basis of gender or racioethnicity.

**utilization analysis** An analysis that compares an organization's workforce (workforce analysis) with the availability of qualified individuals (availability analysis) to determine whether women or racioethnic minority groups are underutilized (i.e., whether there are fewer minorities or women in a particular job group than would reasonably be expected by their availability).

Affirmative action refers to any measure, beyond a simple termination of discriminatory practice, adopted to correct for past or present discrimination or to prevent discrimination from recurring in the future. This controversial social policy has inspired research in

a variety of disciplines. The present article summarizes research on (a) the economic effects of affirmative action on target groups and on organizations, (b) stigmatization of affirmative action target group members by others and by the individuals themselves, and (c) attitudes toward affirmative action.

## 1. WHAT IS AFFIRMATIVE ACTION?

### 1.1. Background

Humans have a propensity to distinguish between the in-group and the out-group, a process that often results in prejudice and discrimination toward individuals in the out-group. Consistent with this tendency, most countries have long histories of discrimination toward women and racial/ethnic minorities. During recent decades, many countries have passed antidiscrimination laws, and some have gone further to require that affirmative actions be taken to enhance opportunities for traditionally underrepresented groups. Although affirmative action may apply to university admissions, this article is restricted to affirmative action in the workplace. It summarizes the economic, behavioral, and psychological research on affirmative action that has been published in English. Most of this work has been conducted in the United States or Canada.

### 1.2. Definitions

What is affirmative action? That question can be answered conceptually, legally, and operationally.

#### 1.2.1. Conceptual Definition

Many conceptual definitions have been offered. For example, the U.S. Civil Rights Commission has defined affirmative action as “a term that in a broad sense encompasses any measure, beyond a simple termination of discriminatory practice, adopted to correct for past or present discrimination or to prevent discrimination [from] recurring in the future.”

#### 1.2.2. Legal Definitions

The legal definitions of affirmative action vary from place to place and change over time. Critical dimensions on which the law varies include which groups are targeted, which actions are required and forbidden, what penalties can be imposed for violations, and

how the law is enforced. In the United States, federal affirmative action law is jointly determined by the U. S. Constitution (5th and 14th amendments), presidential executive orders (e.g., Executive Order 11246 [EO 11246]), legislation (e.g., Civil Rights Acts of 1964 and 1991), implementing regulations promulgated by federal agencies (e.g., Office of Federal Contract Compliance Programs [OFCCP], Equal Employment Opportunity Commission [EEOC]), and court decisions.

Organizations in the United States may be required to establish affirmative action plans (AAPs) for four reasons. First, courts may impose AAPs on organizations that have been convicted of discrimination. Second, to avoid adverse legal decisions, organizations may set up AAPs as required by court-approved consent decrees. Third, EO 11246 and other acts require affirmative action of federal agencies and of most organizations that do business with the government (i.e., federal contractors). Fourth, government “set-aside” programs at the federal, state, and local levels require that some proportion of contracting and procurement expenditures go to designated firms. In addition to these legal bases for affirmative action, some organizations may voluntarily take affirmative actions to diversify their workforces. All AAPs are constrained by the Constitution and by the Civil Rights Acts.

It is not possible to describe here the requirements and constraints of affirmative action in detail. However, because most AAPs in the United States are established by federal contractors as required by EO 11246, some of their key requirements and constraints are described here. First, the organization must not discriminate on the basis of race, color, religion, sex, or national origin. Second, the organization must complete an annual utilization analysis that compares its workforce, in terms of gender and racial/ethnicity, with the available qualified workforce. Third, if underutilization of women or a racial/ethnic minority group is revealed, the organization must generate a plan for eliminating the underutilization. Fourth, the plan should use targeted recruitment, outreach, and other nonpreferential actions that will facilitate target group success. Fifth, the plan must not involve the use of demographic status (i.e., racial/ethnicity, gender) in making selection, promotion, or termination decisions. This last point is important: Preferential treatment of women and racial/ethnic minorities is contrary to EO 11246. Preferences are legal only when they are specifically required by a court order or consent decree to reverse the effects of proven discrimination.

### 1.2.3. Operational Definitions

For current purposes, we must consider two types of operational definitions: organizational implementation of affirmative action and the descriptions of affirmative action used in behavioral research. Organizational AAPs are complex. They potentially include organizational structures, written reports, proscribed and prescribed processes, and more. Although “best practice” reports exist, the author has never seen a report summarizing the frequencies of typical practices. Importantly, although much of the public equates affirmative action with preferential treatment, the actual frequency of such illegal procedures is unknown. Operationalizations used in behavioral research tend to be brief, simplistic, and (often) illegally strong.

### 1.3. Public Beliefs

The American public construes affirmative action in many different ways. Some people focus on the role of affirmative action in decreasing discrimination against women and ethnic minorities. Other people believe it involves quotas or preferential treatment. Very few people think of affirmative action in terms of recruitment and outreach, and virtually nobody has a realistic understanding of the law. Beliefs about what the term “affirmative action” means vary with respondent characteristics, such as race/ethnicity and political orientation, and predict support or opposition to affirmative action.

## 2. ECONOMIC EFFECTS OF AFFIRMATIVE ACTION

There has been a lively debate regarding the economic effects of affirmative action on the targeted groups and on organizations subject to affirmative action regulations. The implicit question is whether the benefits of affirmative action outweigh the costs. That question cannot be answered in the abstract because it depends both on the effects of affirmative action and on the values one assigns to organizational outcomes and to the well-being of the target groups. Although such values are a personal matter, it is possible to study the effects of affirmative action. Both conceptual and empirical studies have been published, and the conclusions are summarized in what follows. However, it is important to acknowledge some unavoidable methodological challenges. Causal conclusions are possible only when the experimenter can randomly assign

experimental units to treatment conditions while holding constant all other potentially relevant factors. That is impossible when dealing with a public policy such as affirmative action, where assignment to “conditions” is determined by the law and by the experimental units (organizations and target groups) themselves.

### 2.1. Target Groups

Many studies assess the employment status of White women and race/ethnic minorities, vis-à-vis White men, in federal contractors versus noncontractors while controlling for relevant variables such as the type of industry. A positive effect of affirmative action is inferred if the relative outcomes of target groups are better in federal contractors than in noncontractors. This research indicates that affirmative action has had a positive effect on the employment status of White women and race/ethnic minority groups, but the results are complex and some are tentative. At a minimum, the effects have varied across target groups, locations, time, and types of positions. The initial effect of affirmative action was substantial for African Americans and slightly positive for White women. The limited work on Hispanic Americans suggests that they were also helped. The clearest effects were on employment, especially in professional positions, although wages may also have been affected. However, this conclusion must be qualified by effects of region and time. The largest initial impact of affirmative action (and of antidiscrimination law in general) was observed in the South, where the extent of discrimination was most extreme and so there was the greatest room for improvement. In addition, changes were most substantial during the decade from 1965 to 1975. Broad progress stagnated during the 1980s, especially for African Americans, although there is some indication of continuing effects on the status of White women.

Other studies use micro-level employer data drawn from a variety of sources, including the organizations themselves. This work has found that target groups profit more from affirmative action than from passive equal opportunity practices. The effects of affirmative action vary across target groups, locations, industries, implementation procedures, and other factors.

A final indication of the effectiveness of affirmative action is the finding that growth in employment of White women and African Americans within organizations is predicted by organizational AAP goals. Although the relation is positive, the actual growth in employment of these groups is just a fraction of the

goals. This demonstrates that affirmative action goals do not function as strict quotas.

## 2.2. Organizations

The public debate surrounding affirmative action often centers on the presumed consequences of affirmative action policies for organizations. Critics claim that AAPs adversely affect performance by (a) forcing organizations to expend resources on implementation, (b) constraining organizational decisions, and (c) forcing organizations to hire underqualified individuals. Supporters of affirmative action argue that AAPs improve organizational performance by (a) motivating both White male and other employees and (b) providing organizations with the resources represented by a diverse workforce.

Theoretical models show that whether the net effect of affirmative action on organizations is positive or negative should depend on several factors, including the extent of preexisting discrimination, exactly how the AAP is implemented, and differences in human capital resources of majority and minority groups. Preexisting discrimination would hinder firm performance by limiting access to the resources represented by the victimized groups. If affirmative action decreased this discrimination, the effect on firm performance would be positive. On the other hand, if discrimination were already absent, the constraints on behavior posed by affirmative action could have a negative effect. Similarly, AAPs such as focused recruitment that serve to increase organizational access to human resources should have a positive effect on performance, whereas preferential selection should have a negative effect. Finally, the magnitude of the negative effect of preferential selection should increase with the size of the human capital gap between the majority and minority applicant groups.

As is clear from the preceding, there are logical arguments for both positive and negative effects of affirmative action on organizations. Thus, we turn to the empirical literature. The strongest research has included a wide variety of organizations and has controlled for industry, job level, and other influential variables. The paper qualifications (e.g., education) of minorities and White women appear to be somewhat lower than those of White males in firms with AAPs, particularly when affirmative action is used in hiring rather than when it is used only in recruiting. (The use of affirmative action in hiring does not necessarily imply preferential selection. Instead, it can involve actions such as the use of different screening mechanisms and the insistence that each

set of interviews must include one person who belongs to an underrepresented group. Many organizations reported that affirmative action was involved in hiring processes that resulted in the selection of White males.) Importantly, the minor decrements in educational qualifications are not associated with decrements in performance. Several explanations are offered for this apparent inconsistency, but all relate to the fact that firms that use affirmative action tend to have superior human resources practices. For example, virtually all AAPs include extensive recruitment, and this increases the overall quality of the labor pool. In addition, these organizations tend to use selection procedures that are unusually comprehensive and valid. Their use of comprehensive selection systems enables them to detect high-potential minority applicants whose competence is greater than their education would suggest. The null effect of affirmative action on organizational performance that has been observed in the multi-industry research has been replicated in the few studies that have assessed a single type of organization (e.g., police departments).

## 3. AFFIRMATIVE ACTION AND STIGMATIZATION

Concerns about stigmatization have been raised by judges rendering decisions in affirmative action cases, by Black and Hispanic intellectuals, and by conservative opponents of affirmative action. Empirical research has addressed several questions, of which three have received the lion's share of attention. Are members of affirmative action target groups stigmatized by others? Do members of affirmative action target groups expect to be stigmatized by others, and if so, what are their reactions? Do affirmative action beneficiaries engage in self-stigmatization? Work in this area draws on concepts from stigma theory and from attribution theory, particularly the principle of discounting.

According to the discounting principle, an individual's confidence regarding a possible cause of some outcome is affected by the presence or absence of plausible alternative causes. For example, if a woman is hired by a firm that has an AAP that targets women, she could have been selected either because she is highly qualified or to help the firm meet its affirmative action goal. In the absence of affirmative action, the only remaining explanation is her qualifications. Thus, a woman hired in the context of an AAP will be

evaluated less positively than a woman hired in the absence of affirmative action or less positively than a man. This reasoning applies both to observers and to the selected individual herself.

Note that the logic of discounting requires the assumption that affirmative action involves preferences. Also note that this reasoning assumes there are only two plausible reasons for an individual's selection—qualifications or affirmative action—and that a White male's selection must necessarily be based on his qualifications. This ignores factors such as traditional gender or racial discrimination, nepotism, personal friendship, similar backgrounds, and the myriad of other factors that are unrelated to merit but nonetheless affect employment decisions.

### 3.1. Stigmatization by Others

In the prototypical stigmatization experiment, the participant is given information about the background and performance of an employee who is either male or female and, if female, either is or is not an affirmative action hire. The affirmative action condition nearly always involves absolute preferences. The participant is asked to evaluate the employee on competence and other dimensions. Comparisons among the three conditions—male, female, affirmative action female employee—provide the key tests of stigmatization. In a second approach, the male participant believes he is working with a female leader on a one-way communication task, with the leadership role having been assigned on the basis of either test performance or gender. Again, the affirmative action condition involves explicit preferences. After successful or unsuccessful completion of the task, the participant is asked to evaluate his partner.

Consistent with the discounting principle, the affirmative action female is stigmatized. In the scenario research, she is evaluated less positively than either the male or the non-affirmative action female. In the communication research, she is evaluated less positively when selected on the basis of gender than when selected on the basis of test scores. Although the target is usually a White female, a few studies have included racioethnic minority targets. The stigmatization effect generalizes to these targets.

Having established the existence of stigmatization, subsequent research has concentrated on determining its limitations. Key conclusions include the following. First, stigmatization can be eliminated by providing incontrovertible information regarding the employee's competence or successful performance on the job.

Second, ambiguous information about employee success does not eliminate stigmatization. Third, stigmatization occurs when affirmative action is explicitly described as involving preferences and also when it is undefined. The latter finding is consistent with the common assumption that affirmative action involves preferences. Fourth, stigmatization does not occur when affirmative action is explicitly limited to the elimination of discrimination.

Much of this work has assessed the impact of respondent gender and racioethnicity and has sometimes found that stigmatization is most evident among respondents who are not members of the target group. With this exception, there is little research on the potential importance of other individual attributes, such as political ideology, prejudice, and belief in a just world, regarding the extent of stigmatization. In addition, there is little use of attribution theory other than the principle of discounting.

### 3.2. Target Group Members' Anticipation of Stigmatization

Conceptual and empirical work on self-fulfilling prophecies and stigmas suggests that the anticipation of stigmatization will affect the individual's affective, cognitive, and behavioral reactions. The belief that others consider one to be inferior will usually be distressing. It can stimulate either assumption-confirming or assumption-disconfirming behavior, depending on the situation and the individual. Indeed, concerns about stigmatization have motivated several Black and Hispanic intellectuals to speak out against affirmative action. However, little research directly addresses this issue in the context of affirmative action. The research that exists suggests that many White women and racioethnic minorities expect to be stigmatized by others even in the absence of affirmative action. They anticipate that affirmative action will increase the stigmatization, although they may express less concern about this possibility than do White males. The anticipation of stigmatization, with or without affirmative action, affects their feelings and behavior. One consequence may be a lessening of support for affirmative action. Nonetheless, they apparently believe that the known disadvantages of discrimination outweigh the possible disadvantages of stigmatization because they consistently support nonpreferential versions of affirmative action. A second possible effect could be an increase in stereotype threat. Stereotype threat refers to the fear that one may behave in such a way as to

confirm negative stereotypes of one's group. It is exacerbated by factors that draw attention to the stereotype and to the individual's membership in the devalued group. Although the presence of an AAP may be such a factor, this possibility has not yet been subjected to empirical investigation.

### 3.3. Self-Stigmatization by Target Group Members

Self-stigmatization refers to doubts regarding one's own competence, subsequent selection of a less demanding task rather than a more demanding task, decreased interest in the position in question, negative affect, and the like. There are two primary reasons to anticipate self-stigmatization as a result of preferential selection. The first relates to the discounting principle mentioned previously. If a woman knows that her selection is based on her gender rather than her qualifications, she may be less confident about her qualifications than she would be if merit were the only plausible reason for her selection. Preferential selection may also lead to self-stigmatization because receiving help can have self-threatening implications. The fact that one is helped suggests that one needed to be helped, and this implies incompetence.

Most research in this area has used the following research procedure. Male and female participants (typically White undergraduates) believe that they are about to partake in a one-way communication task. Assignment to the leader and follower roles is said to be based either on gender or on test performance. In reality, assignment to role is predetermined and the participant is assigned to the leader role. After role assignment, participants are asked about their leadership ability in general, their interest in serving as the leader in the communication task, and the like.

This research reveals consistent evidence of self-stigmatization by preferentially selected White undergraduate women. This effect is clearer for judgments of competence than for reports of motivation or task interest. However, when the AAP involves selection among comparable candidates, self-stigmatization is eliminated. Research on people of color is very sparse, but the work that exists has failed to observe self-stigmatization. This null effect likely stems from the belief by people of color that they are hurt more by discrimination than they are helped by affirmative action. Similarly, virtually all of this research has studied undergraduates; the limited work on employed adults has obtained no evidence of self-stigmatization. Several

studies failed to replicate the effect with White male respondents. This is consistent with the general lack of concern about self-stigmatization among White males, who have profited from traditional race and sex discrimination for centuries. These results suggest that initial self-confidence plays a critical role in reactions to preferential selection, and there is some empirical support for that perspective. Finally, psychological processes such as self-protective and egocentric biases would work against the development of self-stigmatization.

### 3.4. Summary

One observation applies to all of the work that has been discussed in this section: Negative effects are limited to situations in which affirmative action is operationalized as preferential treatment or is assumed to involve preferences. Preferential selection causes observers to question the competence of the selected individual and causes selected White women to doubt their own competence. These effects are eliminated or sharply reduced when affirmative action involves preferences among comparable candidates, and these effects are fully eliminated when affirmative action is restricted to the elimination of discrimination and opportunity enhancement efforts such as targeted recruitment.

## 4. ATTITUDES TOWARD AFFIRMATIVE ACTION

By far the largest body of research on affirmative action deals with the variables that predict attitudes. Support for affirmative action is frequently assessed in large surveys, and many dozens of questionnaire studies have been published.

Although many attitudinal predictors have been studied, this article is limited to the variables that have received substantial empirical attention. These predictors can be divided into three categories: structural aspects of the AAP, characteristics of the respondents, and psychological mediators. The structural factors include AAP strength, identity of the target group, and justification. The individual difference variables fall into two categories: demographic characteristics and opinion variables. Finally, the respondent's personal and collective self-interest and the perception of affirmative action fairness are presumed to mediate the relationship between the individual difference predictors and attitudes.

## 4.1. Structural Factors

The structural factors include details of the AAP. Real AAPs are complex. They include utilization analyses that compare the employer's workforce with the corresponding qualified labor market as well as organizational structures that assign responsibilities for maintaining equity. If target groups are underrepresented, the AAP includes goals for target group employment and a set of processes for implementation of the change. In contrast to this complexity, few research descriptions of AAPs exceed a single paragraph. In addition, much of this research includes AAPs that would be illegal if actually implemented.

### 4.1.1. Affirmative Action Plan Strength

The structural factor that has received the most attention is the manner in which the AAP attends to demographic status. This is typically referred to as AAP strength. The most common AAPs used in research, from weakest to strongest, include (a) the elimination of discrimination against the target group; (b) opportunity enhancement activities such as focused recruitment of the target group and the provision of training; (c) a tie-break procedure (often referred to as weak or soft preferential treatment) in which members of the target group are preferred if, and only if, their qualifications are equivalent to those of nontarget applicants; and (d) a preferential procedure (typically referred to as strong or hard preferential treatment) in which members of the target group are preferred even if their qualifications are inferior to those of nontarget applicants. Quotas would also qualify as strong preferential treatment. Although strong preferential treatment is normally illegal, this is the version most frequently used in behavioral research. Note that these four procedures confound the weighting of demographic status with the relevant phase(s) of the employment process.

Most research has found a negative effect of AAP strength on attitudes. The elimination of discrimination receives strong support. Opportunity enhancement procedures receive somewhat weaker support. Preferences are opposed, with most studies finding greater opposition to strong preferential treatment than to weak preferential treatment.

However, the preceding conclusions must be qualified. To a great extent, they are drawn from research on White respondents. When the effect of AAP strength on attitudes is studied within racioethnic minority groups, the negative monotonic effect observed

among Whites is not always observed. Although all groups express opposition to strong preferences, racioethnic minorities make less of a distinction than do Whites among the weaker versions of affirmative action. Indeed, two studies on minority respondents have found greater support for recruitment than for the elimination of discrimination or any other AAP. Very few studies have included multiple racioethnic groups and AAPs of varying strength, so this conclusion is tentative.

### 4.1.2. Identity of Target Group

A few studies have manipulated the identity of the target group. There is a tendency for White respondents' attitudes to be less positive when the AAP targets minorities (especially African Americans) than when it targets White women or individuals with disabilities. Explanations for this effect could include racial prejudice (which would lower evaluations when the AAP targets racial minorities) and inclusion of the target group in the respondents' scope of justice (which would raise evaluations when the AAP targets White women or individuals with disabilities). However, there is such little work in this area that these conclusions must be tentative. One would expect the impact of target group status on attitudes to vary with demographic status of the respondents, but this interaction has received too little attention to permit any conclusion.

### 4.1.3. Justification

A number of studies have assessed the effects of explicit justification on attitudes. This work has found a modest positive effect of justifying the AAP on the basis of fairness (it is needed to remedy past discrimination) or economic value (it increases diversity, which has a positive impact on the organization's success). Justifying the AAP by stating that the target group is underrepresented appears to have a neutral or very small positive effect.

## 4.2. Individual Difference Variables

Dozens of studies have correlated affirmative action attitudes with a host of individual difference variables. The conclusions drawn in what follows are largely based on a meta-analysis of this research that was presented by David A. Harrison, David A. Kravitz, and Dalit Lev-Arey at the 2001 meeting of the Academy of Management.

### 4.2.1. Demographic Variables

Two demographic variables, race/ethnicity and gender, have received the bulk of the attention, although there is also enough research on education to permit conclusions. Other variables such as age, income, and religion have occasionally been studied, but they have typically been included as controls rather than as variables of central interest. It is important to emphasize that demographic variables cannot serve an explanatory role. If one finds an effect of gender, for example, then one must ask why the effect occurred. The answer necessarily lies in the opinion variables that covary with gender and with the psychological mediators, as detailed in what follows.

**4.2.1.1. Race/ethnicity** One of the most consistent and powerful predictors of affirmative action attitudes is the respondent's race/ethnicity. African Americans are the most supportive, Whites are the least supportive, and Hispanic Americans usually fall between the other two groups. The very limited work on Asian Americans suggests that their attitudes are usually more favorable toward affirmative action than those of Whites but are less favorable than those of other minorities. Despite these mean differences, it is important to note that disagreement exists within each of the groups. Many White Americans support affirmative action, and many African Americans oppose it. Also, as mentioned previously, attitudes vary with the interaction of race/ethnicity by AAP strength. There is a large effect of race/ethnicity in evaluations of strong preferential treatment, a weaker effect in evaluations of weak preferential treatment and opportunity enhancement plans, and no effect in evaluations of the elimination of discrimination.

Why is the effect of race/ethnicity so consistent and substantial, at least in reactions to the stronger AAPs? One simple explanation is that attitudes are a function of self-interest, and affirmative action has different implications for the interests of the various groups. There is no doubt some truth to this explanation, but it is much too simple. For historical and cultural reasons, individuals in the various race/ethnic groups see the world in different ways, and these differences affect their affirmative action attitudes. If we consider Blacks and Whites in the United States, for example, we find that African Americans are more likely than Whites to be politically liberal and to identify with the Democratic party. They are more likely than Whites to have experienced racial discrimination and to believe

that it is an ongoing problem. They are less likely than Whites to be racially prejudiced against Blacks and to believe that affirmative action involves preferences. Each of these differences is associated with attitudes toward affirmative action.

**4.2.1.2. Gender** Numerous studies have assessed gender differences in attitudes toward affirmative action. In general, women are more supportive than men. However, this conclusion must be qualified. First, it is not clear whether it applies equally to all racial groups. It is only among Whites that gender determines target group membership, so self-interest considerations would suggest that the gender effect should be larger among Whites than among other race/ethnic groups. At least one study has reported a reverse effect of gender among African Americans. Second, the author's reading of the literature suggests that the gender effect has shrunk over the past 20 years or so. However, the correlation between effect size and year of publication has not yet been reported, so this supposition is tentative.

**4.2.1.3. Education** The third demographic variable that has received the most attention is education. Some of the most significant work on education has used it as a moderator of other predictors in the context of tests between competing theories. When the main effect of education has been assessed, the effect has varied with respondent race. The relation between education and support for affirmative action is usually positive among African Americans and is usually nonsignificant or negative among Whites. This latter result is anomalous because education is also associated with decreased racial prejudice and increased political liberalism, both of which predict support for affirmative action. This relationship between education and support for affirmative action among Whites merits closer investigation.

### 4.2.2. Opinion Variables

The two theoretical perspectives that have received the most research attention emphasize the importance of political ideology and prejudice. Distinguishing between these two approaches is complicated by the positive correlation between political conservatism and most measures of prejudice. A third theory posits that prejudice and political orientation serve as mediators between a desire for group dominance and affirmative action attitudes. Each of these three perspectives is described in what follows.

**4.2.2.1. Prejudice** Research indicates that prejudice is among the most powerful predictors of White Americans' attitudes toward affirmative action. However, the effect of racial prejudice varies with the measure used. The correlation is larger when racial prejudice is assessed with a measure of contemporary racism than when it is assessed with a measure of old-fashioned racism, and it is smaller yet when it is assessed as pure anti-Black affect. Because of a dearth of research, comparable distinctions among measures of sexism cannot be made.

The distinction between contemporary racism and old-fashioned racism requires some explanation. The core of old-fashioned racial prejudice, which prevailed among Whites until the past few decades, is the assumption of inherent inferiority of non-White racial groups. Blacks, in particular, were assumed to be inferior in intelligence and other characteristics. This belief was accompanied with support for racial segregation and for discrimination. Such beliefs have become relatively uncommon. Instead, many scholars argue, such old-fashioned racism has been replaced with contemporary forms of racism. The author uses the term "contemporary racism" to represent several different but related concepts, including ambivalent racism, aversive racism, laissez-faire racism, modern racism, racial resentment, subtle racism, and symbolic racism. These various conceptualizations differ in important ways, but most incorporate the following components. First, they assume that contemporary racists have underlying anti-Black feelings. Second, they assume that contemporary racists do not accept old-fashioned racist beliefs in inherent inferiority, segregation, and discrimination. Instead, contemporary racists believe that African Americans violate American values such as individualism and self-reliance. Third, contemporary racists acknowledge that racial discrimination existed in the past but believe that it is no longer a problem. Thus, they conclude that any economic disadvantages experienced by Blacks are due to their cultural inferiority. Finally, contemporary racists reject racism and any suggestion that their own beliefs are racist. One implication of this is that they feel most free to express their prejudice when some non-prejudicial reason for their action can be offered.

Given this distinction between old-fashioned racism and contemporary racism, one can understand why the latter predicts opposition to affirmative action more strongly than does the former. The anti-Black affect assumed by contemporary racism predisposes individuals to oppose affirmative action. The assistance that

affirmative action offers to target group members decreases the importance of self-reliance. In addition, affirmative action is designed to eliminate the effects of discrimination that contemporary racists do not believe exists. Finally, because they can justify their opposition to affirmative action by arguing about the importance of self-reliance and the lack of ongoing discrimination, contemporary racists feel free to oppose it. On the other hand, old-fashioned racists who genuinely believe that Blacks are genetically inferior may conclude that African Americans need the assistance offered by affirmative action and so lessen their opposition. In addition, the ideological argument for individualism that is at the center of the contemporary approaches, and that is a core reason for opposition to affirmative action, is not central to old-fashioned racism. When both old-fashioned racism and symbolic racism are included as predictors of affirmative action attitudes, the latter is appreciably more important and typically explains incremental variance beyond the former.

One problem with the contemporary racism approach is that it overlaps conceptually and (in some cases) empirically with political conservatism. Thus, an individual who espouses a conservative ideology but genuinely lacks any prejudice toward Blacks would score relatively high on most measures of contemporary racism.

**4.2.2.2. Political Orientation** Beliefs associated with respondents' political orientation consistently predict attitudes toward affirmative action. Specifically, numerous studies have found that support for affirmative action is greater among Democrats than among Republicans and is greater among liberals than among conservatives.

From the beginning of the modern civil rights era during the 1950s, the major political parties in the United States have taken opposing positions on many racial issues. This divergence was solidified in the 1964 presidential race, when the Republican party nominated Barry Goldwater to run against President Lyndon Johnson. The civil rights movement played a central role in that election, and this established a pattern of Republican opposition to, and Democratic support for, government policies designed to assist racioethnic minorities. Indeed, in recent national elections, opposition to affirmative action, in particular, has been a formal or informal plank in the Republican party platform. Republican politicians typically refer to affirmative action as "preferential treatment" or "quotas" and argue

for its elimination. With this history, it is no surprise that party identification predicts public attitudes toward affirmative action.

Political conservatism is an even stronger predictor of affirmative action attitudes than is political party identification. Political conservatism includes several beliefs that engender opposition to affirmative action. First, conservatism posits that government should not intervene in the marketplace. Affirmative action is a federally mandated policy that constrains business actions. Second, conservatism is associated with support for self-reliance and individualism in both its normative and descriptive meanings. Normatively, individualism argues that individuals should take responsibility for their own success and should not rely on assistance from others (e.g., the government). Descriptively, individualism argues that success in the United States actually is determined by individual actions rather than structural forces. Affirmative action, with its underlying assumption that success is affected by structural factors and its targeted assistance to women and minorities, is contrary to individualism. A related point is that conservatives would support actions designed to ensure equality of opportunity but not those designed to ensure equality of outcomes. For these and associated reasons, conservatism is a strong predictor of opposition to affirmative action, particularly when it is construed as involving preferences.

The past decade has seen the development of a principled conservatism theory of affirmative action attitudes. The argument is that opposition to affirmative action is largely a function of race-neutral conservative principles rather than of racial prejudice. These scholars do not deny the importance of prejudice, but they argue that it is more important among the less educated than among the more educated, who are better able to behave in a manner consistent with their principles. One important point made by some principled conservatism scholars is that members of the public may be less than candid in their responses to survey questions regarding sensitive racial issues. When these researchers use both overt and covert measures of anger about affirmative action, they find the usual effect of political ideology on the overt measures, but this effect is greatly reduced or entirely eliminated on the covert measures. The effect of political orientation decreases because liberals express more anger on the covert measures than they do on the overt measures. A related point is that the effect of prejudice on affirmative action attitudes was stronger among liberals than among conservatives.

**4.2.2.3. Social Dominance Theory** A number of scholars have developed models that emphasize the role of group conflict in motivating attitudes toward affirmative action. These approaches assume that societies are organized as hierarchies, with some social groups being dominant and others being subordinate. To maintain its position, the dominant group creates an ideology that legitimizes the status quo. In the current case, this ideology incorporates meritocracy, political conservatism, classical racism, and a belief in the cultural inferiority of racial minority groups. These beliefs serve as legitimizing myths; they legitimize the differences in status between the dominant and subordinate groups. They also legitimize opposition to affirmative action, which threatens the status hierarchy. This theory explains the correlation between prejudice and political conservatism, both of which serve as legitimizing myths, as well as the correlations of each with opposition to affirmative action.

Social dominance theory differs from some other conflict-based theories in its assumption that people vary in the extent to which they have internalized this perspective. This personality trait is called social dominance orientation. Supporting social dominance theory, there is a consistent correlation between social dominance orientation and affirmative action attitudes.

**4.2.2.4. Comparison among Theories** There has been a rigorous debate within the affirmative action area among scholars who espouse the prejudice, principled conservatism, and social dominance perspectives. The resulting literature is quite complex and cannot be summarized briefly. In addition, much of it suffers from a significant limitation, that is, the equating of affirmative action with preferential treatment. As mentioned previously, attitudes toward affirmative action are strongly affected by AAP strength. More importantly, there is evidence that the importance of certain predictors varies with AAP strength. For example, research suggests that principled conservatism may be especially important in predicting reactions to strong forms of affirmative action that run counter to the basic principles of conservatism. On the other hand, racial prejudice seems to be most important in predicting opposition to weaker forms of affirmative action that are less inconsistent with conservative principles. To the author's knowledge, no research has assessed the relative importance of social dominance orientation across AAP strength.

The consistent correlation between measures of prejudice and opposition to affirmative action supports

both the prejudice and social dominance approaches. However, it is important to emphasize that the racism literature is complex and that the strength of the racism–attitude correlation varies with the conceptualization and measure of racism. The correlation is most substantial in studies that employ measures of contemporary racism that often overlap with measures of political conservatism. A limitation of the prejudice approach is that it does not explain attitudes of target group members.

Much of the research designed to competitively test these theories assesses the impact of potential moderators. For example, research has found that correlations between affirmative action attitudes and all three predictors increase with respondent education and political sophistication. This is fully consistent with the logic of social dominance theory, but the principled conservatism approach would predict a negative effect of education on the correlation between affirmative action attitudes and racism. Also contrary to the principled conservatism approach, the correlation between racism and political conservatism increases with education. However, these results are based on overt measures of racism and affirmative action attitudes. As mentioned previously, the principled conservatism scholars have demonstrated that the use of covert measures can profoundly affect results.

Some results supportive of social dominance theory were mentioned previously. In addition, research has shown that the significant correlation between racism and conservatism disappears when social dominance orientation is controlled. This result is consistent with the assumption that racism and conservatism are merely legitimizing myths that render the dominance hierarchy acceptable. Other supportive research has involved the use of structural equation modeling analyses of the social dominance and principled conservatism theories.

Finally, at least one study found that all three concepts—racism, conservatism, and social dominance orientation—contributed unique variance to the prediction of affirmative action attitudes. It is not now possible to say that one of these approaches is right and the others are wrong; it is likely that each is at least partly correct. It is also likely that the validity of these approaches varies with the operationalization of affirmative action. Unfortunately, little can be said about the moderating effect of AAP strength because, as noted previously, the vast majority of studies in the principled conservatism and social dominance areas have equated affirmative action with preferential treatment.

#### 4.2.2.5. *Beliefs About the Extent of Discrimination*

More than 20 studies have correlated beliefs about the extent of target group discrimination with affirmative action attitudes. The result is clear: Support for affirmative action increases with the perception that the target group has suffered from discrimination. There are at least two reasons for this relationship. First, most people support equality of opportunity for all racial groups and actions designed to ensure such opportunity. Consistent with this reasoning, the association of perceived discrimination with support for affirmative action appears to be most substantial for the strongest AAPs. The second reason for this correlation is that perceptions of target group discrimination are associated with other predictors of affirmative action attitudes. For example, political liberals generally believe that discrimination is still a major problem, whereas conservatives tend to believe that discrimination is now uncommon. Along the same lines, social dominance theory would argue that disbelief in discrimination serves as a legitimizing myth.

#### 4.2.3. *Psychological Mediators*

Finally, research has focused on the psychological mediators of affirmative action attitudes. Precisely how do AAP strength, racioethnicity, political orientation, and other variables affect attitudes? The two answers that have received the most attention are that they affect the respondent's perceptions regarding (a) the impact of the AAP on his or her interests and (b) the fairness of the AAP.

##### 4.2.3.1. *Self-Interest (Personal and Collective)*

Assumptions regarding self-interest underlie the conflict theories, such as social dominance theory, mentioned previously. Self-interest has been assessed both indirectly and directly. An example of an indirect approach would be to use respondent age as an indicator of economic self-interest, assuming that younger workers are more strongly affected by affirmative action policies than are older workers. The direct approach asks the respondent to rate the likely impact of the AAP on his or her outcomes. Each of these approaches is flawed. The indirect measures suffer greatly from construct contamination and deficiency, whereas the direct measures suffer from common method variance and experimental demands. Research using indirect measures of self-interest finds that it is only weakly related to affirmative action attitudes, whereas work employing direct measures finds a strong relation.

Furthermore, research using direct measures finds that self-interest partly mediates the impact of other predictors such as gender and race/ethnicity.

Self-interest has been studied at both the personal and collective levels. Personal self-interest refers to the impact of the AAP on the respondent's own outcomes. Collective self-interest refers to the impact of the AAP on the respondent's demographic group, which is typically defined in terms of race/ethnicity and perhaps gender. The few studies that have included both predictors have found that they correlate similarly with affirmative action attitudes.

**4.2.3.2. Perceived Fairness** Attitudes toward affirmative action are more closely related to perceived fairness of affirmative action than they are to any other variable. Indeed, research on evaluative reactions to affirmative action has often used measures of fairness rather than attitudes as the dependent variable. People report positive attitudes toward AAPs they consider fair and report negative attitudes toward AAPs they consider unfair.

Fairness perceptions mediate many of the effects described previously. Strong AAPs that provide rewards on the basis of demographic characteristics rather than merit are disliked because they are seen as unfair. That is a central argument of the principled conservatism perspective. Justifications affect attitudes in part because they affect the respondent's beliefs about the fairness of the AAP. In short, support for affirmative action is strongly predicted by beliefs about its fairness.

### 4.3. Complications and Limitations

As detailed previously, attitudes toward affirmative action are affected by actual or perceived details of the AAP, by the target group, and by justification for the plan. Attitudes are also associated with various demographic factors and opinion variables. Many of these effects are mediated by judgments of AAP fairness and self-interest implications. Most scholars would acknowledge the importance of all these predictors. There is disagreement, however, regarding their interrelationships and their relative importance. Unfortunately, most studies include relatively few predictors. In these cases, one cannot know how important the predictors would be in the context of additional variables. A related issue is that many of the most important predictors are individual difference variables that are intercorrelated and over which the experimenter has no control. This exacerbates the "missing variables" problem and also

means that causal conclusions cannot be drawn from the results. Of perhaps greater significance, not enough is known about how the predictors interact. We know that the relative importance of race/ethnicity varies with AAP strength, but because of a dearth of research, it is impossible to draw strong conclusions about other interactions. This limitation is due in part to an unfortunate tendency for research to include only a single AAP.

The operationalizations of affirmative action have tended to be brief, simplistic, and (often) illegally strong. In many studies, affirmative action is not even defined. Instead, the stimulus is simply the term "affirmative action." In such situations, responses are determined by the participant's beliefs about what affirmative action entails, and these beliefs are rarely assessed. In addition, this work relies on respondent self-reports. Self-reports on controversial issues may be less than candid, especially if the research procedures do not guarantee anonymity. Furthermore, much of this research has used only White respondents. Even more disturbing, a few studies pool data across all race/ethnic groups, and many studies pool results across minority groups. There is evidence that the minority groups differ in their attitudes. Of greater importance, it appears that minority groups may also differ in the factors that predict those attitudes.

Affirmative action is a complex social policy, and the determination of attitudes toward it is complex. Theories and research designs must be similarly complex if we are to obtain a fuller understanding of attitudes toward affirmative action. There is ample room for progress in this area.

## 5. CONCLUSIONS

One common thread weaves throughout the tapestry of research on affirmative action: Many negative effects are associated with the use of strong preferences but not with weak forms of affirmative action. Negative effects of affirmative action on organizational performance should be limited to situations in which the organization gives preference to underqualified target individuals. Stigmatization of target group individuals by others and by the individuals themselves occurs when affirmative action involves (or is believed to involve) preferences but not when preferences are known to be absent. Opposition to affirmative action and the resulting conflict exist primarily in response to preferences and quotas. The good news is that such

strong preferences are illegal, at least in the United States. The bad news is that a substantial segment of the population equates affirmative action with preferences. A major challenge for the future is to craft affirmative action procedures that are widely perceived to be fair and legitimate and that improve employment opportunities for traditionally underrepresented groups. Furthermore, the nonpreferential nature of these plans must be publicized, and the tendency to equate affirmative action with preferences should be eliminated in public discourse and academic research. Such actions should decrease social conflict while leading to a more equitable society.

### Acknowledgments

I thank Stephen L. Klineberg, Arthur Gutman, Barbara J. Martin, and Michelle Marks for their helpful comments on earlier drafts of this article.

### See Also the Following Articles

Employment Discrimination ■ Prejudice and Discrimination

### Further Reading

- Bobocel, D. R., Son Hing, L. S., Davey, L. M., Stanley, D. J., & Zanna, M. P. (1998). Justice-based opposition to social policies: Is it genuine? *Journal of Personality and Social Psychology*, *75*, 653–669.
- Crosby, F. J. (2004). *Affirmative action is dead; long live affirmative action*. New Haven, CT: Yale University Press.
- Crosby, F. J., & VanDeVeer, C. (2000). *Sex, race, and merit: Debating affirmative action in education and employment*. Ann Arbor: University of Michigan Press.
- Donohue, J. J., III, & Heckman, J. (1991). Continuous versus episodic change: The impact of civil rights policy on the economic status of Blacks. *Journal of Economic Literature*, *29*, 1603–1643.
- Doverspike, D., Taylor, M. A., & Arthur, W., Jr. (2000). *Affirmative action: A psychological perspective*. Huntington, NY: Nova Science Publishers.
- Federico, C. M., & Sidanius, J. (2002). Sophistication and the antecedents of Whites' racial policy attitudes: Racism, ideology, and affirmative action in America. *Public Opinion Quarterly*, *66*, 145–176.
- Heilman, M. E., Battle, W. S., Keller, C. E., & Lee, R. A. (1998). Type of affirmative action policy: A determinant of reactions to sex-based preferential selection? *Journal of Applied Psychology*, *83*, 190–205.
- Holzer, H. J., & Neumark, D. (2000). Assessing affirmative action. *Journal of Economic Literature*, *38*, 483–568.
- Holzer, H. J., & Neumark, D. (2000). What does affirmative action do? *Industrial and Labor Relations Review*, *53*, 240–271.
- Kravitz, D. A., Harrison, D. A., Turner, M. E., Levine, E. L., Chaves, W., Brannick, M. T., Denning, D. L., Russell, C. J., & Conard, M. A. (1997). *Affirmative action: A review of psychological and behavioral research*. Bowling Green, OH: Society for Industrial and Organizational Psychology.
- Kravitz, D. A., & Klineberg, S. L. (2000). Reactions to two versions of affirmative action among Whites, Blacks, and Hispanics. *Journal of Applied Psychology*, *85*, 597–611.
- Major, B., Feinstein, J., & Crocker, J. (1994). Attributional ambiguity of affirmative action. *Basic and Applied Social Psychology*, *15*, 113–141.
- Rubio, P. F. (2001). *A history of affirmative action, 1619–2000*. Jackson: University Press of Mississippi.
- Sears, D. O., Van Laar, C., Carrillo, M., & Kosterman, R. (1997). Is it really racism? The origins of White Americans' opposition to race-targeted policies. *Public Opinion Quarterly*, *61*, 16–53.
- Selmi, M. (1995). Testing for equality: Merit, efficiency, and the affirmative action debate. *UCLA Law Review*, *42*, 1251–1314.
- Sniderman, P. M., & Carmines, E. G. (1997). *Reaching beyond race*. Cambridge, MA: Harvard University Press.
- Spann, G. A. (2000). *The law of affirmative action: Twenty-five years of Supreme Court decisions on race and remedies*. New York: New York University Press.



# Age-Related Issues among Minority Populations

Edna E. Brown and James S. Jackson

University of Michigan, Ann Arbor, Michigan, USA

1. Introduction
  2. The Changing Nature of Aging in America
  3. A Life Course Framework for Aging in Racial and Ethnic Groups
  4. Resource Dependency: Work, Asset Accumulation, and Retirement
  5. Importance of Formal and Informal Care
  6. Receiving Care
  7. Importance of Perceived Control to Physical and Mental Health Outcomes
  8. Receiving Care, Sense of Control, and Psychological Well-Being
  9. Racial and Ethnic Minority Elderly as a Resource
  10. Conclusions
- Further Reading

## GLOSSARY

**biopsychosocial model** A framework for understanding human behavior, health, and well-being across the life span; it incorporates the biological, psychological, and social perspectives on the etiology, understanding, and treatment of illness and disease.

**formal care** Paid care received from private or government social service agencies or organizations.

**informal care** Unpaid care received from family, friends, neighbors, or religious organizations.

**receiving care** Help with impairments in activities of daily living (ADLs), such as eating, walking, using the toilet,

bathing, and getting in and out of bed, or help with limitations in instrumental activities of daily living (IADLs), such as cooking, cleaning, and shopping.

**reciprocity** The exchanges of goods, services, and support of equal value among members of a social network.

**sense of control** The degree to which people believe that they can deliberately bring about desired outcomes and avoid undesirable ones in their environment.

In contrast to just a few years ago, more research on the psychological and social gerontology of racial and ethnic minority groups is being included within the general investigation of ethnicity and cultural factors in aging and human development. The existence of large national data sets and more powerful analytical techniques is increasing the quality and quantity of aging research on these groups in many areas of psychological interest. National longitudinal data collection efforts are improving the available data on the aging experience of minority Americans. Although better data are always needed, the improvement in a relatively few short years has been impressive. Similarly, the approach to research on the ethnic minority elderly has also seen a greater recognition of the heterogeneity among and within these groups. Research is now more focused on the role of culture, socioeconomic status, and gender as important markers of potential process differences within and among aging

groups of color, especially Black Americans. Recent ethnogerontology research is reversing historical trends, and generalizable high-quality findings are emerging concerning health, socioeconomic status, work and retirement, social support, well-being, and especially family patterns of caregiving and care receiving. This article reviews some implications of this work for the applied psychology of aging, with a special focus on issues related to perceptions of control, family caregiving, and care receiving among racial and ethnic minority populations.

## 1. INTRODUCTION

Research on racial and ethnic minority populations has propelled significant interest in the development of life course models and frameworks that assume an important role of race and ethnicity, culture, acculturation, and national origin in how various groups in the United States traverse the individual life course and age. The growth of racial and ethnic groups in the United States, and the relative deprivation of these groups in material, social, and health resources, demands greater attention if we are concerned about characterizing them appropriately and providing effective and adequate services as they age.

The growing internationalization of research on aging demands that we move beyond singular cultural perspectives. Fortunately, the United States is culturally heterogeneous, and comparative research on racial and ethnic groups yields excellent models relevant to understanding basic processes in aging in larger cross-national contexts. Important processes, such as stress and coping, may be more sharply defined in environmentally pressured groups (e.g., Blacks, Latinos, Native Americans). Racial and ethnic group boundaries may provide a personal and group coalescing of culture/environment and biological differences. For example, recent research supports degrees of assimilation and acculturation as important variables for understanding basic processes in mental health and mental disorder outcomes.

## 2. THE CHANGING NATURE OF AGING IN AMERICA

America has always been a nation of immigrants. Several factors will change the face of America in

the future. The population will age dramatically. The racial and ethnic composition of the country will change significantly. Earlier waves of migration came from Europe, China, and Africa, whereas new waves are coming from Central and Latin America, the Middle East, and Southeastern Asia. Although the country is aging rapidly, the rate of aging is faster among the racial and ethnic minority "new" immigrant groups, especially Latinos, Asians, and (to a lesser extent) African Americans and other Black immigrant groups.

This changing racial and ethnic minority population will represent culturally unique groups of aging individuals. As culture has affected their lifetime experiences, it will also influence the nature of their aging experiences. Elements of the biopsychosocial model of aging are fundamentally affected by the cultural experiences of aging individuals, both cumulatively and contemporaneously. Immigrant groups in the United States have unique histories. For example, Cuban immigrants fled Castro and arrived in the United States feeling forcibly expelled from their homeland. They came to the United States, which was largely anti-Castro and sympathetic to their plight. Middle Easterners escaped dictators, war, and oppression, but (at least currently) they are aging in a considerably less sympathetic, if not hostile, United States. Asians (e.g., Koreans) also fled a homeland torn by war and remain embedded in communities with specific cultural values that guide their expectations about aging. The life span experiences of these individuals may include immigrating as children and spending 60 years in the United States, learning another language, living in urban versus rural communities, and being among a minority versus majority culture. Some immigrants left a higher standard of living, whereas others came to the United States to seek a higher standard of living. But particularly relevant to the aging experience is the fact that all immigrants have been influenced by their biological and physical characteristics, their family and friendship relationships, and the communities within which they live (i.e., their biopsychosocial development). Although this article focuses on minority racial and ethnic minority groups in the United States, these biopsychosocial influences are no less important for the majority group. In this case, recognizing the significant influences on racial and ethnic minority groups only serves to highlight the role of these factors for aging among all individuals.

### 3. A LIFE COURSE FRAMEWORK FOR AGING IN RACIAL AND ETHNIC GROUPS

As changing demographics require increased focus on older ages, lessons learned from the life span model of development are particularly germane. A life span framework is needed to explore how environmental stressors influence and interact with group and personal resources to both impede and facilitate the quality of life of successive cohorts of racial and ethnic minorities over the group life course and in the nature of their individual human development and aging experiences. Research on socioecological predictors of health in African American and Caucasian residents of a large midwestern city exemplifies the need for the life course approach. Findings have demonstrated relationships between socioecological factors (e.g., high crime rates, family dysfunction, high noise levels, social isolation) and negative health factors (e.g., hypertension) among minority residents. These factors can affect all members of the minority families and community, thereby possibly initiating poorer health among both younger and older individuals. It is the premise of a life course perspective that current and aging cohorts of ethnic minorities have been exposed to the conditions that will profoundly influence their social, psychological, and health statuses as they reach older ages during the years and decades to come.

Recent work on the incomplete ontogeny of human development and the recognition of the influence of culture in human development represents an important extension of the original work. In this latest consideration of life span developmental theory, the concepts of selection, optimization, and compensation are illustrative of the adaptive strategies that people use to maximize their competency. With the challenges of age, one has an increased need to select, optimize, and compensate to achieve designated goals. At the same time, it is noteworthy that culture is infused in every element of this process. What one selects, how one chooses to optimize, and what one considers appropriate forms of compensation all are culturally influenced if not culturally determined. As one ages, this influence accumulates and appears to have ever-increasing effects on the experiences of aging. The greater emphasis on culture in this life course modeling is very important in understanding the experiences of racial and ethnic minority groups. In fact, some have suggested that although culture becomes increasingly important

with age, its actual influence on development decreases with age. Thus, biological and physical development are said to dominate during childhood when culture has minimal impact. Within most Western cultures, this appears to be true. But of course, one might imagine a culture where one group is disfavored (e.g., girls) and, thus, is provided only very limited access (e.g., to nutritional resources), thereby limiting the biological or physical development that might otherwise dominate. Similarly, among older people, what one values and how one copes with biological or physical losses might also be influenced by culture.

The current authors believe that it is necessary to incorporate both the life span model and the role of culture, especially a view of culture nuanced by racial and ethnic differences, into life course biopsychosocial models, especially as they apply to aging.

### 4. RESOURCE DEPENDENCY: WORK, ASSET ACCUMULATION, AND RETIREMENT

#### 4.1. Socioeconomic Status

Many minority elderly individuals continue to lag behind Whites in social and economic status. For example, indicators of income, education, and health status document the deprived position of Blacks relative to Whites. Some argued that this was a cohort effect, but the lingering poor relative position of Blacks refutes this; new cohorts of the Black elderly are not faring significantly better than prior ones. For example, there is a continuing disadvantage of middle-aged and younger Blacks (and increasingly Hispanic groups), relative to Whites, in housing, income, occupation, health, and education. Recent reports on the circumstances of Blacks across the entire life course continue to show the presence of relatively poorer circumstances, especially wealth, suggesting that new cohorts of elderly Blacks will continue to experience relative disadvantage in comparison with other groups.

On the other hand, even though approximately one-third of the Black elderly continue to live below the poverty level, today's Black elderly are better fed, better housed, and in better health than were those during earlier eras. Most of this improvement is attributable to government assistance programs, which is still the prime support of many black Americans and increasingly among other older age groups as well. Unfortunately,

because of histories of poor occupational opportunities, lack of wealth, and private retirement funds, a large proportion of Blacks are heavily dependent on these government programs. It is still unclear whether future cohorts of older Blacks will enjoy what may be the relative luxury of today's elderly. Recent diminution of federal programs, a change in the economy favoring job creation in "high-tech" and specialized educational intensive sectors of the economy (e.g., computers, communications), and simultaneous growth in low-paying service positions (e.g., fast-food restaurants) provide little room for today's Black adult and middle-aged cohorts, many of whom lacked basic educational opportunities during their formative years. It is now predictable that future cohorts of older Blacks may not be as well off as their White counterparts, and it is still unlikely that they will be as well off as today's cohort of Black elders, many of whom worked in union-intensive industries.

#### **4.2. Health Morbidity and Mortality**

At nearly every point from birth to death, African Americans and other minority groups (e.g., Native Americans) have poorer morbidity and mortality rates than do Whites. It is also well documented that there is increased longevity among Blacks (and some suggest that this is also true among Hispanics) who live to approximately 84 years of age. Many have suggested a possible selection bias favoring the long-term survival of particularly robust and hardy individuals (e.g., immigration) or differential rates of aging within Black, Asian, Hispanic, Native American, and White populations. Others have claimed that this crossover is only an artifact of faulty reporting and exaggerated age claims. The effect has been firmly established, although there is still no widely accepted explanation. The racial mortality crossover appears to be a real phenomenon—one that may involve some type of "survival of the fittest."

Recent research on the "oldest old" continues to document the heterogeneity of the social and psychological health of very old Blacks. This type of data provides strong support for a thesis that views the mortality crossover as involving the survival of hardier old Blacks and not a methodological artifact. Similarly, other work reveals some evidence for greater functional health among older elderly Blacks in comparison with Whites, although the effect seems highly dependent on educational status.

Recent research points to differences between ethnic minority groups and Whites in the nature of self-

reported health. Early in the past decade, it was found that the largest differences were in the validity of the subjective interpretations of health state. These findings challenge traditional thinking and research regarding possible race differences in health. At this point, whether there are differences in the structure of health, the processes of health, or the influence of service use on experienced health problems remain open questions. What is clear is that changing health policies may have increasing negative effects on the ability of Black and many other minority elderly to receive adequate health care in the new century.

#### **4.3. Psychological Well-Being**

Research on psychological well-being has shown an increasing sophistication over the past few years. Structural factors, such as income and education, tend to show small but positive relationships to well-being. Some recent work suggests that psychological well-being may be strongly tied to family and health satisfaction. Some recent evidence also suggests that younger ethnic minority cohorts may be less satisfied than older cohorts at comparable periods during the life span. This is in sharp contrast to Whites, who have shown the opposite pattern. This lowered satisfaction and happiness in younger Blacks may be related to rising expectations and structural constraints that are likely to persist into older age.

#### **4.4. Work and Retirement**

Little empirical research had been devoted to the study of work and retirement among ethnic and minority elderly individuals. Some earlier work had speculated that the entire retirement process, viewed within a life span context, may be very different for Blacks and other disadvantaged ethnic minorities. For example, Blacks and other groups, such as Native Americans and some Latino groups, often have long histories of dead end jobs with poor benefits and bleak expectations, thereby lowering any advantages of retirement. Thus, inadequate income, poor housing, and uncertain futures may confront these older groups at retirement age. Faced with limited retirement resources, many older ethnic minorities may continue to work past customary retirement ages out of desperation. Some research indicates that these individuals are physically, psychologically, and socially worse off than their retired counterparts. As suggested earlier, even the relatively poor but stable government retirement

support (if these individuals are fortunate enough to qualify for such support) may, in contrast, be better than sporadic and poor jobs in the regular labor market. Thus, retirement may provide a small but secure government income, leading to increased psychological and social well-being.

#### **4.5. Family and Social Support**

Historically, research on minority family and social support networks has been based predominantly on anecdotal data. Two myths have dominated this area. The first is a view of older minorities being cared for by loving and extended family members and fictive kin. The other is a view of impoverished, lonely, older minority elderly individuals being abandoned by a disorganized and incompetent family system. National and other large social surveys indicate a reality somewhere in between, documenting the existence of extended family forms but also demonstrating that much of the assistance is reciprocal; that is, ethnic minority elderly individuals often provide help to younger family members and neighbors. Recent research supports the importance of community institutions, such as churches, as sources of physical and emotional support for older Blacks. This work points to the considerable obstacles faced by many ethnic minority Americans in providing services to physically disabled and/or demented (and mentally ill) relatives.

### **5. IMPORTANCE OF FORMAL AND INFORMAL CARE**

#### **5.1. Benefits of Support/Informal Care**

Research has provided impressive evidence of the importance of social support throughout the life span, particularly to the elderly. Social support has salutatory effects on physical and mental health of older adults. Support reduces the negative effects of health problems, helps in coping with stressful life events, and buffers the impact of psychological distress. A lack of social support from family and friends is associated with poor psychological well-being (e.g., depression). The most important social support usually involves long-lasting, significant, and close relationships with a network of family, friends, and church members. The quality, rather than the quantity, of the support relationship has been shown to contribute most

significantly to health and well-being. In ethnic minority families, informal support has served to counteract the overall deleterious effects of adverse environmental, social, and economic conditions.

#### **5.2. Sources and Types of Informal Support/Care**

Studies that examine the beneficial effects that care networks have on health and well-being find that support derives from many sources, including spouse, children, siblings, and friends. Racial and ethnic minority families, in comparison with White families, tend to have larger multigenerational households. On the one hand, investigators report that generations live together out of economic necessity due to low socioeconomic status, poverty, and generally few resources and opportunities available to ethnic minorities. On the other hand, some researchers assert that intergenerational households may be more the product of cultural norms such as familism, filial piety, and familial obligation. In any case, the large family provides a functional, mutually beneficial network from which to give and receive care. Common characteristics of ethnic minority families are the importance of strong family bonds, exchanging of resources, and caring for the elderly.

As they age, many elderly individuals perceive that family and friends can be relied on to help with increasing needs. Family, friends, neighbors, and fictive kin play an important role in the care of ethnic minority elders. Mutual assistance, exchanges, and reciprocity involving informal and formal exchanges characterize the support patterns. A variety of tangible and emotional resources are exchanged within racial and ethnic minority populations, including goods and services as well as financial, emotional, and affective support. Instrumental aid (e.g., food, money, transportation) and emotional support (e.g., advice, counseling, visiting, companionship) are exchanged throughout the life course. These exchanges allow elderly individuals to remain viable and independent in their communities.

The elderly prefer assistance in a manner depending on need. Members of helping networks provide different types of support. Some members may give emotional support, such as listening and advice, whereas others may provide instrumental support, such as sick care and financial help. Family assistance may be more appropriate for long-term care, addressing impairments in activities of daily living (ADLs) and limitations in instrumental activities of daily living

(IADLs). Friends and neighbors, in close proximity to elders, may help with daily or short-term needs. Friends may also be a source of companionship and emotional support.

Research provides a wealth of information concerning formal and informal church support. The special role of church support is separate from its religious role for Black elderly individuals. Black churches have proven to be responsive to the needs of communities that have limited access to general societal support systems. Church members exchange instrumental, financial, emotional, and spiritual assistance with each other. This support includes food, clothing, and sick care as well as advice, encouragement, and information.

### **5.3. Reciprocity**

Overall, older individuals are more satisfied with their relationships and have increased well-being if they are involved in reciprocal relationships. The nature of reciprocity exchanges is dependent on the nature of the relationships. Some relationships require immediate and in-kind or equivalent value return of services or goods. Other more intimate or lifelong relationships have more mutually satisfying exchanges over time and less immediacy to the exchanges. For example, receiving more aid than giving aid can result in feelings of guilt and indebtedness. One theory that addresses this issue is Antonucci's notion of a support bank or support reserves. The concept explains how many elderly individuals are able to maintain positive psychological well-being when they are in need of support but are unable to reciprocate. To illustrate, older minorities usually have fewer resources than do younger network members. By participating in years of giving and mutual exchanges, it becomes psychologically acceptable for the elderly to receive care without giving something in return. There is also evidence that some elderly individuals may reduce the size of their social exchange networks so as not to feel indebted to too many people. It is important to keep in mind that the elderly are not necessarily unable to render some type of reciprocity. For example, they might provide child-rearing assistance, including child care and advice, to younger parents in their networks.

### **5.4. Formal Care**

At times, the elderly may need more professional care than family and friends can provide. When necessary,

family, friends, and church members are instrumental in providing referrals to formal services such as professional home care providers. Ethnic minority elders also use formal services in the absence of spouses, children, or other informal helpers. Traditionally, the minority elderly have had few links to formal service. In the past, service delivery organizations rarely acknowledged cultural or family generational lifestyles differences, thereby perhaps contributing to mistrust of formal service organizations.

In general, the ethnic minority elderly are involved in long-term associations involving reciprocal support and care exchanges. More information is needed, however, about how societal changes affect these relationships. For example, younger people who become more educated and economically successful might need to move farther away from their family homes, thereby affecting the availability of care for elderly family members. Furthermore, people are living longer and having fewer children, and more middle-aged adult daughters are working outside the home. These changes also may contribute to a decrease in the number of people available for caring for elderly family members. Among recent immigrants, assimilation and acculturation issues also may complicate familial and friend relationships. The degree of assimilation varies and has effects on the provision of care. Despite histories of disadvantage, including poverty and poor medical care, ethnic minority groups have been creative at developing resources and effective coping strategies to alleviate these deleterious situations. Consequently, they have high levels of satisfaction with life, even in circumstances of relatively poorer health than that of Whites.

## **6. RECEIVING CARE**

Although it has been noted that ethnic minority elderly have viable support systems, the helping literature suggests that receiving aid may have negative psychological consequences for care recipients who have health limitations. The logistical differences between receiving social support in general and receiving care for functional limitations are vague. Past research indicates that older Blacks have varied social support networks consisting of family as well as friends. But little research has examined the adequacy, satisfaction, or psychological implications of transitioning from self-sufficiency to dependency among older Blacks. The psychological effects of actual dependency differ from the psychological benefits of participating in

social support exchanges. Studies have focused on the psychological and physical effects of providing care to functionally limited elders, but little is known about care recipients' perceptions of the caregiving relationship.

It is often implicitly assumed that if care recipients with IADL limitations and ADL impairments have caregivers who adequately meet their physical needs, they are likely to be functioning satisfactorily both physically and mentally. In general, giving help is beneficial and should be encouraged; however, it is also complex and multifaceted.

Responses to aid depend on many factors, including the relationship and history of the donor with the recipient, the characteristics of the donor and the recipient (e.g., age, race, socioeconomic status, gender), and the appropriateness of aid to the needs of recipients (e.g., the context in which the aid is given). Research indicates that all of these factors can affect the recipient's sense of control. For example, if the help is threatening to feelings of personal control, it can lead to negative psychological outcomes for the recipient. However, if the help is supportive and appropriate to the needs of the recipient, it may reduce any threat to the recipient's concerns about control. Having some sense of control over who provides care and how much is provided may help to limit psychological distress.

Sense of control is a psychological resource that is effective in helping people to overcome environmental threats. Believing that one has control over negative events reduces possible adverse effects and promotes positive psychological outcomes. Control is indicative of psychological resilience and is effective in buffering the effects of stress. Perceptions of control have beneficial effects on individuals, whereas diminished control over undesirable events induces stress and anxiety and may lead to learned helplessness and depression. A sense of not being in control reduces motivation and adversely affects coping strategies, adaptation, and problem solving.

Some studies show an increase in control with age, whereas others demonstrate a decrease in control with age. The elderly usually encounter more loss and negative life events than do younger individuals. Events such as loss of one's spouse, physical limitations, and increased contact with health care providers are often associated with a decrease in control and adverse psychological outcomes. However, it is important to examine control within a specific context. Individuals do not feel the same extent of control over all sectors of their lives; degree of control varies from one aspect of

life to another. For example, control over career and social relations increases with age, whereas control over health and physical functioning decreases with age. How older people adapt to a new care-receiving situation may depend on caregiving factors noted earlier and control beliefs. Successful adjustment means better mental health outcomes. Believing that one has control over challenges and difficulties results in positive outcomes in various domains, such as psychological adjustment. Research has determined that the perception that an individual has control over his or her life is a stronger predictor of psychological well-being than is actual control. Thus, threats to perceived control in important domains of life may have particularly salient effects on overall health and well-being.

## **7. IMPORTANCE OF PERCEIVED CONTROL TO PHYSICAL AND MENTAL HEALTH OUTCOMES**

Control has been studied in relation to a number of psychological constructs and has been found to be a strong predictor of physical and mental well-being. These constructs include life satisfaction, coping, depression, mortality, and morbidity. There is a strong relationship between helplessness (i.e., the opposite of control) and depression. Sense of control over both good and bad outcomes is associated with low levels of depression. Studies also indicate that people with a high sense of control are more satisfied with life and are more willing to face challenges than are those with a low sense of control. Attributing control to luck or chance fosters uncertainty and anxiety and may lead to depression. When people believe that they have a low or no degree of control, they become passive and withdrawn. People who believe that they have a higher level of control develop habits to solve and prevent problems.

Feelings of control are especially important for vulnerable populations. Sense of control not only affects psychological functioning but also affects the actions that people pursue as well as people's cognitions and emotions. For example, control influences how people characterize an event. Those who have low control expectations will attribute declining health to the aging process and, thus, do little to change or prevent further problems. Those with a higher sense of control will participate in intervention and prevention activities to improve their health. Elders receiving care due to physical impairments or health problems may be

feeling particularly susceptible to concerns about loss of control. Sense of control is an important construct in predicting adjustment among the elderly, particularly those with ADL impairments, such as problems with cooking, housework, shopping, and walking.

The infirm elderly may find themselves with care providers who prefer compliant care recipients who easily relinquish control in exchange for care. There may be expectations that care recipients should accept decisions made for them about who provides care and how much care is provided. These uncontrollable stressors, including physical limitations and problems in caregiving, may magnify the sense of loss of control and lead to depressive symptoms for elderly individuals receiving care. To illustrate, research suggests that when older people move into their children's homes, they often experience feelings of depression and increased helplessness and dependency, a diminished sense of personal control over their environment, and a loss of control over their destinies. However, when elderly people in care-receiving situations are given responsibility for some of their daily activities, they experience higher levels of control and fewer symptoms of depression and less dependency. Elderly people experience less psychological distress when they are able to control the course of events in their lives.

For people in poor health, a high sense of control is usually associated with high levels of psychological well-being. However, it should be noted that high perceptions of control have also been linked to poorer psychological well-being in people with health problems. This may be because people with a high sense of control and health problems may feel as though they are losing their independence. Specifically in the caregiving context, recipients with physical impairments may consider the assistance they receive to be intrusive. Nevertheless, the caregiving assistance is needed. Therefore, those who are impaired and have a high sense of control may be at greater risk for depressive symptoms than are those without a high sense of control. Traditionally, people with a low sense of control depend on others for help, whereas a high sense of control is related to greater motivation for making the decisions that affect people's own lives. Care receivers with a low sense of control may be better at adapting to relinquishing control of their daily lives to their caregivers. This relinquishing of control may result in positive psychological outcomes for care recipients as well as feelings of satisfaction with their relationships with caregivers. On the other hand, people with a high sense of control might feel more independent

and perceive themselves as not giving up their autonomy and self-sufficiency in exchange for care. Consequently, these individuals might be at risk for poor care or, worse, for no care providers, eventually resulting in institutionalization.

Levels of sense of control may have unique effects on psychological well-being for ethnic minorities. Research suggests that because of past histories of discrimination, ethnic minorities might interpret some negative life events as uncontrollable. However, these interpretations are often associated with positive outcomes of psychological well-being. Racial and ethnic minorities may be well adapted to dealing with stressful negative life events and to acknowledging and accepting when they are unable to control or change the course of those events. When opportunities and resources are low, or when advantages are few, it may be healthier and more prudent to assume a low sense of control. Usually, an external sense of control is associated with unhealthy physical and mental health outcomes, but for the ethnic minority elderly, an external view may serve as a buffer to the effects of uncontrollable negative life events.

## **8. RECEIVING CARE, SENSE OF CONTROL, AND PSYCHOLOGICAL WELL-BEING**

### **8.1. Caregivers**

#### ***8.1.1. Burden, Stress, Depression as a Result of Providing Care***

The caregiving literature reveals that spouses are the first choice for providing care. Spouses are willing and committed to helping each other, but often the outcome for caregiving spouses is depressive symptoms, anxiety, stress, and physical impairment. Spouses in caregiving situations may be at advanced ages with limited physical strength and stamina for the tasks of caregiving and sole responsibility of performing household tasks and home maintenance that were previously shared. The added responsibilities may lead to burden, stress, and depressive symptoms for caregiving spouses.

Adult children are also preferred over other nonrelative care providers. However, caregiving adult children usually have other obligations, including the needs of their spouses, children, and career goals. These other demands can be negatively affected by providing care to elderly parents. Some adult children providing care for parents often forfeit or delay career advancements due

to the responsibilities of caregiving. Furthermore, adult children may experience role reversal due to caring for their parents. Role exchange, career delay, and multirole responsibilities can have negative psychological consequences, such as role strain, depression, and poor physical health, for caregivers.

Studies comparing caregivers' psychological well-being indicate that Black caregivers report better psychological well-being and less negative psychological consequences of caregiving than do White caregivers. Investigators attribute these differences to cultural norms and beliefs and to patterns of living. For example, some studies show that Black caregivers report lower levels of caregiver burden and stress, and greater levels of caregiving mastery and satisfaction, than do White caregivers. Furthermore, ethnic minority caregivers often have less social and high-status career obligations to relinquish when assuming caregiver responsibilities than do Whites. As a result, they usually have less to lose economically than do their White counterparts.

## **8.2. Care Recipients**

### ***8.2.1. Effects of Caregiving on Care Recipients' Psychological Well-Being***

Care recipients who must learn to master a new environment due to physical limitations or some other transition in their lives find that they would rather have the freedom to choose who helps, when they need help, or when they would rather do tasks themselves. When help is needed, recipients of aid are usually more willing to receive help from family and friends than from strangers. The more intimate the relationship with the helper, the better for the care recipient. Receiving help from family and friends permits care recipients to use little or no formal services and to remain in the community.

Although care recipients rely heavily on family, they also weigh the costs of time and effort to their caregivers. If the cost to the caregiver is low, the care recipient may be more comfortable in asking for or receiving help. If the caregiver is a spouse, it may be less costly than it would be for an adult child who has a family and other job and career responsibilities. However, some research has found that recipients' psychological well-being sometimes is negatively affected in caregiving contexts with both spouses and adult children. Other work has found that the elderly prefer to have less control, particularly during major

health-related difficulties. Giving up control to family or health care professionals can often be a relief if control is not possible. In most day-to-day living circumstances, the elderly desire to have a routine schedule, to be independent, and to make decisions for themselves.

The literature suggests that care recipients with spouses or other family care providers feel dependent on caregivers and depressed. In his research on receiving help, Newsom reported that recipients often have negative cognitive and emotional reactions to receiving care, including feelings of dependency, indebtedness, incompetence, and worthlessness. Furthermore, having family caregivers sometimes exacerbates those perceptions, especially if recipients perceive that family members are burdened, stressed, and depressed by providing care or if care is not appropriate for the IADL limitations (i.e., too much or too little).

### ***8.2.2. Spouse Caregiver***

The caregiving literature indicates that spousal care is less costly in terms of time constraints than is care from others. On the other hand, recipients with spouse caregivers report being more worried, depressed, and dependent than do recipients with younger relative caregivers. These negative effects can be attributed to concerns about overtaxing the elder caregiving spouse with additional responsibilities and tasks. The spouse receiving care may harbor a sense of guilt, realizing that he or she can no longer share the responsibilities of housekeeping tasks and other marital obligations and that the extra burden of caregiving and added household responsibilities could jeopardize the health of the caregiving spouse. In addition, research indicates that the quality of marital relations affects caregiving perceptions and psychological well-being for both spouses. Poorer quality relationships are exacerbated by caregiver burden and stress. These effects on the caregiving spouse may be manifested in the quality of care given and contribute to increased depressive symptoms for recipients with either a low or high sense of control.

Research shows that the quality of spousal care has a strong influence on care recipients' psychological well-being. However, the literature also presents evidence that even spousal care is not always helpful. Unhelpful care can lead to negative self-perceptions and to negative perceptions about caregivers. Spouse caregivers who over- or underestimate the abilities of care recipients can inadvertently criticize recipients'

recovery efforts. These criticisms erode self-esteem, decrease sense of control, and contribute to increased depressive symptoms in recipients. In general, spousal caregivers provide appropriate care; however, the effect of poor-quality care is more detrimental to mental health than high-quality care is beneficial to mental health. Inappropriate help can, at times, reduce the actual and perceived capabilities of recipients. Help implicitly threatens the sense of control of recipients, depending on the nature of the need and the help given.

### **8.2.3. Adult Children**

Most care recipients have good relationships with their adult children caregivers; however, there are elderly individuals with adult children caregivers who perceive themselves as burdensome to their children. In a review of the caregiving literature within Hispanic communities, it was concluded that familial support systems for the impaired elderly could have detrimental consequences for both care recipients and adult child caregivers. The more dependent elders were on their children, the more depressive symptoms the elders reported. Depressive symptoms and dependency can be a result of role reversal and of the feeling of being a burden on children. Receiving support for shopping, cooking, and household tasks from children indicates that individuals are no longer self-sufficient, causing care recipients to feel as though they are losing control of their lives and their independence. Most elderly individuals would prefer to live independently from their adult children and to receive emotional support rather than instrumental support. Care recipients fear burdening their children and experience more psychological problems receiving help when adult children have to adjust their lives, including work schedules, to provide any type of support.

In addition, with family members performing chores and tasks for care recipients, time for social interactions with recipients may be reduced. After completing caregiving tasks, adult children may lack the time or energy to show affection or emotional support, resulting in elder recipients feeling lonely as well as burdensome to their children. Lack of affection can cause negative psychological consequences for recipients, particularly if the stress and burden that family members feel as a result of providing care is manifested in their behaviors, attitudes, and quality of care toward recipients. The elderly desire to continue familial relationships, but they also want to live independently.

### **8.2.4. Other Relatives, Friends, and Church Members**

The literature indicates that the ethnic minority elderly in the community without family often have several sources of support. Studies have examined the similarities and difference in the structure, demographics, and living arrangements of elderly Black and White social and caregiving networks and found that elderly Blacks are more likely to have nonfamily and unpaid caregivers than are elderly Whites. Many elderly individuals may feel less of a burden to other relatives and friend caregivers than to spouses and adult children care providers. Nonfamily members may be providing care out of kindness and genuine concern, but the care might not be consistent, steady, or dependable. For example, neighbors or friends going to the market may offer to purchase groceries or prepare extra food, and church friends may provide emotional support and transportation. However, providing care on a daily basis, such as for bathing, dressing, and taking medication, might require more commitment than friends and neighbors are able to give. If help is required too often, care recipients may begin to feel burdensome to nonfamily members. Receiving help from nonfamily members may be distressing but necessary to overcome care needs. The familiarity found in families is usually not present with nonfamily caregivers. In addition, nonfamily members may begin to feel beset by too many requests for assistance.

Because of the need for help with physical limitations, such as eating, bathing, and walking, recipients with nonfamily caregivers might need more professional help than spouses or family members can possibly offer. Some daily tasks are more appropriate for trained formal caregivers. Having physical or medical needs met, however, does not guarantee that recipients are comfortable with the caregiving situations or with the caregivers. Elderly individuals who are reluctant to voice disapproval or who have no control (whether actual or perceived) over ineffective or excessive care are at risk for depression.

## **9. RACIAL AND ETHNIC MINORITY ELDERLY AS A RESOURCE**

The information in this article is important for practitioners dealing with elderly care recipients. Frequently, psychological well-being, such as depressive symptoms in the elderly, are ignored or misdiagnosed, resulting in early morbidity and mortality among older persons

receiving care. Research indicates that some health care professionals believe that the symptoms of depression are inevitable in the elderly and focus more on physical and medical problems. This article suggests that elderly care recipients' psychological needs are also important and that certain aspects of the caregiving relationship are related, for example, to an increase in depressive symptoms. Unfortunately, most care recipients often feel that they should be appreciative of any help they receive and will not complain, even when the help is inappropriate and causes poor psychological well-being.

Consideration should be given to elder care recipients as a resource. The elderly often possess valuable knowledge, skills, and insights that they can share with others. The elderly generally have more positive perceptions about the quality of their lives when their relationships with family and friends are of a reciprocal or interdependent nature than when they are simply receiving help from others. Elderly individuals receiving care are not necessarily unable to reciprocate in the caregiving relationship. Elderly parents are more satisfied if they can exchange support with care providers. For example, care recipients often provide emotional and financial support to caregivers, serve as confidants and sources of advice, and provide child care and sick care. The relationships between caregivers and care recipients are found to be supportive and reciprocal for both parties when recipients are given some responsibilities. Recipients can help with child care for their adult children who work and provide caregiving. The elderly can also share their skills and knowledge with family caregivers and grandchildren. They can help teach younger children to read and write, and care recipients who have hobbies (e.g., knitting) can teach skills to younger generations, further helping adult children and strengthening family intergenerational bonds. The elderly often help to relate and preserve family heritage and serve as socialization agents. Moreover, although frail or disabled older adults may be dependent on family members for care and support, there is some indication that various caregivers are dependent on care recipients for economic and housing assistance. For example, when adult children are in the divorce process or are experiencing temporary unemployment, parents are usually their first source of help.

Caregivers should also be aware of the importance of the emotional needs of care recipients and how to provide necessary care in a nonintrusive way. Care recipients should have as much control as possible over who provides care. Caregivers should allow care recipients to become involved in the planning of their daily care. For

example, recipients may help to create grocery lists to purchase foods of their liking and help to prepare meals of their choice. The literature indicates that when care recipients are allowed to make decisions about their routines and are given some responsibility for their well-being and the important events in their lives, their emotional and physical well-being improves. Recipients who must learn to master a new environment due to disabilities need the freedom to choose when they need help or which tasks they would rather perform themselves. Their perceptions of their needs may be different from their actual needs, but it should be possible to provide care in a way that is not perceived as intrusive, overwhelming, or useless. To provide the best possible experiences for both receivers and caregivers, practitioners should assess and evaluate emotional and psychological aspects, and not only physical and medical aspects, of caregiving situations.

### 9.1. Implications

One of the major implications of the literature in the area of caregiving and care receiving is that recipients can have poor psychological health even when caregivers are providing for their physical needs and IADL limitations. Recipients' psychological health is an important factor to consider when examining the caregiving environment.

As noted previously, research that has investigated care recipients and caregivers has largely addressed the medical and physical limitations of care recipients. It is erroneously assumed that if caregivers are available to help with those limitations, recipients are functioning satisfactorily. Contrary to expectations, being cared for by family for some limitations and by nonfamily members for others might not always be beneficial to recipients' psychological well-being. Sense of control may be related to decreased psychological well-being, such as depressive symptoms, and there is compelling evidence that receiving and giving care is multifaceted and complex, especially among racial and ethnic minority elders. The authors believe that more research is needed on recipients' psychological well-being, especially sense of control, within the caregiving relationship.

## 10. CONCLUSIONS

This article has reviewed some implications of research findings in the areas of socioeconomic status, social support systems, well-being, and family patterns,

focusing specifically on perceptions of control and family caregiving and care receiving. It discussed how these findings are important for an applied psychology of aging, particularly within a biopsychosocial model of human development and aging. As indicated at the beginning of the article, the population of the United States continues to grow, especially in racial and ethnic diversity. Along with this complex demographic growth are social, health, and well-being considerations that affect the structure and functioning of the entire society. A clearer understanding of the influence of ethnic, racial, and cultural dimensions on social and psychological processes of aging and human development is necessary to address this complicated change in the size, diversity, and needs of the expanding older population.

### See Also the Following Articles

Cross-Cultural Psychology, Overview ■ Elder Caregiving  
 ■ Health and Culture ■ Health Psychology, Cross-Cultural

### Further Reading

- Antonucci, T. C., & Jackson, J. (1997). The role of reciprocity in social support. In B. R. Sarason (Ed.), *Social support: An interactional view* (pp. 173–198). New York: John Wiley.
- Jackson, J. S. (1993). Racial influences on adult development and aging. In R. Kastenbaum (Ed.), *The encyclopedia of adult development* (pp. 18–26). Phoenix, AZ: Oryx.
- Jackson, J. S. (Ed.). (2000). *New directions: African Americans in a diversifying nation*. Washington, DC: National Policy Association.
- Jackson, J. S., Chatters, L. M., & Taylor, R. J. (1993). *Aging in Black America*. Newbury Park, CA: Sage.
- Lustbader, W. (1991). *Counting on kindness: The dilemmas of dependency*. New York: Maxwell Macmillan.
- Rodin, J. (1987). Personal control through the life course. In R. P. Ables (Ed.), *Life-span perspectives and social psychology* (pp. 103–119). Hillsdale, NJ: Lawrence Erlbaum.
- Taylor, R. J., Chatters, L. M., & Jackson, J. S. (1997). Changes over time in support network involvement among Black Americans. In R. J. Taylor, L. M. Chatters, & J. S. Jackson (Eds.), *Family life in Black America* (pp. 293–316). Thousand Oaks, CA: Sage.



# Aggression and Culture

Michael Harris Bond

Chinese University of Hong Kong, Shatin, Hong Kong

1. Culture and Aggression: Contexts for Exercising Coercive Control
2. Culture as Contexts for Influence
3. Holocultural Studies of Aggression
4. Culture and Studies of Aggression in Individuals  
Further Reading

## GLOSSARY

**aggression** Acts of coercive control perceived as illegitimate for the social context in which they were enacted.

**coercive control** Influence applied to another person or group that harms the person's or group's physical, material, or social interests.

**culture** Ecological–social contexts that promote or restrain acts of interpersonal or intergroup control.

**holocultural studies** Research that treats a social unit, such as the nation, as the unit of analysis and that spans as much of the world as possible.

**retaliation** Application of coercive control against a person or group deemed to have attacked oneself or one's group unjustly.

This article conceptualizes aggression as coercive forms of social control, perceived as being applied illegitimately. Mundane social control is exercised through role prescriptions and is routinely met with compliance. Whenever social control is perceived to be exercised illegitimately, however, victims and their representatives respond to restore interpersonal and

social order. That response may be retaliatory coercion. The type of response elicited will be shaped by personal factors such as the emotionality of the victim and the hedonic relevance of the anti-normative action; target factors such as the gender, status, and ethnicity of the perpetrator; social factors such as the degree of interpersonal support for retaliation or harmonizing; historical factors such as the strength of antagonistic ideologies among ethnic groups; economic factors such as the degree of relative inequality between the actors and between their social groups; and cultural factors such as the presence of honor or restraint codes. For the purpose of understanding aggression, culture may be understood as that complex set of external conditions that sustains or retards the development of those social psychological conditions that predispose toward interpersonal violence. This article elaborates these societal conditions and their psychological imprints on social actors.

## 1. CULTURE AND AGGRESSION: CONTEXTS FOR EXERCISING COERCIVE CONTROL

Cultures may be construed as ecological–social contexts that promote or restrain acts of interpersonal or intergroup control. These acts function to influence other persons or the groups they constitute to provide resources desired by the influencing agent while minimizing the costs required for their acquisition. The party being influenced is also acting to maximize benefits and

minimize costs while engaged in this dance of interdependencies. So long as this exchange of influence proceeds within normatively accepted guidelines, the process is regarded as an acceptable negotiation. Once one of the parties violates the other's norms of acceptable procedure or outcome, the other's influence attempts may shift and become more coercive.

Coercion in influence is characterized by the use of punishments or by the withholding of rewards from the target as a consequence of his or her behavior. It is these tactics of contingent control that are closely monitored in society because their use may violate social norms for exercising social influence over another's interests. Any escalation in the strength of influence tactics will typically be construed as intended by the other party but unethical, illegitimate, unjustified, savage, immoral, or wrong and will typically be met with a counterattack. One party triumphs, at least in the short term, leaving the other party subjugated and resentful, immobilized and hopeless, or dead and past harboring plans for revenge.

Harmony or at least a working truce may be restored between these interdependent parties. If that harmony has not been restored in ways that satisfy the aggrieved parties, however, memories of unrequited abuse will fester, motivating acts of revenge for past wrongdoings. These acts of retaliation may erupt if the social calculus shifts through changes in the balance of power, economic recession, environmental degradation, and the like.

In the course of this familiar drama, the label "aggressive" will often be deployed to describe anti-normative tactics of influence exercised by the other party. In general, such acts result in physical or material harm to the other party, such as wounding or removal of possessions, but can also include symbolic impositions, such as insults or threats. However, these same acts are generally not construed as aggressive by those who perform them. Such dramatically different accounts of hurtful acts given by perpetrators and by victims have been well documented. Instead, if called to account by an accepted authority, "aggressors" will justify their use of coercive control in terms of self-defense, doing their job, retaliating for past injustices—that is, getting even; serving their group, organization, or nation; protecting the social order; and the like. Aggression, then, is not a characteristic of the action itself but rather a labeling process arising out of the social context within which it occurs. That label constitutes a disapproving judgment passed on the process by which social influence over another person is exercised. Homicide, robbery, rape, and assault are

universally regarded as acts of aggression. Defending oneself against attack, performing painful surgery, executing a convicted rapist, and spanking a disobedient child are generally not so regarded because those acts are often socially approved.

Of course, the exercise of illegitimate tactics of control does not always result in retaliation by the aggrieved party. The target may instead undertake proxy control, influencing third parties—be they friends, family, or social authorities—to restrain, distract, or attack the "aggressor." Or recipients of such resented manipulation may withdraw from the relationship if an exit option is possible and if alternative relationships providing the same desired goals are available. Alternatively, they may accept such "altercasting" into a subordinate status, thereby acceding to a relationship in which the dominant party may exercise such tactics of control.

The use of any such response to counternormative control depends on the skills and resources available to the target. The target must be aware of alternative responses because not everyone has the social training and intelligence to appreciate alternatives other than retaliation. The target must be willing to use the alternative responses because some may be proscribed by socialization or normatively sanctioned by proximal social groups. The target must be able to mobilize the personal resources and the social agents necessary to exercise that form of influence because not everyone commands the required talents, time, energy, resources, and social credits with others to motivate their intervention.

## 2. CULTURE AS CONTEXTS FOR INFLUENCE

Culture is a polysemous construct of kaleidoscopic richness, variously defined by practitioners of different social sciences. It may be defined for purposes of doing psychology as a set of ecological-social constraints and affordances potentiating some behaviors, retarding others, or even rendering them irrelevant. This general psychological definition culture was elaborated by Bond as

a shared system of beliefs (what is true), values (what is important), expectations, especially about scripted behavioral sequences, and behavior meanings (what is implied by engaging in a given action) developed by a group over time to provide the requirements of living (food and water, protection against the elements, security, belonging, social appreciation, and the exercise of one's skills) in a particular geographical niche. This shared system

enhances communication of meaning and coordination of actions among a culture's members by reducing uncertainty and anxiety through making its members' behavior predictable, understandable, and valuable.

This definition focuses on the socialized psychological "software" constituting the operating system for functional membership in a group setting. That group may be any social unit—a family, a work group, an organization, a profession, a caste, a social class, a political state, and the like. The preceding definition is silent about the characteristics of those social units that predispose toward the development of different types and levels of the shared software called culture. Nor does this definition focus on the types and levels of behavior likely to characterize that social unit and its members. It is precisely those contexts and consequences of culture that need to be elaborated whenever one addresses the link between culture and any behavioral outcome such as aggression.

The cultural factors relevant for explaining differences in aggressive behavior will depend on how aggression has been conceptualized. These factors would then be linked to mediating processes responsible for driving aggressive behaviors. Given the preceding discussion, relevant considerations for a cultural analysis would include institutional supports enforcing procedural norms for coordinating action and for dividing resources; norms about the legitimacy of claims made by outsiders or different others on one's resources, including ideologies of antagonism and social representations of a group's history, focused on specific out-groups and their members; direct or indirect socialization for coercive responses to particular targets (e.g., women, the disabled, other lower status persons); training provisions and social support for conflict resolution strategies other than retaliation; the goals that social group members hold about desirable material and social outcomes and the procedures for achieving those outcomes (i.e., their instrumental and terminal values); attitudes held by citizens about the proper place of out-groups (i.e., social dominance orientation); and beliefs about the efficacy of various influence tactics, citizen neuroticism, and emotionality affecting impulse regulation. Each of these mediators is made salient in the issue of culture's relation to aggression by its hypothesized role in eliciting coercive tactics of interpersonal and social control. Whether a person or group using such coercive violent means of social control would be perceived as behaving aggressively would depend on the position of the actor and the target in the social order of their culture.

### 3. HOLOCULTURAL STUDIES OF AGGRESSION

Dramatic and universally accepted acts of aggression, such as homicide and political violence, can be studied only at the level of larger social units, such as societies and nations, because their incidence is either rare or group encompassing. Such studies are called "holocultural" because they involve social units sampled from the whole range of the earth's cultures. Aggression-relevant data on larger social units are easier to obtain than are individual-level data on cross-cultural comparisons of coercive control. Perhaps this is one reason why there is so little comparative psychological work on aggression despite its social importance. The problem with holocultural studies, however, is that it is more difficult to extract defensible social psychological explanations for their results.

#### 3.1. Waging War and Internal Political Violence

In 1994, Ember and Ember examined the ethnographic atlas for clues about the causes of war in 186 societies. They concluded that societies wage war due to recurrent natural disasters that rendered them vulnerable to food shortages: "Fear of such unpredictable shortages will motivate people to go to war to take resources from others in order to protect against resource uncertainty." The societies they studied were smaller than current nation-states, which are enmeshed in relations of trade interdependence and protected by mutual defense pacts. These state-level connections act to restrain war, although the driving force of resource acquisition remains a powerful incentive for waging war. Looming environmental disasters, with their potential for reducing current national resources such as arable land and potable water, argue for carefully coordinated international programs of disaster relief and foreign aid. In their absence, pressures for waging war will mount as some nations suffer reversals of fortune.

In 1988, Rummel assessed war making during the course of the 20th century. He concluded that the historical record shows that democratic political systems—that is, "those that maximize and guarantee individual freedom"—are less likely to engage in war. Rummel argued that "there is a consistent and significant, but low, negative correlation between democracies and collective violence" and further that "when two

nations are stable democracies, no wars occur between them.” Rummel argued that this relationship is causal and not merely correlational, stating that a democracy “promotes a social field, cross-pressures, and political responsibility; it promotes pluralism, diversity, and groups that have a stake in peace.” Institutional pressures and supports in a democracy, from the opportunity to vote anonymously to legal supports for human rights, lead to the development of psychological processes, such as egalitarian values and trust in others, that prevent political leaders from mobilizing the mass support required for waging war. The same argument applies to restraining internal political violence. In this regard, Rummel provided this sobering reminder:

War is not the most deadly form of violence. Indeed, while 36 million people have been killed in battle in all foreign and domestic wars in our century, at least 119 million more have been killed by government genocide, massacres, and other mass killing. And about 115 million of these were killed by totalitarian governments (as many as 95 million by communist ones).

He hastened to point out that during the 20th century “there is no case of democracies killing en masse their own citizens.”

The preceding discussion suggests that if the big picture on aggression is considered, concerted efforts toward educating populaces for effective democratic citizenship would be the most powerful prophylactic against violence and investment in social capital that the leaders of a nation could make. Of course, the courage and insight required for nondemocratically installed leaders to commit the resources necessary for this task and imperil their own power basis will be hard to find without external political influence being engaged.

### 3.2. Homicide

The most reliable data on levels of homicide in a social unit are provided by the standardized measures collected by the World Health Organization, although judgments of homicide levels in a society have also been taken from the anthropological record provided by the Human Relations Area Files. These homicide levels are typically correlated with societal factors to provide insights about characteristics of social systems that predispose members of their citizenry toward lethal aggression against one another. Such findings are springboards for some social psychological explanation of violence by citizens against fellow citizens. Occasionally, empirical data are

provided to support such speculations about social psychological processes mediating these acts of violence.

#### 3.2.1. Societal Factors

Most research in this area is bivariate, where one societal predictor, such as national wealth, is correlated with the rate of national homicide. Over time, a number of societal factors have been advanced as “causing” higher levels of internal murder. Theory is then advanced to explain each variable’s avenue of influence. One problem with this piecemeal approach is that many of these factors are interrelated and their avenues of influence are overlapping or perhaps identical. Understanding the full range of societal relations to homicide rate requires multivariate studies where a host of potential societal predictors is deployed, so that their separate influences may be assessed. Such studies could then confirm many of the past findings, integrating theorizing around key processes such as social trust. It could also add new predictors to the societal equation, suggesting further lines of theoretical development.

It is possible to use these fewer multivariate studies to integrate the results from the various studies that have been conducted on homicide rates over the past 40 years. Using this approach, the following societal variables have been linked to homicide rate: a society’s recent history of waging war, its level of wealth (negatively), its degree of economic inequality, its percentage of male unemployment, its strength of human rights observance (negative), and a measure of male dominance themes in its national system (labeled “female purity”). More studies have been conducted using the readily available economic indexes, so one can be more confident in their robustness of impact, especially that of relative inequality. The other variables are new entrants in the predictive arena, emerging as more national indicators become available for use by social scientists during the 21st century.

#### 3.2.2. Social Psychological Mechanisms

Investigators of societal variables related to homicide have often used their empirical results to support the probable operation of certain social psychological processes in potentiating homicide. So, for example, Wilkinson argued in 1996 that relative economic inequality predisposes the “have-nots” in a social system to feel routinely status deprived and inferior. The social stress so induced increases the likelihood of violent behavior in everyday social exchanges. Taking a complementary approach, Fukuyama argued in 1995 that the

higher levels of incivility, including homicide, in poorer societies arise through lower levels of social trust held by citizens in such polities. This higher ambient trust among a society's members may be sustained by greater equality in resource distribution because the two features of a social system are highly correlated.

In their 1994 analysis, Ember and Ember did more than speculate on the social psychological processes involved in homicide; they empirically demonstrated that the societal link between a recent history of warfare and a society's internal homicide rate is mediated by the stronger socialization of males for aggression in such societies. This demonstration is accomplished by using regression techniques that enable the interrelatedness of societal and social psychological variables as predictors of homicide to be assessed within the same data set. The richness of the linkages so revealed, however, depends on the types of social psychological variables included in the data set and their overlap with the nations whose homicide rates are being studied. Because these data sets consist of nation scores, considerable data collection must be undertaken to provide the necessary number of "citizen" averages on the social psychological predictors. This multicultural database is becoming more and more available with the increasing activity in cross-cultural psychology, so the linking of societal factors

to psychological dispositions of a citizenry to national rates of violent outcomes is becoming possible.

Using one such database, Lim and colleagues found that national rates of homicide across 56 nations were jointly predicted by a nation's level of economic inequality, its gross national product (GNP) growth over a decade (negative), and its emphasis on female purity (Fig. 1). This finding confirmed the results of many previous studies of economic inequality and homicide but extended them to include additional societal predictors of internal violence. Lim and colleagues also found that citizen scores on the psychological constructs of belief in fate control and the length of typical emotional experiences predicted national rates of homicide. Using regression techniques, they concluded that the effect of female purity was mediated by citizen endorsement of fate control, that is, the belief that one's outcomes were shaped by one's predetermined destiny. This linkage led to speculation that in societies characterized by social divisions and rigidity of social structure, parents typically socialize their children for responsiveness to external control rather than for internal self-regulation. Lacking strategies for internally generated self-control, citizens in such societies are more likely to respond with homicide to the universal frustrations arising out of mundane interpersonal interdependencies.

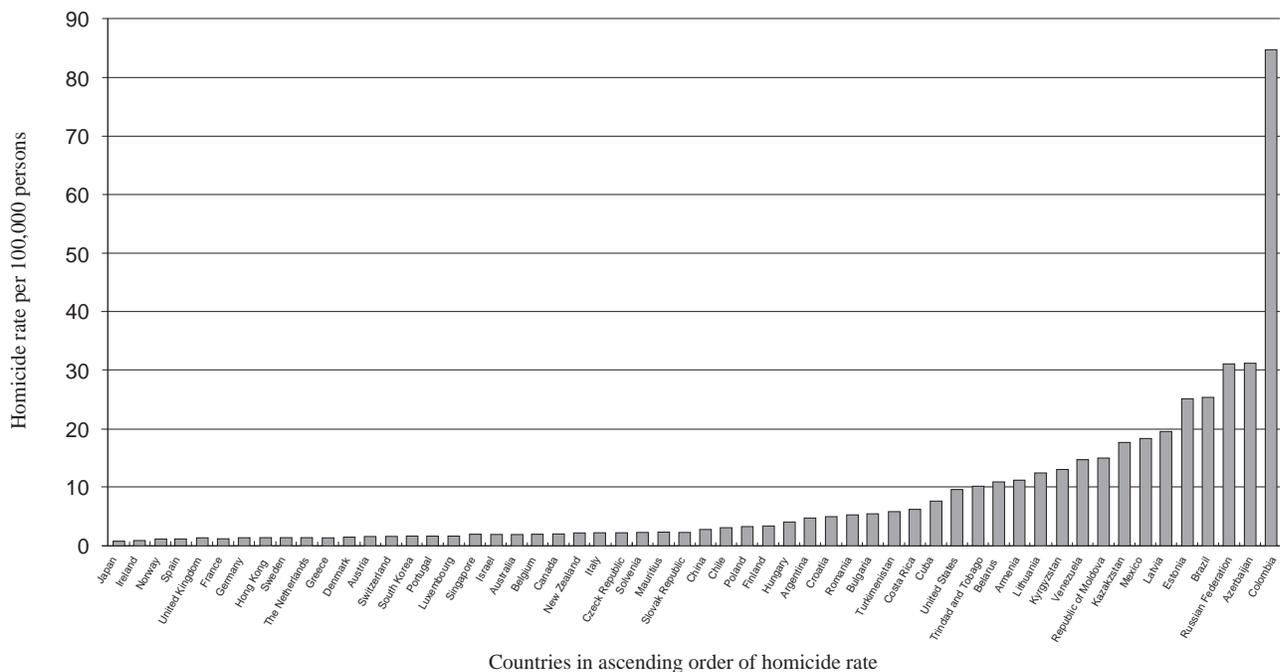


FIGURE 1 Homicide rates in 56 countries.

In nations where citizens routinely experience emotions for longer periods of time, there would be a persistence of emotionality into more of one's daily interactions with others. This wider reach of emotionality would amplify whatever dominant tendencies characterized a particular exchange. It is easy to understand how, if that exchange involved coercive control or resistance to its imposition, the frequency of homicide as one extreme outcome would be higher in such countries.

Such conclusions are speculative because they are drawn from correlations of national-level scores. In the case of homicide, however, there is little choice but to use data from large social units. The plausibility of the conclusions drawn will depend on their connectability to theory and empirical data derived from analyses of individual behaviors related to homicide. For example, a person's capacity for self-regulation has been linked longitudinally to lower frequencies of antisocial acting out. So, it is plausible that belief in control by fate is part of a personality constellation related to poor impulse control. Psychological studies of individuals differing in their beliefs about fate control will be necessary, however, to strengthen the plausibility that citizen differences in fate control link to national homicide rates through the agency of impulse regulation.

#### **4. CULTURE AND STUDIES OF AGGRESSION IN INDIVIDUALS**

The holocultural studies discussed previously alert social scientists to the need for validated theories about the psychological mechanisms involved in generating violent behavior against others. These mechanisms must be measurable across cultural systems, and their role in potentiating harm to others must be demonstrated empirically. This is the mandate for cross-cultural psychology, but its offerings are remarkably scanty. In part, this shortfall arises due to the definitional problems noted earlier (e.g., what behaviors other than homicide, rape, or physical assault constitute aggression?) There are a plethora of American laboratory studies, for example, that use paradigms such as the teacher–learner shock scenario to examine aggression. In such a setup, the participant-teacher trains her confederate-student under various experimental conditions hypothesized to provoke higher levels of hostility (e.g., ambient temperature, expectation of future interaction, prior insult by the confederate-student). This procedure, however, fails to engage the dynamics of aggression as anti-normative

behavior because the experimenter has explicitly legitimized the delivery of noxious stimuli to the student as an educational exercise. Other experimental scenarios for studying aggression and its facilitating conditions likewise encounter the same problem of its implicit legitimation by the experimenter.

A related problem for the cross-cultural laboratory study of aggression involves the meaning of the exchange between the two parties created by the experimenter. In the teacher–learner scenario, one could imagine that in more hierarchical cultures with a tradition of corporal punishment in classrooms, the legitimation of hurtful behavior by the experimenter would appear to be more acceptable to the participant-teacher. The confederate-student would then receive even higher levels of shock for a given set of experimental conditions. One could not, however, conclude that persons of this hierarchical culture were more aggressive given that local understanding of the shock-behavior is so different in the new culture.

Given these conceptual problems, two ways forward have been tried. One approach across cultural groups involves the use of scenario studies where cultural informants are presented with descriptions of behavioral exchanges between persons and are asked to make judgments about these exchanges and the persons involved. Knowledge about the cultures can then be used to predict the outcomes of these judgments. So, for example, Bond argued that in a more hierarchical culture, a direct coercive tactic of influence in a business meeting would be construed differently than it would in an egalitarian culture, depending on its source. As predicted, a superior insulting a subordinate was less sanctioned in Chinese culture than in American culture, and a subordinate in Chinese culture was more strongly sanctioned for insulting a superior than was a subordinate in American culture for insulting a superior. Similarly, knowledge about cultural dynamics can be used to predict the frequency and judged effectiveness of direct coercive tactics of influence, depending on their source and their relationship with the target. In such studies, the behavior being judged is the same, and its different meaning is examined in light of different cultural dynamics. The perceived aggressiveness of the behavior and the sanctions it receives vary as a function of these dynamics.

The second approach involves observation of hostile behavior in common settings such as prisons and schools. These observational studies are time-consuming and require agreement on what behaviors to measure. There has been considerable debate across participants

in the European Union, for example, as to what behaviors constitute "bullying," so that cross-national comparisons can proceed. Typically, a "least common denominator" position that includes only physical or verbal provocation between perpetrator and victim is adopted. Other indirect forms of bullying, such as gossiping and cliquing against others, are excluded. This happens in part because persons from different cultural traditions might not regard these behaviors as bullying, but also because the indirect behaviors are extremely difficult to measure reliably. Given all of these concerns, few cross-cultural studies involving *in situ* measures of behavior have been conducted to date.

Less conceptual and measurement difficulty is encountered when behaviors are measured by asking knowledgeable observers, such as parents and teachers, to rate the frequency with which their charges engage in aggressive behaviors. Chang, for example, employed a 7-item scale measure of aggressive behaviors such as hitting, teasing, and shoving others. The scale forms a coherent single measure of aggressiveness, and independent observers of a given pupil rate that pupil's aggressiveness with high degrees of agreement. It would be easy to examine this measure in other cultural contexts for its equivalence, adding or subtracting behaviors until a comparable measure of pupil aggressiveness is achieved. Cultural groups could then be compared for their levels of aggressiveness.

There is a slowly growing body of cross-cultural research using the scenario and behavior-reporting approaches. The problem with this work is that it is often only descriptive, reporting the cultural differences and perhaps speculating why they occur. There is little theory-driven work that measures the processes hypothesized to drive the aggressive behavior in the cultural groups studied. One exception is a recent study by Chow and colleagues using recollections by cultural informants of upsetting interpersonal events. They argued that everyday aggressive behavior often emerges out of social exchanges where one party harms another party, resulting in retaliation from the aggrieved party. Self-reports of behaviors undertaken by the victim after the harm doing indicated an equivalent cluster of assertive or aggressive behaviors in both Japanese and American informants. This assertive counterattack is driven by psychological factors uniting to generate a motive to retaliate and by social factors relevant to the relationship such as the gender of the two parties and their degree of familiarity.

Chow and colleagues found that both internal psychological and external social factors contributed toward

assertive counterattack in both Japanese and American cultures but that psychological factors weighed more heavily than social factors for the Americans than for the Japanese. This finding confirmed theoretical speculation that collectivists, such as Japanese, are more responsive to external forces, whereas individualists, such as Americans, are more responsive to internal forces. This reasoning was supported by personality measures of individualism that predicted the tendency to emphasize psychological factors more than social factors in directing one's retaliation against the harm doer. Controlling for the level of harm experienced, the investigators found that Americans were more likely to counterattack assertively because their motivation to retaliate was greater than was the Japanese motivation.

The next step in such research on culture and aggression is to examine why the motivation of any cultural group to counter any harm with coercive retaliation is greater than that of another cultural group. One strong possibility is the experience of harsh parenting. Punitive parents model coercive strategies of control as they manage their children, indirectly teaching children to respond to interpersonal frustration with confrontational assertiveness. This linkage between harsh parenting and children's aggressiveness has been confirmed in other settings such as school. Such studies have provided the inspiration for successful intervention studies that reduce the level of child aggression through programs teaching parents to intervene nonviolently.

Chang further discovered that some of the linkage between harsh parenting and child aggression is mediated by lower levels of affect regulation characterizing those pupils who act out. Such internal dispositions may be higher in citizens of some cultural groups than in citizens of other cultural groups due to differential socialization styles and procedures. The different strengths of these aggression-related attributes could then account for observed cultural differences in aggressive responding. Other possible internal mediators include well-studied psychological dispositions such as self-regulatory efficacy, sociopathy, and state anger. Social psychological factors varying across cultures could include orientation toward confrontation, honor sensitivity among males, degree of ethnic identification, social support for retributive as opposed to restorative procedures of justice, and socialization into ideologies of antagonism and indebtedness.

Higher cultural levels of these psychological and relational dispositions may be the legacy of cultural circumstances (e.g., a history of war, tribal conflict, or colonialism), a herding economy, slavery, or internal political violence, any of which may lead parents to

socialize their children to inflict harm on others as a way of solving problems in interpersonal and intergroup coordination. As indicated earlier in the discussion of holocultural studies, psychologists have some ideas about what these cultural circumstances might be. But they need to conduct the sort of theory-driven and empirically validated research necessary to substantiate their speculations about how culture operates to affect aggression. Psychologists must instrument the hypothesized dispositions, measure them in persons, and then observe their behavior when dealing with interpersonal and intergroup relationships. Who chooses cooperation and consensus? Who chooses confrontation and conflict? How then can life be enhanced by encouraging social systems to socialize their citizens for the former rather than for the latter?

### See Also the Following Articles

International Conflict ■ Interpersonal Conflict

### Further Reading

- Bond, M. H. (2004). Culture and aggression: From context to coercion. *Personality and Social Psychology Review*, 8, 62–78.
- Chang, L. (2002). *Emotion regulation as a mediator of the effect of harsh parenting on child aggression: Evidence from mothers and fathers and sons and daughters*. Unpublished manuscript, Chinese University of Hong Kong.
- Chow, H., Bond, M. H., Quigley, B., Ohbuchi, K., & Tedeschi, J. T. (2002). *Understanding responses to being hurt by another in Japan and the United States*. Unpublished manuscript, Chinese University of Hong Kong.
- Dishion, T. J., Patterson, G. R., & Kavanagh, K. (1992). An experimental study of the coercion model: Linking theory, intervention, and measurement. In J. McCord, & R. Trembley (Eds.), *The interaction of theory and practice: Experimental studies of interventions* (pp. 253–282). New York: Guilford.
- Ember, C. R., & Ember, M. (1994). War, socialization, and interpersonal violence: A cross-cultural study. *Journal of Conflict Resolution*, 38, 620–646.
- Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. New York: Simon & Schuster.
- Lim, F., Bond, M. H., & Bond, M. K. (2002). *Social and psychological predictors of homicide rates across nations: Linking societal and dispositional factors*. Unpublished manuscript. Chinese University of Hong Kong.
- Rummel, R. J. (1988, June). *Political systems, violence, and war*. Paper presented at the U.S. Institute of Peace Conference, Airlie, VA. (Available: [www.shadeslanding.com](http://www.shadeslanding.com))
- Segall, M. H., Ember, C. R., & Ember, M. (1997). Aggression, crime, and warfare. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology* (Vol. 3, pp. 213–254). Needham Heights, MA: Allyn & Bacon.
- Sidanius, J. (1993). The psychology of group conflict and the dynamics of oppression: A social dominance perspective. In W. McGuire, & S. Iyengar (Eds.), *Current approaches to political psychology* (pp. 183–219). Durham, NC: Duke University Press.
- Staub, E. (1989). *The roots of evil*. Cambridge, UK: Cambridge University Press.
- Tedeschi, J. T., & Bond, M. H. (2001). Aversive behavior and aggression in cross-cultural perspective. In R. Kowalski (Ed.), *Behaving badly: Aversive behaviors in interpersonal relationships* (pp. 257–293). Washington, DC: APA Books.
- Tedeschi, J. T., & Felson, R. B. (1994). *Violence, aggression, and coercive actions*. Washington, DC: American Psychological Association.
- Wilkinson, R. G. (1996). *Unhealthy societies: The afflictions of inequality*. London: Routledge.



# Aging and Competency

Jennifer Moye

Brockton VA Medical Center, Boston, Massachusetts, USA

1. Clinical Versus Legal Competency
  2. Competency Versus Capacity
  3. Legal Frameworks for Aging and Competency
  4. Decision-Making Abilities in Competency
  5. Clinical Evaluations of Capacities
  6. Categories of Capacities
- Further Reading

## GLOSSARY

**autonomy** A concept in biomedical ethics that refers to an individual's right to self-determination and to make decisions and choices concerning one's own life.

**beneficence** An ethical principle important to consider in guardianship issues that refers to the obligation to intervene to positively benefit another individual and prevent harm.

**capacity** Another word for competency referring to the ability to receive and evaluate information and to make and communicate decisions.

**competency** A legal status presumed for all adults—as opposed to incompetency, wherein a person with a clinically diagnosed condition is unable to receive and evaluate information or to make and communicate decisions to such an extent that he or she lacks the ability to meet essential requirements for health, safety, and self-care; judicial determinations of incompetency are not necessarily permanent and may be revoked if and when the ward's mental status changes.

**conservatorship** A legal mechanism established by a court after a hearing that empowers one party to make financial decisions (e.g., management of assets, businesses, contracts, wills, and gifts) for another party.

**durable power of attorney (DPA)** An alternative to guardianship in which a competent individual designates another person to manage his or her affairs (medical and/or financial) in the event of incapacity.

**guardianship** A legal mechanism established by a court that designates a party to make decisions (e.g., health, financial, residence) for another (incompetent) individual.

**health care proxy** Any person appointed by another competent individual to have the authority to make health care decisions if and when he or she becomes incapable of making such decisions; the proxy's authority may be unspecified (to make all decisions in accordance with the principal's best interests) or may be specified (to make decisions in accordance with advance directives).

**representative payee** An individual appointed by a benefit provider (e.g., veterans disability, social security disability) when the provider questions the benefit recipient's ability to manage the funds; the payee receives and has responsibility for only the funds distributed by that provider.

Many people use the concept of competency in clinical settings, although there continues to be confusion about what it means, when it is relevant, and how it should be assessed. All adults are presumed to be legally competent unless otherwise determined in a court of law. An adult's competency may be questioned when a family member or other interested party files a petition for guardianship of that individual. If incompetency is found in a court hearing, a guardian or conservator may be appointed. In most states, a guardian is responsible for decisions regarding the care of the individual's "person" (e.g.,

day-to-day issues of care, residence, and activities), whereas a conservator is responsible for decisions regarding the care of the “estate” (i.e., management of assets).

individual strengths and weaknesses or functional capacities rather than the all-or-none notion of incompetency.

### 1. CLINICAL VERSUS LEGAL COMPETENCY

Although clinicians may question a patient’s competency status and may request other clinicians to evaluate the patient’s competency, a clinical finding of incompetency should not be confused with a judicial determination of incompetency. The clinical use of the term “competency” refers to a clinical opinion regarding the patient’s decisional capacities. Such an opinion may be used to activate a durable power of attorney or health care proxy if one exists. However, clinical statements about capacities should not be equated with court-determined findings of incompetency.

### 2. COMPETENCY VERSUS CAPACITY

Increasingly, the term “capacity” is preferred to competency. The use of capacity recognizes the possibility of

### 3. LEGAL FRAMEWORKS FOR AGING AND COMPETENCY

In the United States, most statutory definitions of incapacity to care for one’s person or estate include four parts, as shown in Fig. 1. This four-part definition means that a diagnosis is a necessary, but not a sufficient, component of incompetency. Rather, information must be provided as to how the disease affects attention, memory, information processing, and so forth as well as, in turn, how these symptoms affect the person’s ability to do certain things or make specific decisions. The ability level should be considered within the context of the environmental demands and resources as well as specific situational risks and benefits. For example, the threshold for the capacity to reasonably manage a small amount of day-to-day cash spending would be different from the capacity to manage a complex portfolio of investments, and the abilities needed for the capacity to consent to a relatively harmless medical intervention (e.g., a flu shot) would be much lower than those needed to consent to a more risky medical intervention (e.g., heart bypass surgery).

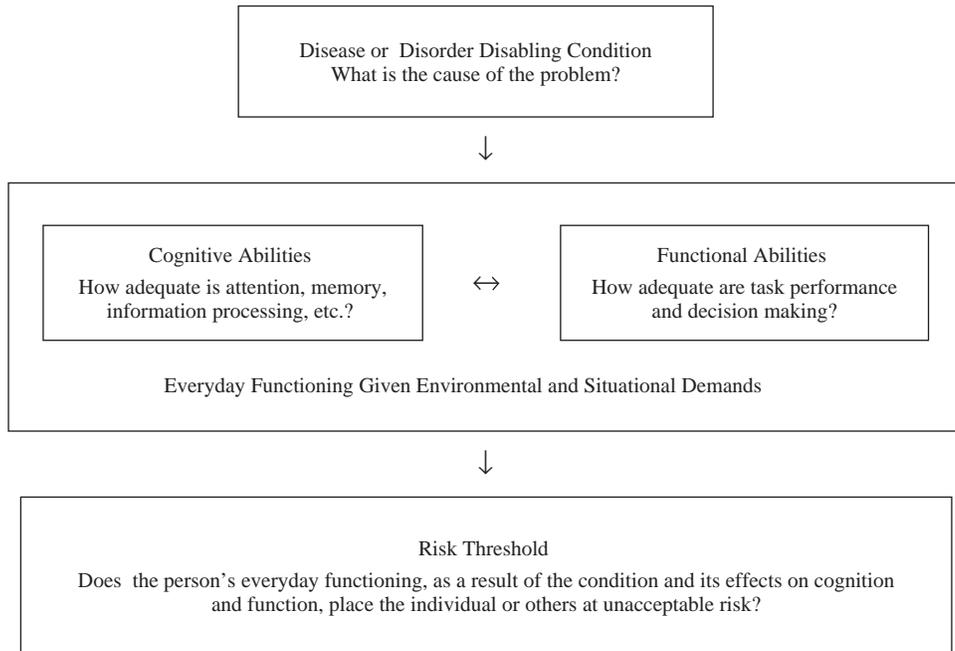


FIGURE 1 Four components of most U.S. statutory definitions of competency.

#### 4. DECISION-MAKING ABILITIES IN COMPETENCY

Much U.S. case law about competency to consent to treatment refers to four decision-making abilities, as shown in Table I. Case law typically refers to these abilities individually as independent legal standards. Together, they form a useful framework for evaluating decisional processes.

#### 5. CLINICAL EVALUATIONS OF CAPACITIES

##### 5.1. Competency and Older Adults

Although the large majority of guardianships concern elderly individuals with psychiatric or neurological diagnoses, it is clear that advancing age or physical frailty themselves are not grounds for guardianship. Competency questions are increasing for older adults due to the age-associated increase in prevalence of dementia that can affect competency during the later stages of the disease.

TABLE I  
Four Decision-Making Abilities Important for Legally Competency

Capacity	Definition
Understanding	The ability to comprehend diagnostic and treatment-related information and to demonstrate that comprehension
Appreciation	The ability to determine the significance of treatment information relative to one's own situation, focusing on beliefs about the actual presence of the diagnosis and the possibility that treatment would be beneficial
Reasoning	The process of comparing alternatives in light of consequences through integrating, analyzing, and manipulating information; also defined as providing "rational reasons"
Expressing a choice	The ability to communicate a decision about treatment that applies to individuals who cannot express a choice and also to those who are consistently ambivalent

##### 5.2. Competency Assessment

In the past, most clinical evaluations of competency relied on a physician's interview or knowledge of the patient. However, such approaches can be unreliable. A more objective evaluation of competency, particularly important in more complicated cases, involves three parts, as shown in Table II.

##### 5.3. Ethical Issues

###### 5.3.1. Consequences of Guardianship

Many adults who receive guardianship experience benefit from the arrangement. In the best case, the

TABLE II  
Components of Competency Assessment

Component	Definition
Clinical interview	
Who	Patient and relevant informants
What	Patient's values, goals, and preferences Situational factors Mental health conditions Symptoms of other illness
How	Open-ended interview Psychiatric rating scales
Structured assessment of cognition	
Who	Patient
What	Attention, memory, executive function, language, visuospatial abilities
How	Cognitive screening tests Neuropsychological tests
Performance based assessment of the specific capacity in question	
Who	Patient
What	Capacity to do the task or make the decision being questioned
How	Activities of daily living and instrumental activities of daily living instruments Capacity assessment tools

individual receives needed care, supervision, and advocacy in accordance with his or her wishes. In the worst case, guardians may take advantage of the individual's assets. In either case, removal of competency results in the loss of the right to make choices about residency, health care, medication, relationships, marriage, contracts, voting, driving, use of leisure time, and how one's own money is spent. Such losses in autonomy may affect the individual's emotional well-being. These potential rights deprivations and personal consequences of incompetency underscore the special care that should be taken in competency assessments and the extra attention that should be devoted to ethical concerns.

### 5.3.2. Informed Consent

An individual who is the subject of a competency evaluation needs to be informed of the nature of the evaluation and the potential consequences. Obtaining informed consent for the evaluation of competency from a person whose decision-making capacity is in question requires special care. Clinicians should fully disclose information about the evaluation and its risks and benefits in clear and direct language and then should assess the individual's ability to understand, appreciate, and reason through the information and options.

### 5.3.3. Individual and Cultural Differences

Persons from different ethnic or age backgrounds may have different cultural and cohort preferences for medical care, finances, living situations, relationships, and so forth. Evaluators must be sensitive so that individual and cultural differences are not misconstrued as incompetent decisional processes. This also applies to the use of standardized tests because some do not have adequate normative data concerning cultural, cohort, and educational influences on test performance, especially in older populations. Test instruments should be given in the primary language of the person being assessed.

## 6. CATEGORIES OF CAPACITIES

### 6.1. Financial Management

As shown in Table III, one important category of competency is the capacity to manage finances, including managing assets and how one spends money, managing debts and how one pays bills, and managing specific issues of contracts, disposition of property, and wills. Such financial capacities may involve knowledge of facts (e.g., where one's bank account is), financial skills (e.g., counting change), and financial judgment (e.g., avoiding

TABLE III  
Categories of Competency

Category	Possible tasks	Potential services
Finances	Managing assets	Bill-paying services
	Paying bills	Money management services
	Writing will	
Health	Medical decision making	Visiting nurse
	Health care management (including medications)	Pill box and pill-dispensing systems Telephone reminder systems
Independent living	Household cleaning and maintenance	Homemaker services
	Laundry	Meals on wheels
	Meal shopping and preparation	Emergency call systems
	Communication	Home health aide
	Personal hygiene	Care management Adult protective services Assisted living
Transportation	Driving	Adult foster care
	Use of public transit	Skill training classes
		Rides to appointments and services Assisted public transportation

fraud). As indicated in the table, services can sometimes provide alternatives to guardianship.

## 6.2. Health Care

Another category of competency focuses on health care, especially the capacity to consent to treatment. Health care capacities may also cover managing day-to-day health such as nutrition, wound care, and medications.

## 6.3. Independent Living

A third and more broad category of competency concerns the capacity to live independently. Tasks in this category include household cleaning and maintenance, laundry, meal shopping and preparation, and communication (telephone and mail). Senior service agencies have a wealth of services to assist vulnerable adults in living independently, including homemaker services to clean, do laundry, and prepare meals; “chore” services to tackle larger tasks such as lawn care and snow removal; and meal delivery programs. Communication aids, including large-button telephones, assistive devices, and animals for the deaf and hard of hearing, may also promote independent living. Medical alert systems can be installed for seniors to enact in the event of a disabling emergency.

## 6.4. Transportation

A more narrow category of functional abilities concerns transportation and the capacity to drive a motor vehicle. This question may arise outside of other competency concerns when an adult with a disabling psychiatric or neurological condition has a series of motor vehicle accidents.

### See Also the Following Articles

Cognitive Skills: Training, Maintenance, and Daily Usage  
 ■ Decision Making ■ Elder Caregiving ■ Ethics and Social Responsibility

## Further Reading

- Department of Veterans Affairs. (1997). *Clinical assessment for competency determination: A practice guideline for psychologists*. Milwaukee, WI: National Center for Cost Containment.
- Edelstein, B. (2000). Challenges in the assessment of decision-making capacity. *Journal of Aging Studies*, 14, 423–437.
- Earnst, K. S., Marson, D. C., & Harrell, L. E. (2000). Cognitive models of physicians' legal standard and person judgments of competency in patients with Alzheimer's disease. *Journal of the American Geriatric Society*, 48, 919–927.
- Grisso, T. (2002). *Evaluating Competencies* (2nd ed.). New York: Plenum.
- Grisso, T., & Appelbaum, P. S. (1995). The MacArthur Treatment Competence Study III: Abilities of patients to consent to psychiatric and medical treatments. *Law and Human Behavior*, 19, 149–174.
- Marson, D. C., Ingram, K. K., Cody, H. A., & Harrell, L. E. (1995). Assessing the competency of patients with Alzheimer's disease under different legal standards. *Archives of Neurology*, 52, 949–954.
- Marson, D. C., McInturff, B., Hawkins, L., Bartolucci, A., & Harrell, L. E. (1997). Consistency of physician judgments of capacity to consent in mild Alzheimer's disease. *Journal of the American Geriatric Society*, 45, 453–457.
- Moye, J. (1999). Assessment of competency and decision making capacity. In P. Lichtenberg (Ed.), *Handbook of geriatric assessment* (pp. 488–528). New York: John Wiley.
- Moye, J., & Karel, M. J. (1999). Evaluating decisional capacities in older adults: Results of two clinical studies. *Advances in Medical Psychotherapy*, 10, 71–84.
- Park, D. C., Morrell, R. W., & Shifren, K. (Eds.). (1999). *Processing of medical information in aging patients: Cognition and human factors perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Seckler, A. B., Meier, D. E., Mulvihill, M., & Cammer Paris, B. E. (1991). Substituted judgment: How accurate are proxy predictions?. *Annals of Internal Medicine*, 115, 92–98.
- Smyer, M., Schaie, K. W., & Kapp, M. B. (Eds.). (1996). *Older adults' decision making and the law*. New York: Springer.
- Zimny, G. H., & Grossberg, G. T. (1998). *Guardianship of the elderly*. New York: Springer.



# Aging and Culture

Toni C. Antonucci and Hiroko Akiyama

University of Michigan, Ann Arbor, Michigan, USA

1. Introduction
  2. Aging and Demographics
  3. Aging, Culture, and Families
  4. Cultural Diversity and Aging
  5. Age, Culture, and Social Relations
  6. Cultural Influences on the Health of Aging People
  7. Summary and Conclusions
- Further Reading

## GLOSSARY

**aging** Becoming old and showing the effects or the characteristics of increasing age.

**culture** The totality of socially transmitted behavior patterns, norms, beliefs, institutions, and all other products of human work and thought.

**demographics** The statistical characteristics of human population.

**families** All members of a biologically or legally related group usually living in one household.

**health** The condition of being sound in body, mind, and/or spirit; it especially refers to freedom from physical disease and pain.

**immigration** Entering and settling in a country or region to which one is not native.

**life course** The interval of time between birth and death.

Although the world population has been aging at an unprecedented rate, the speed of change is considerably

different depending on national history and life circumstances. The association between age and culture can be best understood in the light of the life span perspective and the family. Different cultural backgrounds affect individuals' aging experiences contemporaneously and cumulatively. Family is both the conveyor and interpreter of culture. The life span attachment model and the convoy model emphasize that close relationships accumulate over time and differentially affect the aging process and, consequently, the physical and psychological well-being of individuals who live in different cultural milieus.

## 1. INTRODUCTION

The study of aging has a long and ancient history, with notable comments having come from historical luminaries such as Aristotle, Socrates, and Plato. Despite the increasingly youth-oriented society, the modern study of aging has garnered increased attention during recent years for many reasons, not the least of which is recognition of the significantly changing world demography. This article first presents a brief overview of these changes. It then considers the association between aging and culture, noting the role of family and cultural diversity as well as the implications of both family and culture for health and well-being among the elderly. The article closes with a consideration

of how culture shapes aging and health in the United States as well as around the world.

## 2. AGING AND DEMOGRAPHICS

Changing world demographics are nowhere as dramatic as they are with respect to the age structure of the world population. According to the UN Populations Division and the U.S. Census, the world is graying at an unprecedented rate. For example, comparison of the percentages of people who will be age 60 years or over in the years 2005 and 2030 indicates that approximately 20% of the European population will be at least 60 years old in 2005 and that 30% of the population will share that status by 2030. The situation in the United States is not quite as dramatic, but the increase is still noteworthy. In 2005, 17% of the U.S. population will be age 60 years or over, and in the year 2030, one-quarter of the U.S. population will be in that age bracket. The figures are much smaller for Latin America, where 9 and 15% of the population are expected to be at least 60 years old in 2005 and 2030, respectively. The lowest figures are for Africa, where only 5% of the population is currently age 60 years or over and a mere 7% is expected to reach that age bracket by 2030. The AIDS epidemic is actually reducing the life expectancy in some African countries. These differences make it clear that the age distribution of the world is changing differently, depending on national history and current life circumstances. Interestingly, and perhaps contrary to the assumptions of most people, the United States ranks only 28th in "aging" among the oldest countries of the world, with most European countries and Japan reporting older populations. Although these changing age structures reflect increased longevity, this longevity is due mostly to changes at the other end of the life span. That is, fewer children are dying in childbirth, thereby increasing the numbers of people who will live to old age, and fewer children are being born, thereby reducing the proportion of a population of young people and naturally increasing the proportion of older people in each society. The decline in fertility has reached critical proportions in many countries, including Italy, Greece, Sweden, Japan, and Spain. Although fertility has been declining in the United States, it has not reached the crisis levels being experienced in other parts of the world.

## 3. AGING, CULTURE, AND FAMILIES

The effort to understand the association between age and culture must include consideration of both the life span perspective and the family. The life span perspective incorporates a long-term view of the growth and maturity of individuals throughout their lifetimes on multiple dimensions. Development on these dimensions is influenced by previous developmental periods that can manifest through stability or change as well as through continuity or discontinuity. Consequently, individuals become more different from each other as they develop over time. This is noteworthy because life span development both affects and is affected by the cultural milieu within which individuals live. During recent times, those cultural milieus have been changing in unprecedented ways in the United States as well as around the world. Examples of cultural milieus include specific individual experiences (e.g., ethnic identity) as well as more universal experiences (e.g., the unprecedented advances in technology experienced during recent years).

Although the life span perspective focuses on the individual, any effort to understand the aging experience must also include the families within which most people age. The family usually surrounds an individual with a protective convoy that nourishes and supports the individual as he or she faces the challenges of life. The family also is often the embodiment of the cultural perspective that the individual values. Just as each individual emerges as an older person who has evolved throughout his or her lifetime, so too does the individual bring that lifetime of growth and development to the family. At the same time, the family also evolves. A family might begin as a young couple who are the parents of a small child and evolve to a family composed of an adult child with aging parents and later to a family that includes grandparents with adult children and grandchildren. The family can be thought of as the vessel within which the individual experiences culture. Although people experience culture at multiple levels—as an individual, as a family member, and as a member of society—it is the family that most often is both the conveyor and interpreter of the culture. The United States may not be the oldest country in today's industrialized world, but the increased longevity of Americans means that most individuals will at some point be members of three-, four-, or even five-generation families. Hence, the role of the older person as a

family member in both promoting and maintaining culture has been steadily increasing.

#### 4. CULTURAL DIVERSITY AND AGING

The United States has, at least ostensibly, always embraced a commitment to the blending of many cultures. This phenomenon is perhaps best exemplified in the term “melting pot.” This term was used to refer to the blending of many cultures, but these cultures were mostly Western European. Although immigrants also came from China (e.g., to work on the railroads) and from Africa (e.g., to work as slaves), those immigrants were generally not among the “blended.” An important new development in the cultural influences on aging in the United States is the recent immigration from different parts of the world, namely, Central and Latin America, the Middle East, and Southeast Asia. The nature of this immigration is important not only because it represents different cultural influences but also because the immigration patterns are affecting how the United States ages. As noted earlier, the country is aging rapidly, but particularly important is the fact that among certain ethnic and racial minority groups (i.e., the “new” immigrant groups), aging is taking place at an even more rapid pace because more older people are immigrating to be with their children. Hence, as the U.S. population ages, it is also becoming more diverse both racially and ethnically. This will pose a challenge as different immigrant groups seek to maintain their identities while at the same time embracing the American culture.

Several kinds of aging experiences that will be directly influenced by cultural background and perspective can be identified. For example, some older people came to the United States as young immigrants and are now aging. Having arrived during childhood or early adulthood, they are familiar with the American culture and understand it, often while maintaining some “old world” values. At the same time, their children and grandchildren were most likely born in the United States and have embraced only American values. Some problems may arise if, as these individuals age, they identify more strongly with their cultures of origin, especially in terms of their expectations about the role of families and the experience of aging.

Potentially even more problematic is the phenomenon of immigrating in old age. Some people came to

the United States to join their children, who had immigrated earlier and settled in the country. The motivation could be emotional ties (i.e., a desire to be close to their children as they age), or it could be based on pragmatic need (i.e., the expectation that children will care for their aging parents). The potential for conflict in such families is even greater because the elderly immigrants are less likely to have embraced American values; therefore, the situation may exist where parents and their adult children have fundamentally different expectations about the aging experience.

It is clear that people, whether they recognize it or not, experience age through a cultural lens. Thus, the changing racial and ethnic population of the United States represents a unique group of individuals. These different cultural backgrounds affected their individual lifetime experiences as children, adolescents, and adults. It will, of course, also affect their aging experiences both cumulatively and contemporaneously. Each cultural group has a unique history. For example, consider the case of Cuban immigrants in Miami, Florida, who fled Fidel Castro’s dictatorship and arrived in the United States feeling forcibly expelled from their homeland. When they arrived in the country, they found very sympathetic Americans who believed that Castro was an evil and unjust dictator who was treating Cuban people unfairly. On the other hand, Middle Easterners also have come to the United States to escape an unpopular dictator, but during recent times, Americans have become more hostile, rather than less hostile, toward people from the Middle East. In addition, despite the fact that Middle Easterners also came to escape an unjust dictator, war, and oppression, they are currently viewed with suspicion, unkindness, and resentment. But another current example indicates that these more recent reactions are not ubiquitous. Interestingly, although the United States is experiencing increased hostilities with North Korea, this hostility does not seem to extend to Koreans in this country. They also fled a homeland torn by war, but somehow Americans are more accepting of their presence and their culture. The attitudes of the American public certainly influence the aging experience of both early and later immigrants.

Although these comments have addressed the general concept of culture, there are culture-specific expectations and beliefs concerning health and aging. The next section briefly considers the health implications of culture on aging persons and their family.

## 5. AGE, CULTURE, AND SOCIAL RELATIONS

Individuals age within families, communities, and society. They interact with these organized structures most frequently through social relations. Social scientists have studied social relations extensively, and it is increasingly understood that social relations affect both the psychological health and physical health of people of all ages. This is evident in models emerging from the childhood literature, such as the life span attachment model, as well as in those emerging from the adult literature, such as the convoy model of social relations. Both of these models emphasize that social relations accumulate over time and influence both contemporaneous and longitudinal choices. These literatures have identified several elements of social relations that influence both morbidity and mortality. In the convoy model, as well as in other adult social relations models, these elements include structural characteristics such as the number, age, gender, role relationship, and years known of network members; support exchanged with network members such as love, tangible aid, and advice; and evaluation of these relationships, that is, whether one feels that they are positive or negative and whether one is satisfied and/or happy with them. Although some elements of social relations transcend age and culture, other aspects of social relations are fundamentally influenced by culture and age. For example, most people of all ages have close social relations with their parents, spouses, and children. Europeans tend to include more extended family members among their close social relations, whereas this is less true among Americans, who are more likely to include friends.

Social relations involve the exchange of social support in all cultures. How often these supports are exchanged, as well as who exchanges them, what supports are exchanged, and when the supports are exchanged, often varies by age and cultural expectations. Although it is common across cultures to expect that parents will provide for their children when their children are young, the degree to which parents continue to do so when those children become adults, or the way in which the expectation might reverse such that parents expect support from their adult children, varies by culture and country. In a recent Italian court case, parents were sued by their employed, 35-year-old son because he was reluctant to accede to his parents' wishes that he move out and live on his own. The

Italian court ruled in favor of the son, requiring the parents to continue providing room and board for as long as the son so desired. A similar case would be unheard of in the United States because it is non-normative for an able adult son to live with his parents and because no court would sustain such a petition. Similarly, in the United States, it is common for parents to attempt to provide for their adult children, for example, through gifts, favors, and loans. But in Japan, the culture assumes the opposite; that is, no matter what age, children are understood to be in a state of indebtedness to their parents and, thus, are expected to be constantly in the position of providing to their parents rather than receiving from them.

There are many factors that influence how social relations are evaluated. There are objective characteristics that influence this evaluation, but a solid accumulation of empirical evidence indicates that significant subjective characteristics influence this evaluation as well. Subjective evaluation could be considerably different depending on the age and culture within which the social relations are embedded. The Italian son just described would rate his parental support received as poor, whereas an American son would not. An adult American son might be dissatisfied and unhappy if his parents did not provide a college fund or a downpayment on a house, whereas the lack of either of these in another culture would generate no parallel dissatisfaction or unhappiness. For older family members, it is clear that their expectations vary depending on both age and culture. Interestingly, in the United States, some evidence indicates that younger people expect to provide more support than their elders expect to receive, whereas in other countries, parents report rather high expectations of what their children should provide. Regardless of the cultural convention regarding who is in an individual's social network or an objective assessment of what is exchanged, evaluation of social relations varies considerably. This variation, although influenced by many factors, has been well documented in the influence of culture on health and well-being.

## 6. CULTURAL INFLUENCES ON THE HEALTH OF AGING PEOPLE

Different cultural experiences, traditions, and values have direct implications for the expectations of both elders and their families. These expectations range from minor

considerations (e.g., concerning customs around favorite foods or holiday traditions) to more significant ones (e.g., concerning family roles and responsibilities).

There is clear evidence that culture influences health. Some examples are illustrative. It has been documented that elderly Japanese living in Japan are much more likely to die of a stroke than are elderly Japanese living in the United States, where the leading cause of death has been cardiovascular disease. This difference has been traced to dietary differences in ethnically Japanese people living in Japan versus those living in the United States. Similarly, hypertension is a major health problem among people of African descent living in the United States but is relatively rare among Africans living in Africa. Researchers have suggested that this difference is caused by the stressful life circumstances of racial discrimination experienced in the United States. Another interesting example is menopause. Research in medicine, epidemiology, and anthropology has consistently documented that women in many countries (e.g., Japan) experience fewer and/or lighter menopause symptoms (e.g., hot flash, insomnia, depression) than do American women. Although a consensus explanation has not yet been reached, these differences have been attributed to diet (e.g., soybeans) and cultural attitude toward (e.g., acceptance of) the physical and physiological changes associated with aging. A final and compelling argument for the influence of culture on health is that, regardless of the health differences evident among people of different countries, by the third generation all immigrants exhibit patterns more similar to other Americans than to people from their countries of origin.

Cultural expectations of family roles and responsibilities are critical when they have a direct impact on an older person's health and expectations concerning health care. Some simple examples are illuminating. In some countries (e.g., Japan), it is much more common for older people to live with their adult children. The differences in customs have both positive and negative implications. In Japan, older people often live in their own home, and an adult child—often the eldest son—continues to live at home with his wife and children. Although this is a traditional custom that has been changing in modern Japan, it does mean that traditionally a live-in relative attended to the health care needs of a family's oldest members. This live-in relative was most often not the eldest son but rather his wife, that is, the daughter-in-law. Much has been made of the advantages of coresidence for the elderly in terms of meeting their health care and other needs as they get older. However, social scientists have increasingly

documented that not all such arrangements are ideal. There is often tension between the daughter-in-law and her parents-in-law as well as between the son and his parents. This coresidential custom is much less common in the United States, and when it does occur, it is often the case that a parent moves in with an adult child, although the opposite may be the case when the adult child is impoverished or experiencing a life crisis such as divorce or job loss. Under the coresidential condition where the elder moves in with a child, the power dynamics are reversed in terms of the traditional American family dynamics as well as the situation just described in Japan. Tension can arise in this case as well, although health care needs and expectations are also more likely to be met. Certainly, these examples provide persuasive testimony concerning the influence of culture on the health experience of aging people.

## 7. SUMMARY AND CONCLUSIONS

With the radically changing world demographics leading to a significant increase in the number of older people, it is critical to attend to those factors that positively or negatively affect the experience of aging. Although culture is an important influence, one should be aware that there are also other factors that influence aging and family relations such as socioeconomic status, race, and industrialization. Therefore, culture should be neither overlooked nor overestimated. It remains for researchers and clinicians to identify when culture is a primary factor and when it is not and to determine at what point one should seek to optimize the experience of aging both within and across cultures.

### See Also the Following Articles

Age-Related Issues among Minority Populations ■ Aging and Competency ■ Anxiety Disorders in Late Life ■ Cognitive Aging ■ Cognitivism ■ Depression in Late Life ■ Elder Caregiving ■ End of Life Issues ■ Psychotherapy in Older Adults

### Further Reading

Antonucci, T. C. (2001). Social relations: An examination of social networks, social support, and sense of control. In J. E. Birren, & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 427–453). San Diego: Academic Press.

- Blieszner, R., & Bedford, V. (1995). *Handbook on aging and the family*. Westport, CT: Greenwood.
- Jackson, J. S. (Ed.). (1998). *The Black American elderly: Research on physical and psychosocial health*. New York: Springer.
- Jackson, J. S., Antonucci, T. C., & Gibson, R. C. (1995). Ethnic and cultural factors in research on aging and mental health: A life-course perspective. In D. K. Padgett (Ed.), *Handbook on ethnicity, aging, and mental health* (pp. 22–46). New York: Greenwood/Praeger.
- Keith, J., Fry, C. L., Glascock, A. P., Ikels, C., Dickerson-Putman, J., Harpending, H. C., & Draper, P. (1994). *The aging experience: Diversity and commonality across cultures*. Thousand Oaks, CA: Sage.
- Sokolovsky, J. (Ed.). (1997). *The cultural context of aging*. Westport, CT: Bergin & Garvey.



# Aging, Cognition, and Medication Adherence

Odette N. Gould

Mount Allison University, Sackville, New Brunswick, Canada

1. Extent and Cost of Medication Nonadherence in the Elderly
2. Types of Nonadherence
3. Adherence: A Complex Cognitive Task
4. Prospective Memory and Episodic Memory: Age-Related Changes and Effects on Adherence
5. Knowledge, Beliefs, and Communication: Age-Related Changes and Effects on Adherence
6. Measuring Adherence: A Difficult Challenge
7. Interventions: How Cognitive Psychology Can Help Further Reading

## GLOSSARY

**external memory strategies** Actions undertaken to support prospective memory whereby the physical environment is changed (e.g., the use of lists, notes, calendars, and pillboxes).

**internal memory strategies** Mental actions undertaken to support prospective memory (e.g., linking planned activity with daily events).

**medication nonadherence** Failure to take medications as prescribed; this can include taking too few or too many doses, taking medication at incorrect times or in incorrect quantities, or not following special instructions related to the medication (e.g., regarding food or beverages).

**prospective memory** Memory for actions to be performed in the future (e.g., remembering to take prescriptions at a prescribed time).

**reality monitoring** The ability to distinguish between actions that were planned and actions that were actually carried out.

**retrospective memory** Memory for events that occurred in the past (e.g., remembering the instructions received for when and how to take medication).

Medication adherence is increasingly recognized as a complex task involving many cognitive abilities, and adherence errors are common in both younger and older adults. Patients must integrate often complicated instructions on how to take medication into their daily routines, remember to take the medication at the appropriate time, and remember that the medication has been taken. Although age-related declines in many cognitive abilities are clear, older adults manage familiar tasks such as medication taking quite well, at least until very late age. Many mnemonic supports have been found to function quite effectively to remind patients to take their medication. However, more and more, researchers and health care providers are focusing on multifaceted interventions that are tailored to the specific needs, abilities, and beliefs of older patients.

## 1. EXTENT AND COST OF MEDICATION NONADHERENCE IN THE ELDERLY

Individuals over 65 years of age make up an increasingly large percentage of the population in industrialized

countries around the world, and the trend is expected to continue for many decades to come. Indeed, the World Health Organization has projected that the percentage of older adults in developed countries will increase from 14.3% in 2000 to 19.2% by 2020. The majority of older adults suffer from at least one chronic illness; consequently, older adults are the largest consumers of medications of any age group. Many of the illnesses that afflict older adults can be treated or at least controlled quite successfully through the use of prescription medication—if medications are taken as prescribed. Unfortunately, rates of adherence to medication regimens are believed to be quite low. It is quite difficult to establish adherence rates precisely, but in general, as few as 50% of patients treated for chronic conditions are believed to take their medications as prescribed. Rates of adherence are also thought to be particularly low in asymptomatic conditions such as hypertension that afflict many older adults. Surprisingly, most studies have not found evidence that older adults in general are more nonadherent to medication than are middle-aged or younger adults. Recent work by Park and colleagues even indicated that older adults, between 65 and 74 years of age, made the fewest number of adherence errors of any age group, although adults over 75 years of age made the most.

The risks of nonadherence are numerous and serious. Up to 40% of hospital admissions in adults over 65 years of age are due directly to medication errors. Even though older adults do not seem to be more nonadherent than younger adults, they are more likely to experience serious health consequences because their bodies are less tolerant of medication errors and drug interactions. When treatment regimens are not followed, the risk of toxicity is heightened, medical conditions are likely to worsen, and expensive and often invasive treatments become necessary. Clearly, enhancing medication adherence is a priority for the effective treatment of chronic diseases, for enhancing the well-being of elderly individuals, and for reducing health care costs.

## **2. TYPES OF NONADHERENCE**

Medical nonadherence occurs when patients do not carry out medical regimens as prescribed and includes any of the following: (a) not taking the medication at all, (b) taking smaller or larger doses than prescribed, (c) taking fewer or more doses than prescribed, or (d) not following recommendations regarding food or beverages while medicated. Such behaviors can be either intentional (i.e., the patient chooses not to take

the medication as prescribed) or unintentional (i.e., the patient intends to take the medication correctly but does not). In general, unintentional overuse or underuse of medication is due to cognitive variables such as forgetfulness and misinterpretation of instructions. Intentional noncompliance, on the other hand, can be seen as relatively more complex and is linked to either external circumstances or motivational issues. External circumstances include factors such as not being able to pay for the medication, not being able to reach the pharmacy to get the prescription filled, and not being able to open resistant packaging (e.g., childproof bottle tops) due to physical handicaps. Motivational issues refer to the fact that many individuals, both young and old, hold views about their illness and their medication that are incompatible with adherence.

## **3. ADHERENCE: A COMPLEX COGNITIVE TASK**

In the past, medication adherence has been presented as an example of a simple everyday prospective memory task, where prospective memory is defined as remembering to do things in the future. Recently, however, there is a clear consensus emerging that medication adherence is a complex process involving a host of variables. Although many of these factors are clearly cognitive, others (e.g., motivation to adhere to the medication regimen) involve the interplay between cognitive and social factors.

From a cognitive standpoint, the patient must understand the instructions for how to take the medication. These instructions are often complex, especially when multiple medications, each with its own regimen of doses and times, are presented. Thus, combining the information across multiple medications, and integrating all of this information with daily activities to form a plan of action, can be quite challenging. Once the plan of action is created, the patient must remember to take the medication at the correct time and must remember what dose and what medication to take. Moreover, the patient must keep track of whether or not the medication was taken so that extra doses are not taken inadvertently (i.e., reality monitoring). Reality monitoring can be particularly challenging when a medication for a chronic condition has been taken at the same time of day for months or years. Finally, the patient should monitor his or her health for a range of possible side effects from the medication, including symptoms such as confusion and forgetfulness.

#### **4. PROSPECTIVE MEMORY AND EPISODIC MEMORY: AGE-RELATED CHANGES AND EFFECTS ON ADHERENCE**

Clearly, cognitive factors such as language comprehension (e.g., understanding the medication instructions), long-term memory (e.g., remembering what to do), working memory (e.g., juggling the competing demands of everyday tasks and medication taking), problem solving (e.g., integrating complex medication instructions with the daily routine), prospective memory (e.g., remembering to take the medication at the correct time), and metamemory (e.g., monitoring memory performance and judging whether external support is needed for adequate adherence) all are involved in determining whether correct adherence will occur. Although age-related losses in many of these areas have been clearly established, many older adults are able to maintain high levels of performance, especially on everyday tasks such as medication taking.

Many language abilities are maintained into very old age. However, the comprehension of complex medication instructions may depend heavily on working memory capacity as well as language abilities. Age-related losses in working memory and problem solving are well documented. It is clear that if information is presented quickly or is presented using an unfamiliar or disorganized format, many older adults will have difficulty in comprehending and recalling the information. Many studies have shown that older adults are at a distinct disadvantage, compared with younger adults, when asked to combine complex medication instructions into a plan of action, especially if the instructions require inferences. Such problems have also been found when older adults are asked to transfer medication into weekly pillboxes, especially if the pillboxes do not contain multiple compartments for each day.

To take medications correctly, the patient not only must understand the instructions when they are given but also must remember these instructions over the days and weeks following the meeting with the physician. In general, age-related losses are also observed when the ability to recall verbal information after a relatively long interval (i.e., hours or days) is measured. However, age differences can be greatly reduced, at least for relatively healthy older adults, if environmental supports for recall are present. For example, if information is presented clearly, if multiple formats are used (e.g., written and verbal presentation), if recall is based on habitual or

familiar activities, if cues for recall are present in the environment, and if the person is not distracted by competing stimuli at time of recall, age differences in episodic memory are greatly reduced or even eliminated.

Although it is well documented that there are age-related losses in working memory, problem solving, and long-term memory, prospective memory does not show the same rates of decline with age. For example, when asked to make a telephone call or mail a postcard at a certain time, older adults attain performance levels similar or superior to those of younger adults through the use of memory strategies. This may help to explain why older adults do not make more medication errors than do younger adults. Park and colleagues have speculated that adults in their 60s and 70s may have few adherence errors because (a) they are sufficiently focused on their health to adopt mnemonic strategies for medication (unlike middle-aged adults) and (b) they have the cognitive ability to adopt and use mnemonic strategies (unlike much older adults).

In laboratory-based studies of prospective memory, it has been suggested that prospective memory tasks can be divided into time-based and event-based tasks. In a time-based task, the individual must accomplish the task at a certain time (e.g., "Take your pill at 8 AM"), whereas in an event-based task, the individual must accomplish the task when cued by a specific event (e.g., "Take your pill before breakfast"). Computer-based simulations of prospective memory tasks have shown that event-based tasks are easier for older adults to accomplish, especially if the accompanying background task is not particularly difficult. This may also explain why many older adults report that the main strategy they use to remember their medication is to mentally link the medication with routine daily activities. It may also explain why some studies have found that tailoring medication regimens to patients' routines enhances adherence significantly.

#### **5. KNOWLEDGE, BELIEFS, AND COMMUNICATION: AGE-RELATED CHANGES AND EFFECTS ON ADHERENCE**

At least three major foci in the research on the motivational factors of adherence in old age can be identified. First, the beliefs and attitudes that the patient has about his or her illness have been shown to predict adherence. Some people may deny that they are ill or might not believe that their conditions have serious consequences.

Patients who are more knowledgeable about their conditions are more likely to follow recommended medication regimens. This may also explain why social support has been found to predict adherence. It may be that significant others play a role in motivating an individual to adhere. Providing cues or reminders to follow the prescription also may play a role.

A second factor concerns the beliefs that the patient holds about his or her treatment. Beliefs about the effectiveness of the medication, about whether the regimen prescribed is appropriate, about whether medications cause uncomfortable side effects, and about how important it is to adhere to medication regimens as prescribed have been shown to contribute to treatment adherence. Thus, some older adults may simply not believe that their medication is helpful. Others may choose underadherence to avoid side effects or may overmedicate because they believe that the prescribed dose is ineffective. (It should be noted that because many health care professionals lack extensive knowledge about the effects of medications—especially multiple medications—on the older body, partial adherence may be beneficial in at least some cases.)

Finally, a third set of factors revolves around the relationship with the health care provider. This relationship is critical in many ways. Medication adherence can be seen as the outcome of a process that begins before the physician and patient even meet. Namely, the attitudes and beliefs that each holds about the other are likely to affect patient–physician interactions in at least two ways. First, the quality of the interaction and the satisfaction that the patient feels with his or her health care are very important predictors of motivation to adhere to a medication regimen. Second, the relationship may also affect the quantity and quality of the information that is exchanged during the patient–physician encounter. For adherence to occur, the patient must be told how to take the medication in detail and must be informed about his or her condition and its associated risks. Unfortunately, many studies suggest that health care providers are less likely to provide this information to older patients than to younger ones, and older patients are also less likely to question their physicians. Thus, even though older adults are likely to be taking more medications and to have more complex regimens, they are also likely to receive less information about their medications. Finally, the relationship between the health care provider and the patient also affects the quality of the information that the patient provides. In other words, the patient is more likely to provide accurate and detailed information about his or her symptoms and adherence

to existing medications when the patient perceives the health care provider to be caring and approachable. Such accurate information is key for the health care provider to be able to make accurate diagnoses and treatment decisions.

In conclusion, even though the relationship between the health care provider and the patient is commonly addressed from a social psychology perspective, this relationship has important cognitive repercussions as well. The patient's knowledge about the illness and medication and his or her beliefs about the importance of adherence are often highly dependent on how well and how comfortably the patient communicates with the health care provider.

## **6. MEASURING ADHERENCE: A DIFFICULT CHALLENGE**

The difficulty of measuring medication adherence is well established and often discussed in the literature. Both technical and nontechnical approaches have been used. One technical approach has been the evaluation of body fluids (e.g., blood, urine) for the presence of the prescribed medication. However, this technique only recognizes drugs ingested during the recent past and does not indicate whether the patient has taken the medication regularly. Pharmacy refill rates and pill counts have also been used; however, with these methods, flawed data are obtained if the patient is using multiple pharmacies, disposing of pills, hoarding pills, and/or sharing pills with others. The most accurate data come from a specially designed pill bottle with a lid containing a microchip that records every time the bottle is opened. However, even this system still does not provide an accurate measure of whether the drugs were actually taken when the pill bottle was opened and does not verify whether the correct doses were taken. Although these special pill bottles are useful for research purposes, their general use remains limited because they are very costly.

The most convenient and the most often used means of assessing adherence is self-reports. However, it is clear that self-reports underestimate medication errors. Patients are likely to overestimate their adherence levels, not only because they do not recall errors but also because they hope to avoid confrontations and/or embarrassment. Cognitive factors also contribute to the accuracy of self-reports. When asked to recall the frequency of a behavior, respondents are likely to reduce their efforts

by estimating rather than counting, and this is especially true of older adults. Moreover, with older adults, memory complaints are often not predictive of objective memory performance. Thus, the older adults who complain about having bad memories might not be the ones who make adherence errors due to forgetting. Despite the potential pitfalls of self-reports, recent work by Hertzog and colleagues suggests that when older patients are asked specific questions about memory problems relating to medication adherence rather than general memory performance, self-reports can be quite useful.

### 7. INTERVENTIONS: HOW COGNITIVE PSYCHOLOGY CAN HELP

Much of the intervention research has focused exclusively on compensating for either cognitive or motivational deficits. Many studies have focused on providing older adults with external mnemonic strategies. Such strategies involve effecting changes in the environment so that it provides prospective memory cues. Many studies have been carried out to investigate the effectiveness of aids such as pillboxes and pill bottle alarms, memory training, voice mail reminders, and organizational charts. Most of these studies found such memory aids to be at least somewhat effective in enhancing medication adherence. However, many of these studies were conducted over a short period of time, and neither young nor older adults are likely to continue using memory strategies assigned to them after the training period has ended. It is also interesting to note that the few studies that examine what strategies older adults use spontaneously tend to find that internal strategies (i.e., using mental actions to help encode or retrieve information) are reported quite often. However, very little is known about the effectiveness of internal strategies for routinized, long-term prospective tasks such as medication adherence. Also, little is known about what drives the process of determining that a mnemonic strategy is necessary, deciding which one to use, and evaluating its effectiveness.

One external strategy that is being adopted by many older adults is the use of pillboxes with daily compartments. However, as noted previously, at least some older adults have difficulty in transferring pills to the container correctly, at least when the smaller versions are used (i.e., only one compartment per day), and the larger containers have the disadvantage of being bulky and might not be practical unless patients are

housebound. Recently, attempts to reduce adherence errors have focused on using technological advances in how medication is presented to patients. Some have speculated that as pharmaceutical science continues to advance, medications will be released into the body in a controlled fashion (e.g., slow-release pills), thereby making frequent doses and complex regimens unnecessary. More immediately, the use of individualized packaging is becoming more affordable and common. For example, when blister packs are used, the time to take the medication can be listed on the package, and monitoring whether the medication has been taken becomes much easier. However, when patients must take many different medications, each in its own blister pack, the packaging may increase rather than decrease the confusion (and be large and bulky to transport as well). Finally, increases in adherence levels have also been obtained in studies that addressed motivational factors. These studies focused on improving the relationship between the physician and the patient or tried to educate the patient as to the seriousness of the illness or the importance of the medication.

In conclusion, many different interventions can be useful, but there is a growing recognition that the intervention applied must be tailored to the characteristics of the patient. Clearly, it is not helpful to provide mnemonic strategies to someone who is nonadherent by choice, and neither is it helpful to convince someone of the importance of medications if the person cannot remember to take them. Thus, cognitive psychologists today are working in collaboration with health care providers to develop exciting multifaceted approaches that tailor cognitive and motivational interventions to the needs, abilities, and beliefs of individual older adults.

### See Also the Following Articles

Age-Related Issues among Minority Populations ■ Aging and Competency ■ Cognitive and Behavioral Interventions for Persons with Dementia ■ Cognitive Skills: Training, Maintenance, and Daily Usage ■ Elder Caregiving ■ Psychotherapy in Older Adults

### Further Reading

Brandimonte, M., Einstein, G. O., & McDaniel, M. A. (1996). *Prospective memory: Theory and applications*. Mahwah, NJ: Lawrence Erlbaum.

Johnson, M. J. (2002). The medication adherence model: A guide for assessing medication taking. *Research and Theory for Nursing Practice: An International Journal*, 16, 179–192.

- McElnay, J. C., & McCallion, R. (1998). Adherence in the elderly. In L. B. Myers, & K. Midence (Eds.), *Adherence to treatment in medical conditions* (pp. 223–253). Amsterdam: Harwood Academic.
- Murdaugh, C. L. (1998). Problems with adherence in the elderly. In S. A. Shumaker, E. B. Shron, J. Ockene, & W. L. McBee (Eds.), *The handbook of health behavior change* (2nd ed., pp. 357–376). New York: Springer.
- Park, D. C., & Jones, T. R. (1997). Medication adherence and aging. In A. D. Fisk, & W. A. Rogers (Eds.), *The handbook of human factors and the older adult* (pp. 257–287). San Diego: Academic Press.
- Park, D. C., Morrell, R. W., & Shifren, K. (1999). *Processing of medical information in aging patients: Cognitive and human factors perspectives*. Mahwah, NJ: Lawrence Erlbaum.



# Agreeableness

Brad E. Sheese and William G. Graziano

*Purdue University, West Lafayette, Indiana, USA*

1. Describing Agreeableness: What Is It?
2. External Correlates: What Is Agreeableness Related To?
3. Etiology/Antecedents: Where Does Agreeableness Come From?
4. Conclusion  
Further Reading

## GLOSSARY

**effortful control** The ability to suppress a dominant response for a subdominant response (e.g., the ability to cope with frustration).

**knowledgeable others** Peers, teachers, parents, and others who know the person being evaluated.

**latent variable** An underlying system.

**personality dimension** A set of related traits.

Agreeableness is a personality dimension (or set of related traits) that describes a class of individual differences that generally have to do with being pleasant, likable, and harmonious in relations with others. Empirical research finds that Agreeableness as a dimension hangs together consistently (persons who are “kind” also tend to be “considerate”), is relatively stable and enduring over time, and is related to many kinds of human activity.

## 1. DESCRIBING AGREEABLENESS: WHAT IS IT?

Agreeableness has become the label most frequently used for this personality dimension, but it is only one of many such labels. Some of the other labels used to describe the dimension (or closely related dimensions) are tendermindedness, friendly compliance versus hostile noncompliance, love versus hate, likability, communion, and conformity. It has been argued that none of these labels, including Agreeableness, adequately captures either the breadth or the substantive content of this dimension of personality. As a label, Agreeableness has been criticized specifically for being too narrow and perhaps for overemphasizing acquiescence. Theorists have suggested that it may be more appropriate to refer to the dimension either with numerals (the Roman numeral II has been used in the past) or simply with the letter A (for agreeableness, altruism, and affection).

At a theoretical level, Agreeableness describes an underlying system (latent variable) of individual differences. It is one of five broad personality dimensions that appear in all versions of the five-factor approach to personality (i.e., the Five-Factor Model). The five-factor approach describes personality at perhaps its broadest and most abstract (decontextualized) level. Trait adjectives that are positively associated with Agreeableness include kind, warm, cooperative, unselfish, polite, trustful, generous, flexible, considerate, and agreeable. Trait adjectives that are negatively associated with

Agreeableness include cold, unkind, uncooperative, selfish, rude, distrustful, stingy, stubborn, and inconsiderate. Overall, Agreeableness describes a broad, but related, set of individual differences in how a person relates to others. Specifically, Agreeableness appears to describe differences in being predominantly prosocial or other-oriented versus antisocial or self-oriented in social interactions.

Self-report measures of Agreeableness are those most frequently used for assessment. The use of ratings of Agreeableness by knowledgeable others (e.g., peers, teachers, parents) in addition to self-ratings is desirable from a psychometric standpoint but is less common due to the increased demands associated with collecting this information. Currently, there are several measures available to assess Agreeableness within the framework of the five-factor approach to personality. The three most frequently used measures in research are Goldberg's adjective markers, the questionnaire-format Big Five Inventory (BFI) developed by John, and the Neuroticism Extraversion Openness Personality Inventory-Revised (NEO) developed by McCrae and Costa. These measures vary in their construction, length, and content, yet empirical research shows that the five-factor dimensions assessed by these measures, including Agreeableness, are nearly identical across measures. Focusing on Agreeableness, there is some discrepancy between the NEO and the other two measures. The main difference is that the NEO includes warmth as a facet of Extraversion, whereas in the other two measures warmth is associated with Agreeableness. Overall, however, outcomes of studies using different measures of Agreeableness show convergence. For most purposes, the various measures of Agreeableness are functionally equivalent.

## 2. EXTERNAL CORRELATES: WHAT IS AGREEABLENESS RELATED TO?

Over the past decade or so, cross-cultural research has uncovered the same or similar five-factor personality structure in many cultures with many different languages and in both adults and children. In all cultures studied, Agreeableness has been seen as a major dimension of personality. At a minimum, this research indicates that Agreeableness describes a set of individual differences that are of interest worldwide. The dimension is not an artifact of the English language, college-age participants, or Western cultural norms. Recent

research also suggests that despite the clear preference that most groups have for persons with agreeable qualities over persons with disagreeable qualities, Agreeableness differences are also not an artifact of social desirability, self-presentation, or self-deception, nor are they merely a characteristic of observers' evaluations. Instead, research suggests that Agreeableness describes a stable enduring set of qualities in individuals that is related to important social outcomes.

Agreeableness differences can be studied in many different ways. One of the most common ways is to bring persons who differ on Agreeableness into a laboratory, where their behaviors can be assessed in various kinds of social situations. Typically, these persons are college students. In laboratory studies of group processes, Agreeableness is related to lower within-group conflict and to positive within-group evaluations. For example, more agreeable people like their fellow group members more and are liked more in return. Research also shows that Agreeableness is negatively related to competitiveness in group interactions and is positively related to expectations of enjoying group interactions. These findings suggest that Agreeableness may contribute to group cohesion and may serve as a protective factor against group dissolution. These findings also provide empirical support for the notion that Agreeableness can be mapped onto the social motives described by communion.

In some circumstances, being a more cooperative and generally more prosocial group member may contribute positively to group and individual functioning. In these circumstances, one would expect that Agreeableness is related to both group- and individual-level performance. In some situations, however, being considerate, trusting, and compassionate may undermine a group's performance or an individual's performance. In these situations, Agreeableness may be unrelated, or even negatively related, to performance. One meta-analysis found that Agreeableness was negatively related to job performance for managers but also found that Agreeableness was positively related to job performance for professionals and skilled laborers. A later meta-analysis found that in most studies, Agreeableness was unrelated to job performance but was positively related to teamwork. In a competitive business environment, Agreeableness may be a more desirable quality for group members than for managers or business leaders. To a certain extent, goals and tasks that emphasize competition, profits, and self-gain may run contrary to the selflessness and positive social orientation associated with Agreeableness. This may, in part, explain findings that, across the life span, Agreeableness is negatively related

to extrinsic success (e.g., salary) and career satisfaction. Perhaps more agreeable individuals would find more satisfaction, if not more success, in helping professions or public service than in corporate job environments.

Other research (some laboratory and some field studies) has examined Agreeableness and dyadic relationship processes and outcomes. One area of interest has been the resolution of dyadic conflicts. If more agreeable individuals are more motivated to maintain positive social relations with others than are their less agreeable peers, one would expect more positive patterns of attributions and behaviors from agreeable individuals than from their peers when they have disagreements with others. We might also expect that a concern for others would lead to less disagreement or conflict with others in general. Research outcomes have shown that in both adolescents and adults, Agreeableness is positively related to constructive conflict resolution tactics (e.g., negotiation) and is negatively related to destructive conflict resolution tactics (e.g., physical force). Agreeableness is also negatively related to the number of daily conflicts that individuals report in daily diary records. After a conflict has occurred, agreeable individuals say that they are more upset by the conflict but are less likely to have vengeful attitudes about the conflict or to use vengeance as a problem-solving tactic.

In general, more agreeable individuals appear to behave in ways that are constructive rather than destructive to their relationships with others. This suggests that Agreeableness should be positively related to relationship quality (e.g., fewer conflicts, more satisfaction) and perhaps to relationship quantity (e.g., more friends, more robust social networks). Agreeable individuals may incur costs in maintaining their relationships, but they should also gain the benefits associated with having a strong social network. During adolescence, research found that Agreeableness was positively related to the quality of teacher and peer relations as well as to peer social status. In adults, research on social support showed that Agreeableness was positively related to the perceived support satisfaction, and negatively related to the frequency, of negative social exchanges.

These positive relationships with others may either stem from or contribute to a positive relation between Agreeableness and evaluations of the self. As discussed previously, agreeable individuals view others favorably and are viewed favorably by others. Research also shows that agreeable individuals view themselves favorably. During adolescence, Agreeableness and self-esteem are positively related in both self-ratings and teacher ratings. In adults, Agreeableness is positively related to

happiness and subjective well-being. Agreeableness is negatively related to anger, hostility, and (indirectly) depression in both daily and general dispositional self-reports. In general, agreeable individuals earn higher scores on measures of adjustment than do their less agreeable peers. Conversely, research shows that less agreeable individuals are more likely to exhibit various kinds of externalizing problems during both adolescence and adulthood.

### 3. ETIOLOGY/ANTECEDENTS: WHERE DOES AGREEABLENESS COME FROM?

In considering the development of individual differences in Agreeableness, researchers have taken two distinct but related approaches. The first approach is to examine the specific contribution of genes to individual differences in Agreeableness. Over the years, behavioral genetic studies have attempted to determine the contribution of both genetic and environmental (shared and nonshared) influences on personality. Results have varied, with at least one study suggesting that among the Big Five dimensions, Agreeableness is the most susceptible to environmental influences. However, most of the behavioral genetic studies do not find that Agreeableness is any less heritable than the other Big Five dimensions. Some theorists suggest that all of the Big Five dimensions, including Agreeableness, represent entirely genetically driven, biologically based dispositions. Supporting this claim are a variety of studies showing that constructs related to Agreeableness, such as altruism, prosocial behavior, and empathy, also show moderate to high heritabilities.

The second approach to identifying the root of individual differences in Agreeableness has been to link adult personality to infant and child temperament. One problem is that the individual differences identified through modern temperament approaches rarely map directly onto adult personality traits. Agreeableness, in particular, has no clear antecedent among temperament constructs. This may, in part, have to do with the emphasis of Agreeableness on the quality of peer relations and other-oriented responsiveness. During infancy, it may be difficult to detect individual differences in the concern for the quality of relationships with others given that infants have a comparatively limited repertoire of both behaviors and communication abilities. In addition, in comparison with other personality

dimensions such as Extraversion and Neuroticism, if Agreeableness is already present as a disposition, it may require the development of more elaborate cognitive processes (e.g., perspective taking, theory of mind) before it can be observed.

Among the temperament traits that have been observed, theorists have suggested that individual differences in effortful or executive control may be an antecedent of individual differences in Agreeableness. Effortful control refers to the ability to suppress a dominant response for a subdominant response. In particular, effortful control is related to the ability to cope with frustration and respond adaptively to goal blockage. More generally, researchers have proposed that effortful control is linked to both cognitive (e.g., attention) and emotional self-regulatory capabilities. In this light, individual differences in Agreeableness may reflect differences in the ability to monitor and control a variety of self-centered or even antisocial impulses in social situations. For example, in daily interactions, less agreeable individuals report feeling (and probably exhibit) more anger and hostility than do their more agreeable peers. These feelings of anger and hostility may stem from an underlying inability to regulate or suppress emotions to avoid damaging relationships.

Even if there were large heritable or temperamental contributors to Agreeableness, it would still be necessary to know how social learning histories and life experiences shape these tendencies into adult individual differences. There is good evidence that cultures differ in the values they assign to Agreeableness; accordingly, we might expect average differences in Agreeableness among children reared in different cultures. Nevertheless, differences within cultures remain. Agents of socialization may leave their marks, but persons enter the socialization environment with something more complicated than blank slates. Some individuals will be receptive to agents of socialization, absorbing and internalizing cultural values, whereas other individuals will resist. In this approach, then, socialization can produce large effects on Agreeableness, yet adult differences in Agreeableness within cultures are the end-state products of Person  $\times$  Socialization interactions.

Consistent with this line of reasoning, recent research has demonstrated complex links between childhood self-regulation and adult Agreeableness. Childhood self-regulation has, in turn, been linked both to the quality of child-rearing practices and to temperamental characteristics. It is possible that temperamental characteristics and socialization experiences may interact in

a variety of ways to produce individual differences in Agreeableness. For example, temperamental characteristics may contribute to the selection of particular social ecologies that restrict the range of socialization experiences available to any given individual. Children who lack self-control and are easily frustrated may make for poor relationship partners and may subsequently be rejected by their peers. Consequently, these children might miss important socialization experiences provided by peer relationships. Temperament may interact with parent-child relationships to affect socialization experiences in a similar way. As children get older, they may also consciously choose to place themselves in particular environments that are consonant with their temperamental predispositions. Similarly, as adult personality characteristics emerge, they may interact with situations to reinforce underlying predispositions. For example, if given a choice, highly agreeable children may choose to avoid highly competitive situations. If forced into competitive situations, highly agreeable children may attempt to convert the situation into a more cooperative one or otherwise adapt to the social norms associated with highly competitive interactions.

#### 4. CONCLUSION

Agreeableness is a dimension of personality associated with motives for maintaining positive relations with others. It is a superordinate descriptor for a system of lower order affective, behavioral, and cognitive tendencies. Cultures may differ in the emphasis they put on socializing children into seeking harmonious relations with others, but differences remain within cultures. Agreeableness differences in people seem to be relatively stable over time and across settings. Agreeableness may have developmental origins in temperament and in skills in coping with frustration. These differences are associated with corresponding tendencies in dealing with conflict, emotional responsiveness to others, group interaction, interpersonal relationships, and self-evaluation. For many life events, Agreeableness is associated with positive outcomes such as good social relations, friendships, better adjustment, and social recognition (with the exception of external rewards, e.g., salary). There is some suggestion that agreeable persons are less effective as managers, but there is no evidence (at least so far) for claims that agreeable persons are less attractive, more conforming, more compliant, or more submissive than their peers.

**See Also the Following Articles**

Emotion ■ Personal Initiative and Innovation ■ Social Networks ■ Traits

**Further Reading**

- Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 795–824). San Diego: Academic Press.
- Kohnstamm, G. A., Halverson, C. F., Mervielde, I., & Havill, V. L. (1998). *Parental descriptions of child personality: Developmental antecedents of the Big Five?* Mahwah, NJ: Lawrence Erlbaum.
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement in interpersonal behavior. In D. Cicchetti, & W. Grove (Eds.), *Thinking critically in psychology: Essays in the honor of Paul E. Meehl* (pp. 89–113). New York: Cambridge University Press.



# Alcohol Dependence

Norman S. Miller

Michigan State University, East Lansing, Michigan, USA

1. Who Becomes Alcoholic?
  2. Other Drug Use Associated with Alcohol Dependence
  3. Dependence Syndrome and Screening Instruments
  4. Alcoholism Treatment: Pharmacological Approaches
  5. Psychiatric and Medical Complications
  6. Alcoholism Treatment: Behavioral Approaches
  7. Long-Term Recovery from Alcoholism
  8. Beneficial Versus Harmful Effects of Alcohol
  9. Prevention of Alcohol Use and Problems
- Further Reading

Practically everyone has consumed alcohol in his or her lifetime. Fully 90% of the U.S. population has had at least one alcohol-containing beverage. In addition, 30% of the population drinks regularly with adverse consequences (15% develops alcohol dependence and another 15% drinks heavily). Costs due to problems from alcohol use and dependence reach \$100 billion annually in the United States. Despite the available knowledge about the etiology and course of alcohol consumption, the costs from adverse consequences continue to rise each year for a wide range of age groups (including children, adults, and the elderly), most ethnicities and cultures, and both genders.

## GLOSSARY

**anhedonia** Lack of pleasure or satisfaction in experiences that are normally pleasant or satisfying.

**delirium tremens** A kind of alcohol withdrawal occurring during abstinence or reduced alcohol use after excessive long-term alcohol consumption; symptoms include delirium and hallucinations.

**DSM-IV-R** Abbreviation for *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, revised (published January 2000).

**DSM-IV-TR** Abbreviation for *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, text revision (published June 2000).

**gateway drug** A substance that, when abused, is known to lead to the abuse of other substances.

**phylogenetic** Relating to evolutionary relationships within and between groups.

## 1. WHO BECOMES ALCOHOLIC?

The best predictor of who will become an alcoholic is a positive family history of alcoholism. Family history studies of alcoholism show that alcoholism is common among families of alcoholics and that a given alcoholic is likely to have a family history for alcoholism. Moreover, family history, rather than environmental factors, is associated with the development of alcoholism. Contrary to popular belief, histories of deprivation and poverty do not predict alcoholism nearly as well as do biological factors. Furthermore, although mental illness is commonly associated with alcoholism, most alcoholics do not have a mental illness beyond their alcoholism.

Adoption studies reveal that the biological parents, and not the adoptive parents, predict alcoholism in the offspring. Thus, it is the relationship of sons and daughters to parents through blood linkage, and not who raised them or their environment, that predicts the onset of alcoholism. In addition, identical twins who share the same genes (DNA) have a greater prevalence of alcoholism than do fraternal twins who share half of the same genes (DNA), again implicating the biological origin of alcoholism. Importantly, although the prevalence of alcohol dependence typically has been greater in males, studies have shown that the biological risk for developing alcohol dependence is the same among women as among men.

When children are asked for reasons for initiating alcohol (and drug) use, peer influences and exposure to alcohol are the best predictors. Those children who associate with other children who consume alcohol are more likely to try alcohol and use it more often. Other factors, such as poverty and social deprivation, do not predict alcohol use; in fact, certain racial populations such as black males are negatively correlated with onset of alcohol dependence (i.e., they have lower rates of alcoholism). Thus, we can predict alcohol use and onset of dependence with a family history and exposure to alcohol, particularly if an individual is biologically related and exposed to alcohol by peers. Physicians can do well in assessing the risk for onset of alcohol dependence, and who might actually be alcoholic, by inquiring about patients' family histories of alcoholism and patients' association with other drinkers.

## 2. OTHER DRUG USE ASSOCIATED WITH ALCOHOL DEPENDENCE

Other drug use is commonly associated with alcohol use and dependence. For instance, more than half of those identified as alcoholic will have at least one other drug dependence. This is somewhat age dependent; younger alcoholics are more likely to have another drug dependence. At least 80% of alcoholics are dependent on nicotine, with the vast majority smoking cigarettes. Conversely, 30% of those with nicotine dependence are dependent on alcohol. Cannabis (marijuana) dependence is probably the next most common drug dependence among alcoholics, particularly among younger alcoholics. Most cocaine addicts use alcohol, and many qualify for alcohol dependence, as do most heroin addicts.

Longitudinal studies tend to demonstrate alcohol as a gateway drug, followed by nicotine, for those who go on to develop other drug addictions such as cocaine and heroin. Surveys of high school-age students show that 80 to 90% of students have consumed alcohol during the past year and that 5 to 10% drink daily. Also, the mean age of onset of alcoholism in the U.S. population is 21 years in males and 24 years in females, indicating that alcoholism is a youthful disorder that begins to be prevalent during the teenage years. Initiation of other drug use also occurs during these vulnerable developmental years, when biological and especially peer factors are operative.

## 3. DEPENDENCE SYNDROME AND SCREENING INSTRUMENTS

### 3.1. Syndromes

The dependence syndrome, as described in the fourth edition (text revised) of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR), defines a maladaptive pattern of alcohol use leading to clinically significant impairment or distress, as manifested by three or more of the following occurring at any time during the same 12-month period:

- Tolerance (need for increase in dose or loss of effect)
- Withdrawal (typical of alcohol withdrawal: anxiety, dysphonic mood, muscle aches, insomnia, nausea, vomiting)
- Use of larger amounts for longer periods of time than intended
- Persistent desire or unsuccessful attempts to cut down or control alcohol use
- Preoccupation with acquiring alcohol (e.g., frequently in the presence of alcohol; giving up or reducing important social, occupational, or recreational activities due to alcohol use)
- Alcohol use despite knowledge of adverse physical or psychological problems caused or exacerbated by alcohol (e.g., depression; anxiety; restricted options in daily living; pursuit of drugs to the exclusion of social, occupational, and normal interpersonal relationships; poor motivation to solve pain problems through evaluation and treatment; resistance to alternative resolutions to life problems)

Pervasive to these criteria is a loss of control over the use of alcohol, whether an inability to resist initiating alcohol use or an inability to modulate the amount

used during a given episode of alcohol consumption. Accordingly, the dependence syndrome mirrors the following basic definition of behavioral addiction:

- Preoccupation with acquiring alcohol
- Compulsive use of alcohol (continued use despite adverse consequences)
- Pattern of relapse to alcohol use despite adverse consequence

The neuropharmacological basis for behavioral addiction or dependence on alcohol lies in the phylogenetically ancient portion of the brain in the limbic system, which is responsible for instinctual drive states and emotions. The brain center responsible for behavioral addictive use (and dependence) lies in the mesolimbic pathway, which extends from the ventral tegmentum in the midbrain to the nucleus accumbens in the limbic forebrain. Scientific studies show that stimulation of the mesolimbic pathway with alcohol and other drugs reinforces compulsive use of alcohol that closely resembles the out-of-control addictive use seen in patients.

### 3.2. The CAGE and MAST

As screening instruments, the CAGE (Cutting down, Annoyance by criticism, Guilty feeling, and Eye-openers) and MAST (Michigan Alcohol Screening Test) are useful in identifying problematic use of alcohol and probable alcohol dependence in individuals. In the CAGE, individuals are asked whether they find it hard to cut down, whether others are annoyed by their alcohol use, whether they feel guilty about drinking, and whether they use alcohol as an eye-opener in the morning or early in the day to relieve distress (usually symptoms of withdrawal from alcohol). A positive answer to one question means a possible diagnosis of alcohol dependence, and a positive answer to two questions means a probable diagnosis. However, the CAGE is a screening instrument that requires a follow-up clinical assessment for a revised DSM (DSM-IV-R) or similar objective diagnostic scheme for definitive diagnosis.

The MAST is a self-administered questionnaire that comes in various lengths (e.g., brief, standard), depending on the number of questions indicating a diagnosable alcohol dependence according to the number of positive answers. As with the CAGE, the MAST requires a follow-up clinical interview to confirm the diagnosis of alcohol dependence.

## 4. ALCOHOLISM TREATMENT: PHARMACOLOGICAL APPROACHES

We know well that pharmacological treatment of withdrawal from alcohol or medical detoxification reduces morbidity and mortality from chronic alcohol use. The mortality from untreated delirium tremens ranges between 20 and 50%, and untreated alcohol withdrawal can complicate other conditions associated with alcohol withdrawal such as cardiovascular problems (e.g., abnormal heart rhythm, hypertension).

The reliable and commonly used medications to treat alcohol withdrawal are from the benzodiazepine class and include diazepam (Valium) and lorazepam (Ativan). Benzodiazepines share cross-tolerance and dependence with alcohol and are effective in suppressing signs and symptoms of withdrawal from alcohol in the brain at the gamma amino butyric acid (GABA) receptors on nerve cells. These are particularly effective medications to treat withdrawal when used in the short term under clinical supervision for alcohol withdrawal.

Moreover, pharmacological management of alcohol withdrawal improves compliance with behavioral approaches to alcohol dependence in promoting abstinence from alcohol. Alleviation of the noxious symptoms of withdrawal, such as anxiety, depression, malaise, and intense cravings for more alcohol, will reduce the risk for continuous alcohol consumption and provide a transition to the abstinent state. Behavioral management is typically not effective in the setting of active alcohol use due to the poor insight and judgment and the reduced motivation in the intoxicated and withdrawal states from alcohol use.

## 5. PSYCHIATRIC AND MEDICAL COMPLICATIONS

Alcohol regularly induces or causes a plethora of psychiatric symptoms in varying degrees of severity. Depression and anxiety can be expected to occur in virtually any chronic drinker (e.g., daily or weekly consumption). Correspondingly, these symptoms usually subside within a few days following cessation and during abstinence from alcohol. Predictably, chronic drinkers experience sadness, loss of energy, anhedonia, hopelessness, and helplessness that reaches degrees equivalent to those of major depression. In addition, suicidal thinking is common among drinkers; alcoholism is the most common diagnosis among those attempting or completing

suicide. Moreover, 25% of alcoholics commit suicide. Often, these suicidal thoughts subside within days of abstinence from alcohol.

Anxiety is a prominent state during alcohol withdrawal, and most chronic drinkers spend some portion of the day in withdrawal as the blood alcohol level drops. Also, tolerance to alcohol diminishes with chronic intake, such that greater symptoms of anxiety and depression reflect increasing dependence on alcohol. Usually, anxiety and depression subside with abstinence within a few days.

Occasionally, visual and auditory hallucinations, as well as delusions (mostly paranoid in type), can occur in chronic drinkers. Typically, these subside with abstinence, although short-term treatment with antipsychotic medications may be required.

Importantly, 50% of those with significant psychiatric diagnoses will use alcohol and often qualify for an alcohol dependence diagnosis. Although a presumptive dual diagnosis can be made on the basis of history and mental status examination, abstinence from alcohol over time is often required to make a definitive diagnosis to exclude the alcohol-induced effects on mood and thought.

## **6. ALCOHOLISM TREATMENT: BEHAVIORAL APPROACHES**

Although there are many studies on a variety of behavioral approaches for alcohol problems, these methods are not designed for “typical” alcoholics as defined in the DSM-IV-R. Consequently, brief interventions do not work well in individuals who have actual alcohol dependence with the concomitant loss of control over alcohol use. Moreover, studies of behavioral management approaches typically do not provide long-term follow-up of their effectiveness beyond a few months. These approaches have a variety of names, such as relapse prevention and motivational enhancement, but most are derived from cognitive behavioral techniques. However, one approach termed “12-step facilitation” shows long-term positive results principally due to referral to 12-step groups such as Alcoholics Anonymous (AA).

An approach called “controlled drinking” is a misnomer and has caused substantial confusion, and ultimately harm, to those who suffer from alcohol dependence. In studies on controlled drinking, abstinence was the best predictor of those who experienced the fewest adverse consequences from alcohol use

following controlled drinking exercises. Moreover, the mortality from these controlled drinking techniques during follow-up was high among the study participants. Carefully constructed studies of actual alcoholics revealed that once loss of control over alcohol use manifested itself, the return to controlled drinking was not possible. Thus, abstinence is required to prevent adverse consequences from alcohol once alcohol dependence is established in individuals.

Importantly, behavioral approaches that employ the 12-step approach to the treatment of alcohol dependence show significant abstinence rates from alcohol. These approaches incorporate cognitive behavioral management with a focus on alcoholism as an independent disease that responds to treatment and recovery methods over time. Typically, treatment of alcoholism occurs in inpatient and outpatient settings following detoxification from alcohol and concentrates on increasing the individual's ability to refrain from relapse to alcohol and to maintain the abstinent state.

Outcome studies show that following a treatment experience in a structured inpatient or outpatient treatment program, 60% of patients achieve abstinence at 1 year no matter what follow-up care they may receive. However, they achieve abstinence rates of 70% if they attend continuing care in the treatment program (20–30%) or attend 12-step groups regularly (50% attended at least one AA meeting per week). In addition, they improve their abstinence rates to 80% if they both receive continuing care and attend AA meetings.

## **7. LONG-TERM RECOVERY FROM ALCOHOLISM**

Although spontaneous recovery from alcoholism occurs, the only assurance of long-term abstinence supported by empirical data and clinical experience is continuous attendance in AA. Surveys of its members reveal an 83% probability of abstinence for the next year if an individual is abstinent between 1 and 5 years with regular attendance of at least one AA meeting per week and a 91% probability of an additional year if the individual is abstinent in AA for more than 5 years.

Clinical experience shows that relapse to alcohol use is very uncommon for someone who attends AA meetings regularly, particularly after being abstinent in AA for more than a year. However, for those who relapse, their attendance at AA meetings had ceased for a substantial period prior to the relapse, typically years

before the actual relapse to alcohol use. Also, use of other addicting drugs, such as cocaine, cannabis, heroin, and nicotine, and/or use of addicting medications, such as opiate analgesics and benzodiazepines, render continuous abstinence difficult and reliably predict relapse to alcohol.

## 8. BENEFICIAL VERSUS HARMFUL EFFECTS OF ALCOHOL

Some studies have indicated a modest but positive effect of reducing the risk of cardiovascular disease or heart attacks in those who regularly consume one to two drinks of alcohol per day. These studies are featured in influential medical journals, implicating alcohol as a therapeutic agent. Despite the overwhelming evidence of the harmful effects of alcohol on health, these studies advocate the use of alcohol without instructing individuals about its untoward or adverse effects.

For instance, studies show that the same amount of alcohol that is described to be beneficial (one or two drinks per day) will significantly reduce intelligence quotients (IQs), resulting in poor attention, poor concentration, and impaired short-term memory. Studies of alcoholics show that chronic alcohol use causes reduced cognitive capacities that tend to reverse with abstinence from alcohol. Importantly, the reduced IQs correlated with the age of the drinker; thus, the intended audience for putative efficacy on incidence of heart disease is those who are at high risk for the dementing properties of alcohol—namely older adults.

## 9. PREVENTION OF ALCOHOL USE AND PROBLEMS

Currently, our public health policy and controls for alcohol consumption are archaic and counterproductive. The example of advocating alcohol to reduce heart disease in high-risk populations is indicative. Thus, alcohol abuse is not commonly thought of as a

health issue but rather is regarded as a legal and moral issue. The policies usually do not support educating young people and adults on the responsible use and the potential pitfalls and risks of alcohol use, particularly by those at high risk. Most discussions focus on the legal age for drinking and the legal limit of intoxication. Not well discussed are that 80% of homicides occur in the setting of alcohol and that alcohol use and dependence is the leading risk factor for suicide at any age.

Despite these well-established adverse consequences and firm knowledge that onset of alcohol use and alcohol problems occur at young ages, there is blatant advertising clearly and convincingly targeted to these high-risk populations. Our controls over exposure to alcohol for vulnerable populations, and our education in elementary, middle, and high schools, are woefully lacking if we are to protect these vulnerable and at-risk populations. Currently, we do not hold producers of alcohol accountable for the high rate of alcohol-related consequences (in the billions of dollars) due to our lack of policy to protect the public against adverse consequences from alcohol use.

### See Also the Following Articles

Diagnostic and Statistical Manual of Mental Disorders ■ Drug Abuse ■ Drug Dependence

### Further Reading

- Grant, B. F. (1997). Prevalence and correlates of alcohol use and DSM-IV alcohol dependence in the United States: Results of the National Longitudinal Alcohol Epidemiologic Survey. *Journal of Studies in Alcohol*, 58, 464–473.
- McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *Journal of the American Medical Association*, 284, 1689–1695.
- Miller, N. S., & Goldsmith, R. J. (2001). Craving for alcohol and drugs in animals and humans: Biology and behavior. *Journal of Addictive Diseases*, 20, 87–104.



# Amnesia

Ginette Lafleche and Mieke Verfaellie

Memory Disorders Research Center, Boston University School of Medicine, Boston, Massachusetts, USA

1. Introduction
  2. The Contribution of Patient H.M.
  3. Impaired Memory in Amnesia: Declarative Memory
  4. Preserved Aspects of Memory in Amnesia: Nondeclarative Memory
  5. Amnesic Syndromes
  6. Neuropathology of Amnesia
  7. Approaches to Treatment of Amnesic Disorders
- Further Reading

## GLOSSARY

**anterograde amnesia** The inability to learn and recall events and facts following the onset of amnesia.

**declarative memory** Memory for concepts, facts, and personal events; declarative memories are accessible to consciousness and can be retrieved at will.

**episodic memory** A component of declarative memory; it refers to memories of personally experienced events, placed in specific spatial and temporal contexts.

**nondeclarative memory** A collection of unconscious forms of memory, including procedural memory, simple forms of conditioning, and priming.

**priming** A component of nondeclarative memory referring to the facilitation in processing information as a consequence of prior exposure to that information or to related information.

**procedural memory** A component of nondeclarative memory referring to the acquisition of skills and habits through repeated practice.

**retrograde amnesia** The inability to retrieve information about facts and events known prior to the onset of amnesia.

**semantic memory** A component of declarative memory referring to one's fund of knowledge of facts, concepts, and word meanings; these memories are impersonal and contain no reference to the time and place in which they were formed.

Through the study of amnesic patients, cognitive neuroscience has advanced our understanding of the processes that underlie learning and remembering. Memory is now understood to be a collection of functional systems, with each contributing uniquely to the storage and retrieval of information. The study of neurological populations with lesions that interfere with the function of specific components of memory has revealed distinct roles for these memory systems. This understanding has provided the theoretical foundation for evidence-based treatment approaches.

## 1. INTRODUCTION

In 1606, Shakespeare's Lady Macbeth described memory as "the warder of the brain," as a function that somehow stands apart from the brain's other cognitions while bearing witness to them. She made this reference while planning to induce a temporary amnesia in the king's attendants by getting them drunk. Her aim in doing so was to ensure that the attendants would not recall the events surrounding her husband's intended murder of the king. She knew that after the

deed was done and their alcohol-induced stupor passed off, they would be left witness to a blank, to a discontinuity in their experience, and not to the murder. For memory to be capable of keeping an account of the swift interplay of sensations, perceptions, and thoughts that make up events and thereby give experience its continuity, memory must itself have an underlying continuity. If it did not, there would be nothing for people to “go back to” that would allow them to witness an earlier event again, and Lady Macbeth would have had nothing to fear from any witnesses to her husband’s intended crime. Now, four centuries later, modern cognitive neuroscience is going beyond description. It investigates how memory enables people to actively bring the past into the present moment, thereby allowing them to experience their personal persistence through time.

The fact that memory can fail has been known for millennia, but its failure was not specifically called “amnesia” (from the Greek for forgetfulness) until relatively recently in 1786. In current common use, amnesia refers loosely and interchangeably to either one of two quite different states or to both of them. First, it refers to a state in which memories that were available for recall in the past have been lost and are no longer available in the present. Second, it refers to a state in which current experiences are not being kept account of and duly recorded, with the result that they will not be available for recall in the future. The first usage is reflected in the Oxford English dictionary definition of amnesia as “loss of memory” and roughly coincides with retrograde amnesia, whereas the second usage is reflected in the Merriam–Webster dictionary definition as “forgetfulness” and roughly coincides with anterograde amnesia. The loss of previously consolidated memories, on the one hand, and the inability to create and retain new memories, on the other, are states that can occur separately. However, more often than not, both occur together in the same individual following a single event, hence the tendency to refer to either or both of them without distinction by means of the same term “amnesia.”

In addition to its common meaning, the term “amnesia” has both a general sense and a specific sense in current scientific literature. In general, it refers to any pathological loss of the ability to acquire or recall information. In that sense, it encompasses psychogenic amnesia, posttraumatic amnesia following closed head injury, memory loss that occurs along with other major cognitive impairments in progressive degenerative neurological disorders such as Alzheimer’s disease,

and isolated global memory loss that characterizes the amnesic syndromes. In its more restricted use, it refers specifically to these latter syndromes, and it is in this sense that the term is used in what follows.

Early investigation of the amnesic syndromes proceeded both by means of clinical case descriptions and through the correlation of lesion analyses with deficits found through neuropsychological investigations. The most widely known of the early case studies was of the patient “H.M.,” and it is through studies of his memory function that the current understanding of amnesia began to unfold.

## 2. THE CONTRIBUTION OF PATIENT H.M.

At 9 years of age, H.M. suffered a head injury followed by seizures that became more frequent and more severe over the subsequent 18 years. In 1953, as a treatment of last resort, he underwent an experimental neurosurgical operation in which a large portion of the medial temporal region of his brain was removed bilaterally. That extensive resection was successful in reducing his seizures, but it left him with an unexpected devastating memory loss.

One striking finding that came to light in the early studies of H.M.’s memory loss was that his profound forgetfulness and loss of personal memories had occurred without any damage to his general intellectual ability, attention, and language. This discovery established that normal medial temporal lobe functioning is necessary for the accumulation of memories. However, the fact that its dysfunction does not interfere with the attention, thoughts, and perceptions that give rise to experiences in the first place suggests that memory is specifically linked to the function of the medial temporal lobe. Further studies of H.M.’s memory impairment led to the current understanding of the characteristics of hippocampal amnesia. They also revealed that memory, rather than being a unitary capacity, is composed of several functional systems, not all of which are impaired in amnesia. Those systems fall into two major classes. The first group involves nondeclarative memory, that is, the form of memory that functions without awareness. The second group involves declarative memory, that is, the form of memory that supports conscious retrieval of experiences. It is the latter form of memory that is impaired in amnesia.

### **3. IMPAIRED MEMORY IN AMNESIA: DECLARATIVE MEMORY**

After H.M.'s operation, it became apparent that he had suffered damage to the memory system that supports the conscious (intentional) retrieval of experiences. Not only did he fail to remember the events of his then current day-to-day life, but he also could not recall memories he had accumulated over much of his life prior to the onset of his amnesia. This clinical profile of severe anterograde amnesia together with an extensive retrograde memory impairment is typical of hippocampal amnesia.

#### **3.1. Retrograde Amnesia**

H.M.'s inability to recall memories that he had formed prior to the onset of his amnesia extended back approximately 11 years. The density of his retrograde amnesia followed Ribot's law, a semiquantitative description offered in 1881 that states that the "loss of memory is . . . inversely [related to] the time that has elapsed between any given incident and the fall (injury). . . . The new dies before the old." H.M. showed just such a temporal gradient. His inability to recall autobiographical and public events from shortly before his operation was total, but his amnesia was found to be less dense as testing probed further back in time and he was able to remember more and more.

Investigation of patients with a variety of amnesic syndromes has demonstrated that such a temporally graded retrograde amnesia is common and that its severity and duration tend to vary with the size of the brain lesion. Retrograde amnesia can affect both the recall of autobiographical and public events (episodic memory) and the retrieval of facts and concepts (semantic memory) that were known before the onset of amnesia. However, episodic and semantic memory are differentially affected across syndromes. For example, some patients have severe retrograde amnesia for events together with relatively preserved memory for facts and concepts, whereas others show the opposite pattern with severely impaired semantic memory but relatively preserved episodic remote memory.

The finding of a graded decrease in the density of memory loss as probing moves backward through time starting from the onset of amnesia establishes two important facts regarding how memories are stored and retrieved. First, it establishes that a

functional medial temporal region is required for permanent learning of new information. Second, it indicates that the medial temporal lobe is not the ultimate repository of long-term memories given that long-established memories can survive and be available for recall despite the presence of medial temporal damage. It has since been found that long-term memories are stored in the neocortex, specifically in the lateral temporal regions. This is confirmed by the fact that in patients with lesions in these areas, even old memories can be permanently lost, so that their retrograde amnesia is not temporally graded. Thus, the medial temporal lobes and neocortical areas make unique contributions to information storage. When an event is originally experienced, separate aspects of the event, such as its perceptual characteristics, emotional connotations, and associated thoughts, are processed in initially unconnected sites throughout the cortex, as determined by the functional specialization of various parts of the brain. The medial temporal lobes, and especially the hippocampus, provide the glue that allows these diverse aspects to be linked together into a single representation. When a part of any event is recalled again, the hippocampus is engaged, and it simultaneously reactivates the spatially separated cortical traces that together make up the brain's representation of the whole event. With repeated activation over time, links are established among the involved neocortical sites. Eventually, these separate neocortical sites that collectively represent a single event become interlinked to the extent that hippocampal activation is no longer necessary for recall of the event as a whole. It is at that point that the consolidated memory can survive medial temporal lobe damage.

#### **3.2. Anterograde Amnesia**

The nature of the impact of H.M.'s neurosurgery on his cognition gradually became apparent when it was noticed that H.M. consistently failed to recall having eaten a meal and that he failed to develop any ability to get to know the people who provided his daily care. This anterograde amnesia was so pervasive that he could hold onto new information for only a few seconds and then only if there was no distraction. If his attention was diverted, the information faded away and was not available for recall. As was noted earlier, the normal functioning of the hippocampus and surrounding medial temporal lobe structures is crucial for forming and storing new memories. In general, the severity of anterograde amnesia is proportional to the extent of

damage to medial temporal structures. Furthermore, because the hippocampus receives input from a variety of neocortical regions that process information in the various sensory modalities, such anterograde loss is invariably global; that is, it involves both verbal and nonverbal information in all possible modalities of presentation. Such dense global anterograde memory impairment is the hallmark of medial temporal lobe amnesia.

Anterograde amnesia is evaluated in the clinic through tasks that measure recall and recognition of information to which the individual is exposed during a testing session. Recall tasks require that recently learned information be consciously and deliberately retrieved in response to a question, whereas recognition tasks require that previously encountered information be distinguished from, and so recognized among, foils. The recognition format is generally easier for normal individuals because it requires less retrieval effort than does recall. Recognition is thought to be easier for individuals with intact memories because recall depends on conscious and effortful recollection, whereas recognition can be supported not only by recollection but also by familiarity. Familiarity refers to the sense of knowing that arises when a stimulus is processed with ease. Recollection and familiarity can be distinguished from one another experientially. For example, consider the situation where a man hears his name being called out while he is in a crowd. He turns and sees the woman who called his name and has a short conversation with her. The woman is familiar to the man, but he cannot recall any context that accounts for that familiarity. Several days later, the man receives a telephone call from the woman in which she identifies her place of work and, at that moment, he knows who she is and is flooded with memories of details and events. Sudden recollection replaces felt familiarity. In the laboratory, patients with amnesia secondary to medial temporal lobe damage generally show impairments on both recall and recognition tasks, suggesting that familiarity and recollection both are impaired. However, some amnesic patients show relatively preserved recognition memory, and this may reflect their residual ability to use familiarity as a basis for remembering.

Patients with lesions involving the frontal lobes, either directly or indirectly, can also have severe deficits in memory. However, although their performance is impaired on tasks of free recall, they show normal performance on tasks that require the recognition of previously learned information. This performance disparity between recall and recognition is

due to the distinct roles of the medial temporal lobes and the frontal lobes. The medial temporal lobes contribute to the encoding, storage, and retention of newly learned information, whereas the frontal lobes support the organizational and strategic aspects of memory necessary for developing encoding and retrieval strategies, monitoring and verifying memory output, and setting order within the recalled memories.

#### **4. PRESERVED ASPECTS OF MEMORY IN AMNESIA: NONDECLARATIVE MEMORY**

H.M.'s severe amnesia notwithstanding, some aspects of his memory functioning were found to be intact following his surgery. In contrast to his inability to consciously retrieve information acquired before and after onset of his amnesia, it was found that H.M. showed evidence of learning on tasks that did not require his awareness of the learning episode.

##### **4.1. Procedural Memory**

Careful experimental study brought to light that H.M. was able to acquire specific perceptual–motor skills as well as do normal individuals. One such motor task involved his tracing the outline of a star with a stylus while gaining feedback of his ongoing performance through the reflection of his performance in a mirror. H.M. improved with repetition exactly as did normal individuals. But for him, each trial was experienced as if it were his first trial, despite his steady improvement. He had no way of knowing that he was getting better at it, as normal people did. This finding was one of the first to show that procedural memory (i.e., the ability to know how) is distinct from declarative memory (i.e., the ability to know that something occurred). It also established that procedural memory does not depend on the integrity of the medial temporal lobes. Procedural memory develops gradually through repetition without any requisite awareness that the skill is being learned.

##### **4.2. Repetition Priming**

Priming is another form of learning that occurs without conscious awareness. It is defined as a bias or facilitation in identifying or responding to certain information to which one has been exposed previously. In contrast

to procedural learning, the effect of priming can be seen following a single exposure. In a typical priming task, an individual is asked to view a list of words or pictures. The individual is then asked to perform a seemingly unrelated task such as naming words or identifying degraded pictures. Unbeknownst to the individual, some of the stimuli in the second task were part of the list of stimuli that were seen in the first task, whereas others are presented for the first time. Priming is in evidence when naming or identification of previously seen stimuli occurs faster or more accurately than does identification of stimuli not seen previously. Amnesic patients show intact priming across a broad variety of tasks. Sometimes priming reflects facilitation of the processes that support the perception of stimuli, whereas other times priming reflects the facilitation of processes that support conceptual analysis (i.e., meaningful interpretation) of stimuli. Both perceptual priming and conceptual priming are independent of medial temporal structures, and both depend on distinct neocortical regions.

## 5. AMNESIC SYNDROMES

Amnesia can result from a variety of etiologies that cause damage to various regions of the brain. For example, diencephalic amnesia, the memory disorder of patients with Korsakoff's syndrome, arises from damage to structures that make up the diencephalon, particularly the mammillary bodies, the anterior thalamic nuclei, and the medial dorsal thalamic nucleus. Basal forebrain amnesia, in contrast, arises in patients with a history of rupture and repair of an anterior communicating artery (ACoA) aneurysm and is caused by damage to structures of the basal forebrain, the septum, and frontal brain regions. The following is an overview of the clinical characteristics of five of the more common amnesic syndromes.

### 5.1. Herpes Simplex Encephalitis

Herpes simplex encephalitis (HSE) is an acute inflammation of the brain caused by a herpes simplex virus infection. Some patients recover fully following infection, but most are left with a cluster of cognitive deficits that include a memory disorder. The heterogeneous pattern of these cognitive impairments reflects the extent and variability of damage to the brain. Among those patients who recover, a small number are left with a circumscribed amnesic syndrome similar to that of H.M. The fact that these two forms of amnesia share a clinical

profile is not surprising given that the herpes simplex virus preferentially affects medial temporal brain structures, including the hippocampus and adjacent entorhinal, perirhinal, and parahippocampal cortices. In all postencephalitic patients, the extent of any anterograde memory impairment is proportional to the amount of damage to the medial temporal lobes. Although some patients are incapable of any new learning, others can learn and can benefit from increased study time, external cues, and/or repeated exposure. Because the herpes virus can affect the brain asymmetrically, the pattern of anterograde memory loss will also depend on the laterality of damage. The greater the damage to the right temporal region, the greater the difficulty in performing nonverbal/visual memory tasks (e.g., memory for faces), whereas the greater the damage to the left temporal region, the greater the deficit in verbal memory.

Postencephalitic patients with damage extending to the lateral temporal lobes, the region where memories are permanently stored, will have dense retrograde amnesia for autobiographical and public information with little or no temporal gradient. Damage to the right anterior temporal region interferes primarily with retrieval of autobiographical memories, whereas damage to the left temporal cortex impairs semantic memory.

### 5.2. Anoxia

Amnesia can result from damage to, or death of, brain tissue due to a lack of oxygen supply to the brain. This can be caused either by reduced blood flow, such as in cardiac arrest or strangulation, or by normal perfusion with hypoxic blood secondary to respiratory distress or carbon monoxide poisoning. When the oxygen supply to brain tissue is disrupted, compensatory mechanisms that maintain cerebral homeostasis are triggered immediately. These protective autoregulatory mechanisms, although effective in adjusting for sudden short-lived changes, will eventually fail in the face of sustained oxygen deprivation. Oxygen deprivation for 3 to 8 minutes will trigger the release of excitatory neurotransmitters that can result in damage to the hippocampus. Shorter events may also damage regions that are perfused by terminal vascular branches or that lie in the watershed regions because these areas are deprived of oxygen early on. Longer lasting events will result in neuronal damage that extends to the cerebellum, the basal ganglia, the thalamus, and neocortical areas. The outcome following an anoxic event depends on many factors, including the cause and duration of the event as well as the age and health status of the individual.

Therefore, it is not surprising to find variability in the clinical profiles associated with anoxic events.

Most individuals who have suffered an anoxic brain injury experience a memory disorder along with other cognitive impairments. On occasion, an isolated amnesic syndrome is documented secondary to a lesion in the hippocampus. When the lesion is limited to the CA1 area of the hippocampus, there is moderately severe anterograde amnesia together with very mild retrograde loss. When the lesion extends beyond the CA1 area of the hippocampus but remains limited to the hippocampal formation, there is severe anterograde memory impairment and more robust retrograde memory impairment that can extend back more than 15 years. More commonly, however, anoxic patients have a memory impairment that is akin to that seen in patients with damage to the frontal brain region due to disruption of frontal–subcortical circuits as a result of damage to watershed zones in the cerebral cortex. Rather than encoding, storage, and retention deficits, these patients demonstrate impairments at the level of the organizational and strategic aspects of memory.

### **5.2.1. Anterior Communicating Artery Aneurysm**

Memory deficits, along with associated behavioral disorders, frequently follow rupture and surgical repair of an ACoA aneurysm. Such deficits vary from a very mild impairment to a severe amnesic disorder called basal forebrain amnesia. This wide variability reflects the fact that the ACoA perfuses a broad anatomical brain region, all of which is vulnerable in the event of rupture to damage from infarction, either directly or secondary to subarachnoid hemorrhage, vasospasm, or hematoma formation. Basal forebrain amnesia is usually due to damage to the septal nucleus and the subcallosal area. Disruption of hippocampal functioning may play a role in the amnesia of at least some patients with ACoA aneurysm because several basal forebrain nuclei contain a large number of cholinergic neurons that innervate the hippocampus as well as large neocortical regions. Disruption of frontal network systems can occur and may also contribute to the quality of the memory deficit seen in some ACoA aneurysm-related forms of amnesia.

The severe deficit in recall found in basal forebrain amnesia shares superficial similarities with the anterograde memory impairment that characterizes medial temporal lobe amnesia. However, the deficit in basal forebrain amnesia is due to inefficient encoding and not to the failure of consolidation that characterizes

medial temporal amnesia. As a result, ACoA patients will benefit from the use of encoding strategies, whereas patients with medial temporal lobe damage will not. Another difference is found in the performance of ACoA patients on recognition tasks, where they are found to succeed more often than their medial temporal lobe counterparts. This recall–recognition disparity is attributed to the frontal executive component of the amnesia that follows ACoA aneurysm, particularly the disruption of strategic effortful search processes that give access to information stored in memory. Another executive contribution to the disorder is the tendency of some ACoA patients to score many false positives on recognition tasks, a finding that suggests an additional impairment in the ability to monitor the outcome of a memory search. Retrograde memory is nearly always impaired in amnesia secondary to ACoA aneurysm, and it invariably shows a temporal gradient. However, patients do appear to benefit substantially more from cueing than do other amnesic groups, suggesting that impaired retrieval efficiency contributes to both the anterograde and retrograde aspects of their amnesia.

### **5.2.2. Stroke**

Amnesia is a common consequence of infarction of the posterior cerebral artery (PCA). It results from neural tissue damage caused by the interruption of blood flow to a large part of the medial temporal lobes, particularly the posterior two-thirds of the hippocampus, the parahippocampal gyrus, and other critical pathways that connect the hippocampus to surrounding brain areas. A more posterior extension of the lesion will result in other neuropsychological deficits in addition to the memory disturbance, for example, visual field defects and other visual disturbances that may affect reading and cause problems with color identification, space perception, and/or object naming. The typical memory disturbance associated with bilateral PCA infarction is an inability to establish new memories in the presence of preserved intelligence and attention. Retrograde memory problems are also often present. There have been several reported cases of bilateral PCA infarction that spared the medial temporal lobes proper but that involved the occipital lobes bilaterally as well as the deep white matter of both the occipital and temporal lobes. These patients present with a visual amnesic syndrome that results from the disconnection between occipital cortices involved in visual processing and temporal brain regions supporting memory. There have been other reported cases in which the PCA

infarction was unilateral. Patients with infarction of the left PCA present with a selective verbal memory deficit, whereas patients with infarction of the right PCA have preserved verbal memory but impaired visual processing skills and impaired visual memory.

Amnesia can also be caused by thalamic infarction. In such cases, the severity of the memory impairment is related to the site of damage within the thalamus. Lesions that damage the mammillo–thalamic tract, in particular, have been associated with severe anterograde amnesia. Infarction of the medial dorsal thalamic nuclei has also been associated with memory impairments, but it appears that the damage must extend beyond the medial dorsal nucleus to include the mammillo–thalamic tract or anterior nucleus for the development of a severe amnesic disorder. Because the thalamus has rich connections with the frontal lobes, this anterograde amnesia is also accompanied by an increased sensitivity to interference and by impairments in executive functioning. As with other amnesic syndromes, left-sided lesions result in impairments on tasks of verbal learning, whereas right-sided lesions result in nonverbal/visual memory impairments. Retrograde memory deficits following thalamic infarction are variable; some patients are found to have little impairment in remote memory, whereas others demonstrate severe long-term memory impairments.

### **5.2.3. Wernicke–Korsakoff Syndrome**

Wernicke–Korsakoff syndrome (WKS) is seen in patients with a history of long-term alcohol abuse in association with poor nutrition and a lack of Vitamin B1 (thiamine). In acute Wernicke's encephalopathy, patients exhibit confusion, a gait disorder (ataxia), and abnormal eye movements (oculomotor palsy). Treatment with large doses of thiamine may result in improvement in, or even reversal of, some of these symptoms. However, most patients are left with a permanent dense amnesic disorder referred to as Korsakoff's syndrome. This amnesic syndrome arises from damage to the thalamic nuclei, the mammillary bodies, and the frontal system.

Patients with WKS suffer from both anterograde and retrograde amnesia. Several explanations have been proposed to account for their episodic memory impairment. Although early models emphasized their superficial and deficient encoding strategies or their failure to inhibit competition from irrelevant material at the time of retrieval, current views agree that an explanation of their learning deficits is best accounted for by a theory that integrates both encoding and retrieval processes.

The retrograde amnesia in WKS has a steeper temporal gradient than that found in medial temporal lobe amnesia. The concomitant presence of frontal dysfunction in Korsakoff's patients is believed to account for their poorer performance on remote memory tests. Intelligence is usually preserved in Korsakoff's patients, but there are often associated cognitive and neurobehavioral deficits that are unique to this patient population. In particular, some combination of impaired planning and initiation, passivity, apathy, confabulation, and limited insight is nearly always found. These symptoms are thought to arise from associated frontal dysfunction.

## **6. NEUROPATHOLOGY OF AMNESIA**

For centuries, any lapse in the functioning of the warden of the brain's cognitions was described, and eventually defined, as simple forgetfulness. During the first half of the 20th century, it was thought that memory, as the cognition that keeps account of other cognitions, was not dependent on the activity of a circumscribed brain region (as are those other cognitions) but rather was directly dependent on the functioning of the whole brain. There also emerged an opposing view that held that memory function was localized in the brain. There was no conclusive experimental evidence for either view until, at midcentury, the consequences of H.M.'s surgical resection decided the matter in favor of a localized brain representation for declarative memory. Later evidence similarly established distinct localizations for priming and procedural learning, two forms of nondeclarative memory. Cognitive neuroscience has since clarified the nature of those localizations in substantial detail.

### **6.1. Declarative Memory**

Declarative memory is mediated by a group of interconnected structures that are part of the extended limbic system. Within the limbic system, two interacting memory circuits can be identified: the Papez circuit and the basolateral circuit. The Papez circuit is composed of the hippocampus, the fornix, mammillary bodies, the anterior thalamus, and the posterior cingulate gyrus (with additional connections to the basal forebrain via the fornix). The basolateral circuit is composed of the amygdala and surrounding perirhinal cortex, the dorsomedial thalamus, and the prefrontal cortex. Although damage to any part of either circuit in isolation will impair memory

function, damage to both circuits will result in a profound amnesic disorder. The contribution that each circuit makes to memory function remains a matter of active debate. On one side of the discussion are those who believe that both circuits are involved in all aspects of declarative memory. Lesions affecting both circuits would, therefore, be expected to result in more severe amnesia simply because more of the relevant neural tissue would be dysfunctional. Others suggest that the two circuits make qualitatively different contributions to memory. In particular, they suggest that the hippocampus and related structures in the Papez circuit support the recollection of episodic information, whereas the perirhinal cortices and related structures in the dorsolateral circuit support judgments of familiarity. In this case, damage to both circuits would impair more functions than would damage to either circuit alone and, therefore, would explain current clinical findings. Future studies of patients who present with selective lesions will be needed to resolve this debate.

The finding of temporally graded retrograde amnesia in association with damage to the medial temporal lobe reveals that this brain region plays a critical role in the establishment of memory and also suggests a subsequent slow transfer of memory to other brain regions. The finding that old memories, both autobiographical and semantic, are left untouched by damage limited to the hippocampus suggests that memories are not stored there. Long-term storage takes place in neural networks in the neocortex.

## 6.2. Nondeclarative Memory

Nondeclarative memory systems are supported by widely varying brain regions, depending on which sensory mode is involved in a given task and whether or not performance of the task involves higher associative functions. For example, evidence from neuroimaging studies, together with clinical data from individuals who have suffered focal cortical damage, has established that priming finds its substrate in the neocortex. Specifically, the substrate for perceptual priming is the relevant unimodal cortex (i.e., visual priming in the occipital visual cortex, auditory priming in the auditory cortex), whereas the substrate for conceptual priming is located in multimodal association cortices. The same brain regions that are involved in the initial processing of information are also involved in the more fluent processing that follows repetition, and priming is accompanied by a reduction in neural activity in these regions.

Although priming manifests after a single exposure to a stimulus, procedural memory requires a series of repetitions for its manifestation. Consequently, the involvement of relevant brain regions and their changes over time is more complex for procedural memory. The initial acquisition of motor skills engages the motor/prefrontal cortices, basal ganglia, and the cerebellum. Over time, however, two opposite tendencies are evident. On the one hand, as procedural learning develops and a skill can be performed with less effort, the prefrontal cortex and cerebellum gradually become less activated. This repetition suppression effect mirrors that seen in the neural substrate for priming. On the other hand, repetition leading to increased skill engages the higher order motor cortex more than it had initially been engaged. This latter finding indicates expanded cortical involvement in the retention of procedural memory for motor tasks that mirrors the increased neocortical involvement in the long-term retention of declarative memories.

## 7. APPROACHES TO TREATMENT OF AMNESIC DISORDERS

The treatment of memory disorders aims at enhancing day-to-day memory functioning and routine so as to increase an individual's level of independence. The choice of treatment will depend on both cognitive and noncognitive factors. Noncognitive factors to be considered include psychosocial context (e.g., family situation, educational background, lifestyle habits) and emotional factors (e.g., level of insight, motivation, neuropsychiatric symptoms). Of critical importance among these factors are level of insight and motivation. Research has shown that interventions are unsuccessful in patients who fail to appreciate that their memory is impaired and/or who are unmotivated. In these patients, efforts should initially focus on increasing level of insight and/or motivation. Cognitive factors that must be taken into account include premorbid abilities and skills and postmorbid neuropsychological deficits, including a clear delineation of those areas of memory that are impaired and preserved. Within the context of a holistic individualized approach, a memory remediation program can be developed based on current understanding of the processes that support learning and memory.

Several treatment approaches for densely amnesic patients capitalize on nondeclarative memory

processes because this form of memory remains intact in most patients. The “vanishing cues” technique is an example of a treatment method that recruits preserved implicit perceptual memory processes to teach patients domain-specific facts or concepts. The technique takes advantage of patients’ preserved ability to complete studied items in response to word fragment cues. In a typical vanishing cue paradigm, patients are given a definition and are then presented with as many letters as is necessary to produce the target word. With training, the letter cues are gradually reduced until the patients can spontaneously generate the sought after information. Success has been achieved using this technique in teaching patients computer-related vocabulary, business-related terms, and novel concepts. Learning by means of this technique is slow and laborious but can lead to surprisingly good retention, particularly if the information to be learned uses a knowledge base that is already familiar to the individual. However, a caveat arises from the inherent reliance of the vanishing cues method on the perceptual cues given during learning. As a result, generalization has often been limited, and benefits have been found to be best when the information is used in situations similar to those where learning occurred.

Attention to training contexts may be important when using techniques that take advantage of preserved implicit memory processes. Because amnesic patients have no recollection of the learning episodes, they fail to remember their mistakes and consequently fail to benefit from ongoing error correction. Instead, incorrect responses made during learning are often unconsciously repeated, leading to errors becoming primed and more likely being repeated subsequently. To avoid the perpetuation of errors through priming, some investigators have emphasized the importance of “errorless learning” for patients who have explicit memory impairments. In errorless learning, the possibility of making errors is eliminated by using cues and prompts or by providing the correct answer. The approach has met with some success in teaching memory-impaired patients both new skills (e.g., use of a memory book, programming of an electronic organizer) and new knowledge (e.g., learning of new words). Errorless learning is thought to operate by strengthening residual explicit memory, either alone or together with implicit memory. Its applicability as a method to facilitate learning appears to be broad and promising because errorless learning principles can be applied to a variety of remediation methods.

Other treatment approaches capitalize on preserved procedural learning. Through repetition, skills and habits that are important for activities of daily living or occupation can be taught. Such skills can range from simple assembly tasks to more complex multistep tasks such as learning to type. A variety of compensatory aids (e.g., notebooks, scheduling books, diaries, alarms) and augmentative technologies (e.g., computers, personal digital assistants, paging systems) rely on procedural memory. The notebook is an example of a low-tech aid. It is usually created at the early stage of rehabilitation and contains sections aimed at drilling overlearned personal information (e.g., date of birth, age, address), information about immediate family members (including their telephone numbers), daily schedules, and a daily record of activities. The book is tailored to the individual and can be gradually increased in complexity. Electronic organizers, the most favored external memory aid among normal individuals, are now also being used by memory-impaired patients. Training in the use of such technology requires lengthy practice sessions, within and outside the rehabilitation environment, to foster generalization. Because the acquisition of new skills is time-consuming for everyone involved, careful consideration needs to be given as to whether an electronic device is appropriate for an individual before investing the time and effort. Factors that would argue against it include dense amnesia associated with poor insight, lack of initiative, impaired visual attention, poor motor control, and limited problem-solving skills. Patients who are good candidates for this technology are generally younger, have experience in using electronic devices, and have achieved higher educational levels. Devices used premorbidly are preferable because familiarity increases the likelihood that they will be used effectively outside the clinic.

The preceding examples illustrate approaches that focus on preserved memory systems to teach new skills and habits. Another approach focuses on enhancing impaired forms of memory to improve day-to-day episodic memory by means of internal strategies. Internal strategies require awareness of the learning method and recall of the strategy itself; therefore, they are of limited use to patients who are moderately or densely amnesic. However, they are useful for patients who have mild memory deficits secondary to impaired effortful encoding or retrieval, who have good awareness of their deficits, and who have adequate motivation. These patients are more likely to generalize their training to situations that go beyond the clinic setting.

Examples of internal strategies include mental retracing, feature–name association, and verbal elaboration by means of a story or an association—all of which are skills that promote the use of imagery. The choice of technique will depend on the memory process that is targeted for remediation. For example, techniques that focus on strengthening encoding, and therefore storage, are effective at remediating consolidation problems. Story elaboration is effective in linking together a list of unrelated words through the development of a scenario that features the target words. The use of verbal associations is often effective in recalling a surname. In all of these instances, repeated use of the strategy is important and spaced repetitions, at different times and in different contexts, generally increases the likelihood that information will be learned and become conceptually integrated within a matrix of old memories. Strategies that are most effective when the deficit is at the level of strategic processes that enhance encoding are those that increase the organizational structure of incoming information. For example, learning how to “chunk” incoming information is helpful in streamlining and organizing that information. Organizing information according to themes or categories can also structure learning so that a thematic cue can serve to trigger recall when necessary.

These remediation methods are guided by knowledge regarding the cognitive processes that support learning and memory. More than 50 years ago, the early clinical findings with H.M. and other patients informed the theoretical understanding of the functional systems that comprise human memory. Current cognitive neuroscience has evolved from those early findings and now, in turn, can inform clinical approaches to remediation that enable amnesics to function more effectively in their daily lives.

### See Also the Following Articles

Aging and Culture ■ Cognitive Aging ■ Cognitive and Behavioral Interventions for Persons with Dementia ■ Dementia in Older Adults ■ Posttraumatic Disorders

### Further Reading

- Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral and Brain Sciences*, *22*, 425–489.
- Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 7041–7045.
- Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus?. *Nature Review Neuroscience*, *2*, 51–61.
- Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in H.M. *Seminars in Neurology*, *4*, 249–259.
- Moscovitch, M., Goshen-Gottstein, Y., & Vriezen, E. (1994). Memory without conscious recollection: A tutorial review from a neuropsychological perspective. In C. Umiltà, & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious processes in cognition* (pp. 619–660). Cambridge, MA: MIT/Bradford Press.
- Rempel-Clower, N. L., Zola, S. M., Squire, L. R., & Amaral, D. G. (1998). Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. *Journal of Neuroscience*, *16*, 5233–5255.
- Sagar, H. J., Cohen, N. J., Corkin, S., & Growdon, J. H. (1985). Dissociations among processes in remote memory. In D. S. Olton, E. Gamzu, & S. Corkin (Eds.), *Memory dysfunctions: An integration of animal and human research from preclinical and clinical perspectives* (Vol. 4, pp. 533–535). New York: Academy of Sciences.
- Scoville, W., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*, 11–21.
- Squire, L. R., & Kandel, E. R. (1999). *Memory: From mind to molecules*. New York: W. H. Freeman.
- Tulving, E. (1998). Neurocognitive processes of human memory. In C. von Euler, I. Lundberg, & R. Llinás (Eds.), *Basic mechanisms in cognition and language* (pp. 261–281). Amsterdam: Elsevier.
- Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus*, *8*, 198–204.
- Verfaellie, M., & Keane, M. M. (2002). Impaired and preserved memory processes in amnesia. In L. R. Squire, & D. L. Schacter (Eds.), *Neuropsychology of memory* (3rd ed.). New York: Guilford.



# Anxiety and Optimal Athletic Performance

John S. Raglin

*Indiana University, Bloomington, Indiana, USA*

1. Introduction
2. Traditional Theoretical Perspectives on the Anxiety-Performance Relationship
3. Sport-Specific Theories of Anxiety and Performance
4. Other Ideographic Theories
5. Summary  
Further Reading

## GLOSSARY

**anxiety** An emotion consisting of unpleasant thoughts and sensations as well as physical changes; it is a response to a situation or stimulus perceived to be threatening or dangerous.

**arousal** Traditionally defined as a generalized and undifferentiated physiological response to a stressor that is closely associated with negative emotions such as anxiety; research indicates, however, that the physiological responses to standardized stressors can differ significantly across individuals and are often unrelated to negative emotions such as anxiety.

Of all the psychological variables implicated in sport performance, anxiety is regarded to have the greatest impact. Putative explanations for why anxiety can degrade sport performance include diminished cognitive resources, a narrowing of the visual field, diminished motivation, and excessive muscle contraction or

co-contraction of opposing muscle groups that impairs coordination or results in physical injury. As a consequence, a wide variety of techniques have been used to assist athletes in controlling or reducing anxiety. Yet in spite of the general consensus within the field of sport psychology that anxiety harms sport performance, there are major disputes as to which theory best describes this relationship and as to what the most appropriate means are to measure anxiety in athletes. This article presents an overview of anxiety and sport performance literature. Limitations of traditional theoretical explanations are described, as are more recent sport-specific theories based on the premise of individual differences in the responses of athletes, with particular emphasis on the Individual Zones of Optimal Functioning (IZOF) model.

## 1. INTRODUCTION

Anxiety has been defined as an emotion consisting of dysphoric thoughts, unpleasant sensations, and physical changes that occur in response to a situation or stimulus perceived to be threatening or dangerous. According to most theories, anxiety consists of state and trait components. State anxiety indicates the intensity of anxiety experienced at a given moment and can fluctuate widely in intensity over a short time span.

Trait anxiety is a more stable factor that assesses the general tendency of an individual to experience elevations in state anxiety when exposed to stressors such as sport competition. Persons scoring high in trait anxiety should experience greater increases in state anxiety when exposed to a stressor than should persons scoring low in trait anxiety. The experience of anxiety as a response to a stressor such as sport competition is contingent on both an individual's perception of the stimulus and his or her ability to effectively cope with it. Because of this, sport competition may be perceived as threatening to some individuals, neutral to others, and enjoyable to still others.

### **1.1. Assessment of Anxiety**

In sport psychology research, a variety of approaches have been used to quantify anxiety, including the observation of overt behavior, biological activity (e.g., galvanic skin activity, heart rate, stress hormones), and self-reports. No single method is entirely reliable. Assessments of behaviors implicated in anxiety (e.g., pacing) may be an anxiety-reducing strategy for some individuals or may be entirely unrelated to anxiety in other instances. Physiological variables that have been used as biological correlates of anxiety (e.g., electromyogram [EMG]) may be difficult to assess prior to competition, or they may provoke an increase in anxiety in some cases (e.g., sampling blood to assess stress hormones).

Because of these problems, anxiety is most commonly determined by means of self-report questionnaires. In sport research, the most frequently used general measure of anxiety has been the State-Trait Anxiety Inventory (STAI), a 40-item questionnaire that assesses both state and trait anxiety. Despite its proven validity, the efficacy of the STAI and other general measures of anxiety in the context of athletics has been questioned, leading to the creation of more than 30 sport-specific anxiety scales. Among these, the most widely used is the Competitive State Anxiety Inventory-2 (CSAI-2), a multidimensional anxiety measure that assesses self-confidence, somatic anxiety, and cognitive anxiety.

Despite advantages such as the ease of administration and interpretation, self-report measures are not without limitations. The validity and reliability of self-reports are delimited by verbal ability and self-awareness of emotional states. Administering questionnaires near the time of competition can be impractical or disruptive and might even result in increased anxiety by directing attention to internal emotional states. A more serious problem is response distortion, which occurs when

individuals respond falsely to questionnaires for reasons such as social desirability, the demand characteristics of the experiment, and personal expectations. Response distortion can be detected through the use of lie scales, but this form of control is rarely used in sport psychology research.

## **2. TRADITIONAL THEORETICAL PERSPECTIVES ON THE ANXIETY-PERFORMANCE RELATIONSHIP**

It has been a long-standing belief in sport psychology that high levels of anxiety experienced during competition are harmful for performance and, if unabated, may even result in some athletes quitting their sport. A variety of interventions have been employed by sport psychology practitioners to reduce anxiety, including hypnosis, progressive relaxation, visualization, biofeedback, autogenic training, meditation, negative thought stopping, and confidence enhancement. However, it also has been posited that anxiety can facilitate performance under particular conditions. This perspective originally stemmed from drive theory, otherwise known as Hullian theory. According to Hullian theory, performance is a function of drive (i.e., physiological arousal or anxiety) and habit strength (i.e., skill). High levels of anxiety should increase the likelihood of correct behavior for well-learned skills, as would be the case for an emotional pep talk presented to a group of talented athletes. Evidence for drive theory in sport settings is lacking, however, and the theory currently has little status in the field of sport psychology.

Theoretical explanations in which high anxiety adversely influences performance have a higher standing in sport psychology, none more so than the Yerkes-Dodson law, familiarly known as the inverted-U hypothesis. The hypothesis stems from the classic work by Yerkes and Dodson, who in 1908 examined the influence of stimulus intensity on habit formation in experiments where mice were timed in maze running. Discrete levels of difficulty were created by manipulating the level of illumination of the maze and subjecting the mice to several intensities of stimulation via electrical shocks. The highest intensity shocks were found to slow learning under the most difficult (i.e., dimmest) maze trial, suggesting that moderate stimulation was best for such conditions. These results have since been widely reported in both general psychology and sport psychology textbooks, and they have been generalized to

a number of constructs such as drive, motivation, learning, arousal, and anxiety.

In sport psychology, the hypothesis is presented as a relationship between athletic performance and either arousal or, more commonly, anxiety. Optimal performance should occur when anxiety is within a moderate range of intensity, whereas deviations above or below this range should result in progressively worsened performance. Hence, anxiety and performance exhibit a relationship describing the shape of an inverted U. In basic terms, optimal performance is most likely to occur when anxiety is neither too high nor too low, but because of the stressful nature of sport competition, it is assumed that it is far more likely for athletes to experience too much anxiety.

### 2.1. Sport Modifications

The inverted-U hypothesis has been adapted to account for sports with different physical requirements (e.g., fine motor skill vs gross motor skill) and the expertise of the athlete performing the task. Sport tasks that require precise motor control and minimal physical effort (e.g., putting in golf) are posited to benefit when anxiety or physiological arousal is low prior to and during the tasks. As tasks require greater physical effort but less fine motor control (e.g., tackling in football), progressively higher anxiety intensities should enhance performance. The second modification predicts that as expertise and talent in sport tasks increase, the optimal range of anxiety will also be higher compared with athletes who are either beginners or intermediate in skill. It is assumed that more talented athletes possess the necessary motor skills and coping strategies to harness higher anxiety, whereas less skilled or experienced individuals should exhibit worsened performance at the same anxiety intensity. Given information about sport and skill level for a given athlete, it should then be possible to establish an inverted-U function for the individual.

Despite the continued popularity of the inverted-U hypothesis, reviews of the literature in general psychology and sport psychology have concluded that its empirical support is weak or even entirely absent. It has been concluded that much of the research supporting the hypothesis has little bearing on sport because it was conducted in laboratory environments rather than realistic sport settings or because it used nonathletes as test participants. Studies of the inverted-U hypothesis have also failed to support the propositions that optimal anxiety is altered by the motor skills required for

a sporting event or that comparably skilled athletes competing in the same sport benefit from similar anxiety levels.

Another concern expressed in reviews of the inverted-U hypothesis literature and other traditional theories is the assumption that arousal and anxiety are closely related or synonymous constructs. Arousal was originally defined as a global physiological response to a stressor that is closely associated with negative emotions such as anxiety. Subsequent research, however, has shown arousal to be a far more complex phenomenon. Standardized stressors evoke physiological responses that often exhibit little or no intercorrelation, and the pattern of physiological activation can vary considerably across individuals and situations. Physiological variables commonly employed as indicators of arousal (e.g., heart rate, respiration rate) are only weakly associated with self-report measures of anxiety, and this also is true for sport-specific measures of perceived arousal or somatic anxiety. Despite these findings, the conceptualization of arousal as an undifferentiated physiological response closely associated with anxiety persists, particularly within the field of sport psychology.

Finally, reviews of both the general and sport literature conclude that the inverted-U hypothesis and other traditional explanations indicate that athletes should respond uniformly to anxiety. They do not allow for the occurrence of interindividual differences to anxiety despite the fact that it has long been recognized that some athletes perform optimally under high intensities. In 1929, for example, the pioneering U.S. sport psychology researcher Coleman Griffith wrote, "Some of the most distressing cases of anxiety and fear in the dressing room have led to outstanding achievements during the game."

## 3. SPORT-SPECIFIC THEORIES OF ANXIETY AND PERFORMANCE

### 3.1. The IZOF Model

The lack of efficacy of traditional anxiety-performance theories, as well as the acknowledgment that individual differences contribute to this relationship, has spurred the development of sport-specific theories that incorporate this concept. Among these theoretical explanations, the Individual Zones of Optimal Functioning (IZOF) model is believed to have the strongest empirical basis. The IZOF model was developed by the Russian psychologist Yuri Hanin from studies of athletes in a wide variety of sports and competitive settings where anxiety was

**TABLE I**  
**Basic Tenets of the IZOF Model of Anxiety and Athletic Performance**

- 
1. Each athlete possesses an anxiety zone that is associated with optimal sport performance.
  2. The optimal anxiety zone may exist anywhere on the continuum of anxiety intensity, from low to extremely high.
  3. The optimal anxiety zone is not systematically influenced by either the nature of the sporting event or the skill and expertise of the athlete.
- 

assessed using the Russian-language version of the STAI. The results of Hanin's research supported an ideographic explanation, in which anxiety was associated with performance at the level of the individual athlete, rather than a nomothetic explanation, in which an entire team would respond similarly to anxiety. According to the IZOF model, each athlete possesses an optimal anxiety zone or range that is beneficial for performance. This optimal range may exist anywhere within the continuum of anxiety, from as low to as high as is measurable. When anxiety values fall outside the optimal zone, performance should decrease. The IZOF model further posits that the intensity of the optimal anxiety zone is not predictably influenced by either the type of sporting event or the skill of the individual athlete. As a result, an alternative method is needed to identify a more appropriate required anxiety level for an athlete than has been employed in traditional theories where the responses of athletes to anxiety are assumed to be more uniform (Table I).

### 3.2. Determining Optimal Anxiety

Two techniques have been developed to identify the optimal anxiety zone of an athlete: the direct method and the indirect method. For the direct method, state anxiety is assessed shortly before competitions until an athlete achieves a personal best performance. The optimal anxiety zone is then created by adding and subtracting 4 units to this anxiety score, which is approximately one-half standard deviation based on the STAI. The span of the optimal anxiety zone was established from initial research, but more recent studies have demonstrated that it can vary in width across athletes, indicating that some individuals can tolerate a wide range of anxiety intensity before experiencing a decline in performance. Unfortunately, the direct technique requires measuring anxiety prior to competitions

until the athlete achieves an outstanding or personal best performance, a process that may require months or even years and that is compounded when dealing with large teams consisting of many athletes.

As a more efficient alternative, Hanin developed an indirect method based on retrospection of past competitions. With the indirect method, athletes complete the state portion of the STAI according to how they recall feeling prior to their own best past performances or prior to performances judged to be optimal. Again, 4 anxiety units are added and subtracted from this "recalled best" anxiety score to establish the optimal zone. The accuracy of the indirect method has been tested by correlating recalled anxiety scores with anxiety values actually obtained at the time of the recalled events, and effect sizes ranging from .60 to .80 have been reported consistently. Research indicates that the levels of accuracy in recalling anxiety are comparable in athletes who performed either better or worse than expected despite concerns that performance outcome could bias accuracy. Although the effect size for recalled and actual precompetition anxiety supports the use of the indirect method for determining optimal anxiety ranges, occasional reports of inaccurate recall have been noted in the literature. In such instances, the direct method should be used.

### 3.3. Interindividual Variability in Optimal Anxiety

The IZOF model predicts not only that athletes of a similar caliber in a given sport will differ in the intensity of optimal anxiety but also that a significant proportion will benefit from high anxiety. The findings from studies of athletes in a number of sports and different levels are consistent with these propositions, indicating that between 25 and 50% report that best performances occur when anxiety levels are elevated. As predicted by the IZOF model, the proportion of athletes who perform best at high anxiety intensities is not related to skill or even age. For example, in studies of elite U.S. distance runners, it was found that 30% of female runners reported that best performances were most likely when anxiety was significantly elevated, but percentages as high as 51% have been noted in nonelite college track and field athletes. Even in track and field athletes as young as 9 to 12 years, more than one in four indicated that they performed best with high anxiety. These results and other findings provide support for the IZOF model, but they are not consistent with group-based explanations of anxiety and performance. For example, according to both

drive theory and the inverted-U hypothesis, the proportion of individuals who benefit from high anxiety should consistently be higher for elite athletes than for less talented competitors.

The evidence of wide-ranging variability in precompetition and optimal anxiety complicates the use of intervention strategies designed to regulate anxiety. Group-based interventions, in which an entire team of athletes receives a single treatment such as relaxation, are easily administered but would be ineffective or counterproductive for those athletes who perform best at moderate or high anxiety intensities. On the other hand, it can be time-consuming to use IZOF procedures to assess anxiety at the time of competition and then compare these values with each athlete's optimal zone to determine who will be in need of some form of intervention.

### 3.4. Predicting Precompetition Anxiety

In an effort to provide a means for coaches and psychologists to efficiently identify those athletes who will require anxiety intervention, as well as to determine the appropriate direction of the intervention (e.g., increase or decrease anxiety), IZOF studies have tested the ability of athletes to anticipate the intensity they will experience prior to actual competition. In this research, athletes completed the state portion of the STAI several days prior to a competition with instructions to respond according to how they anticipated they would feel immediately before the competition. Athletes then completed the STAI again just before the competition under the standard instructional set (i.e., "right now"), and that score was then compared with the predicted anxiety score. To minimize intrusiveness, the questionnaires were sometimes completed at a prescribed time before competition (e.g., 60 minutes). The results of this work reveal that predicted anxiety scores correlate quite closely with actual values. Correlations between predicted and actual precompetition anxiety range between .60 and .80, with higher coefficients occurring for difficult or important competitions.

Predicted precompetition anxiety scores are useful in situations where it would be difficult or impossible to assess anxiety just prior to the actual competition. The discrepancy between predicted precompetition anxiety and IZOF values can be contrasted to identify athletes who are likely to be too relaxed or too anxious, and this can be done several days before the competition.

From a practical perspective, the extent to which anxiety deviates from the optimal zone can help dictate how much anxiety must be increased or decreased to reach the optimal zone. Simple techniques that are easily implemented by the coaching staff can be an expeditious means to manipulate anxiety. These include emphasizing or deemphasizing the importance of the competition or the expectations of an athlete's performance.

### 3.5. Impact of Optimal Anxiety on Performance

Research based on IZOF procedures indicates that deviations in precompetition anxiety from the optimal zone have a significant impact on performance. Studies of athletes in sports such as swimming, rhythmic gymnastics, and ice skating indicate that precompetition anxiety of successful performers is closer to their own optimal anxiety values than is the case with their less successful teammates. In most cases, performance differences are evident only in difficult competitions, and it has been speculated that optimal anxiety is not necessary to achieve adequate performance in easy or unimportant competitions. In research that has examined the net impact of anxiety on sport performance, it has been found that performances were approximately 2% worse on average in cases where precompetition anxiety fell outside the IZOF and that the decrement was approximately equal whether anxiety was lower or higher than optimal.

Despite evidence that supports the major tenets of the IZOF model, the model has been criticized on a number of grounds. One primary concern is that the IZOF model does not indicate what variables contribute to the differences in optimal anxiety observed in homogenous groups of athletes. To date, studies of the factors contributing to interindividual variability in optimal anxiety have been rare, and the bulk of IZOF research has focused on examining the validity of the major tenets of the model. A second line of criticism contends that IZOF research based on the STAI or other nonspecific and general measures of anxiety is inadequate, whereas more complete results would be yielded by employing a sport-specific measure assessing multiple aspects of anxiety, particularly the CSAI-2. However, the results of CSAI-2 research on the IZOF model have been less consistent than research based on the STAI, and the IZOF model has been found to be less accurate than the STAI in assessing both recalled and predicted precompetition anxiety.

#### 4. OTHER IDEOGRAPHIC THEORIES

Two other major theories of anxiety and sport performance have been adopted recently for sport: reversal theory and catastrophe theory. Although the concept of individual differences are not central to these theories, like the IZOF model, they acknowledge that anxiety can either facilitate or harm sport performance. Unlike the IZOF model, both reversal and catastrophe theories incorporate specific anxiety or perceived arousal scales.

Reversal theory predicts that self-reported arousal is important to performance despite a lack of evidence that self-reports can provide an objective indication of physiological activity. Arousal is interpreted on the basis of an individual's current emotional state that results from the interaction of oppositional high-arousal preferring (paratelic) and low-arousal preferring (telic) states that are also assessed via self-reports.

Catastrophe theory uses an arousal-related measure referred to as somatic anxiety, which is assumed to exhibit an inverted-U relationship with performance. In addition, measures of self-confidence and cognitive anxiety are assessed using a modified version of the CSAI-2 that assesses facilitative and debilitating anxiety. Cognitive anxiety is posited to be negatively related to performance, whereas self-confidence exhibits a positive relationship. When these dimensions are considered collectively, they form a complex three-dimensional figure referred to as a butterfly or catastrophe cusp. Reviews of the efficacy of reversal and catastrophe theories have been mixed, and this may stem in part from the challenges of validating these complex theories empirically.

#### 5. SUMMARY

The growing realization that traditional theoretical explanations of anxiety and performance fair poorly when applied to sport has led to the development of several models and theories of anxiety specific to athletes. Further research is needed before definitive judgments can be made about the relative efficacy of these theoretical explanations, but they all indicate that influence of anxiety on sport performance is more complex than is predicted by traditional explanations such as the inverted-U hypothesis. In particular, the results of IZOF model research reveal that the anxiety intensity associated with optimal sport performance varies considerably across athletes, even for those competing in the same competition. This research also indicates that a

substantial percentage of athletes actually benefit from elevated anxiety and that in these cases, interventions aimed at reducing anxiety may be counterproductive.

#### See Also the Following Articles

Arousal in Sport ■ Performance Slumps in Sport: Prevention and Coping ■ Psychological Skills Training in Sport

#### Further Reading

- Duffy, E. (1957). The psychological significance of the concept of "arousal" or "activation." *Psychological Review*, 66, 183–201.
- Fazey, J., & Hardy, L. (1988). *The inverted-U hypothesis: A catastrophe for sport psychology?* Leeds, UK: White Line Press.
- Griffith, C. R. (1929). *The psychology of coaching*. New York: Scribner.
- Hackfort, D., & Schwenkmezger, P. (1993). Anxiety. In R. M. Singer, M. Murphey, & L. K. Tennant (Eds.), *Handbook of research on sport psychology* (pp. 328–364). New York: Macmillan.
- Hackfort, D., & Spielberger, C. D. (Eds.). (1989). *Anxiety in sports*. New York: Taylor & Francis.
- Hanin, Y. L. (Ed.). (2000). *Emotion in sports*. Champaign, IL: Human Kinetics.
- Hull, C. L. (1943). *Principles of behavior*. New York: Appleton.
- Martens, R., Vealey, R. S., & Burton, D. (1990). *Competitive anxiety in sport*. Champaign, IL: Human Kinetics.
- Neiss, R. (1988). Reconceptualizing arousal: Psychobiological states in motor performance. *Psychological Bulletin*, 103, 345–366.
- Ostrow, A. C. (1996). *Directory of psychological tests in the sport and exercise sciences* (2nd ed.). Morgantown, WV: Fitness Information Technology.
- Oxendine, J. B. (1970). Emotional arousal and motor performance. *Quest*, 13, 23–32.
- Raglin, J. S., & Hanin, Y. L. (2000). Competitive anxiety. In Y. L. Hanin (Ed.), *Emotion in sports* (pp. 93–111). Champaign, IL: Human Kinetics.
- Raglin, J. S. (1992). Anxiety and sport performance. In J. O. Holloszy (Ed.), *Exercise and sport sciences reviews* (Vol. 20, pp. 243–274). New York: Williams & Wilkins.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482.



# Anxiety Disorders in Late Life

**Cheryl N. Carmin**

University of Illinois at Chicago,  
Illinois, USA

**Jan Mohlman**

Syracuse University, Syracuse,  
New York, USA

**Amy Buckley**

University of Illinois at Chicago,  
Illinois, USA

1. Epidemiology of Anxiety Disorders in Older Adults
  2. Assessment and Treatment of Anxiety Disorders
  3. Conclusions
- Further Reading

## GLOSSARY

**cognitive behavior therapy** A collaborative form of psychotherapy that involves facilitating patients' changing their belief system via a Socratic dialogue focusing on those distressing thoughts that exacerbate negative emotions and maladaptive behaviors.

**ego-syntonic/ego-dystonic obsessions** Terms typically used in describing whether obsessions are congruent or incongruent with one's beliefs; an ego-dystonic obsession may include the belief that washing one's hands for 15 minutes will prevent cancer with the understanding that, despite the persistence of the thought, it is illogical.

**epidemiology** A branch of medical science that deals with the incidence/prevalence, distribution, and prevention of a disease or condition as well as the identification of at-risk populations.

**Likert scale** A means of assigning a numerical intensity rating that is anchored by descriptive terms to a construct (e.g., I would describe my worry as 1 = *very mild*, 2 = *mild*, 3 = *average*, 4 = *excessive*, 5 = *very excessive*).

Recent epidemiological studies have underscored the ubiquitous nature of anxiety disorders, with

approximately 25% of adults being affected over the course of their lifetimes. Furthermore, the economic burden of these disorders has been estimated at \$42.3 billion, comprising 31% of psychiatric treatment costs. Given the prevalence of anxiety disorders, it is not surprising that an increasing amount of attention has been given to investigating the prevalence and treatment of these conditions. What is surprising, however, is how little attention has been given to anxiety disorders in what is the fastest growing segment of the population, namely the elderly. This article summarizes how the existing research literature informs us with respect to the epidemiology of anxiety disorders in the elderly and then examines the treatment outcome literature with regard to the individual anxiety disorders.

## 1. EPIDEMIOLOGY OF ANXIETY DISORDERS IN OLDER ADULTS

The United States is aging. The U.S. Census Bureau predicts that more than 72 million adults will be age 65 years or over by the year 2030. Recent epidemiological studies underscore that anxiety disorders are the most prominent of the psychiatric conditions, with lifetime estimates nearing 25%. Despite the widespread prevalence and extensive cost to society of anxiety disorders, there is currently little data on the rate and phenomenology of anxiety disorders in the elderly population.

### **1.1. General Prevalence Data**

It is generally believed that anxiety disorders occur less frequently in the elderly population than in younger adults. However, there have been only a small number of epidemiological studies that have tested this contention directly. A recent epidemiological study in the United States that included persons over 65 years of age is the Epidemiological Catchment Area (ECA) study, which included more than 18,000 noninstitutionalized adults. The ECA study systematically examined the rate of anxiety disorders at five sites across the country. Fully 30% of the ECA subjects were age 65 years or over. It was ascertained that the 1-month prevalence rate of anxiety disorders in the elderly was approximately 5.5%, lower than the 7.3% estimate for all adults surveyed. Furthermore, elderly women were nearly twice as likely to have anxiety disorders as were elderly men. The percentage of elderly adults with anxiety disorders was higher than that with any other psychiatric illness, including cognitive impairment, underscoring the need for appropriate identification and treatment in this population. Furthermore, rates of anxiety disorders were much higher than the 2.5% prevalence rate of affective disorders in this population. Other reviews of smaller scale epidemiological studies found that rates of anxiety disorders ranged between 0.7 and 19.0%. Similar 6-month prevalence rates of 3.5% have been found in Europe.

### **1.2. Epidemiology of Individual Anxiety Disorders**

Rates for specific anxiety disorders also vary across studies and differ between elderly males and females. Data from the ECA study suggest that phobias are the most frequently experienced anxiety disorder in the elderly, affecting an estimated 4.8% of older adults. Panic disorder (PD) and obsessive-compulsive disorder (OCD) in the elderly population are expected to occur only infrequently. Rates for PD are estimated to not exceed 0.3%, and when panic does occur in elderly adults, it tends to be in women. Similarly, rates of OCD among the elderly are low, with prevalence estimates being 3.5% at a maximum and with elderly adults residing in institutional settings accounting for the upper end of these prevalence estimates.

Data regarding the prevalence of generalized anxiety disorder (GAD) suggest that it has a much more variable occurrence in the elderly population, ranging between 0.7 and 7.1%. It is important to note that data on

the prevalence of GAD are not included in the ECA data and that other studies that find higher prevalence rates of GAD may be due to differences in their respective methodologies, for example, the decision rules applied in rendering a diagnosis.

### **1.3. Reliability of Epidemiological Studies**

There are several potential explanations for the wide variability of estimates of anxiety disorder prevalence in the elderly. The existing epidemiological studies use different methodologies, and this often makes it difficult to compare prevalence rates across individual studies. Some of the studies include institutionalized elderly adults, whereas others survey community-dwelling residents only. In addition, measures of anxiety symptoms validated on younger adults might not be applicable to older adults given that the experience and expression of anxiety may change with age. Likewise, the instruments used to assess anxiety in older adults might not have norms or other psychometric data that are established for older adults, let alone for very old adults. Further research is needed to delineate factors that contribute to the difference in prevalence estimates and to clarify the nature of anxiety in the elderly. Until then, estimates of prevalence rates of anxiety disorders in the elderly should be considered preliminary.

### **1.4. Comorbidity**

It is well known in the literature that anxiety disorders often co-occur with other diagnoses, including depression and other anxiety disorders. However, it is important to examine whether this is true for elderly persons as well. Research is beginning to shed light on the co-occurrence of anxiety and other psychiatric disorders in older adults. As with younger adults, depression most frequently co-occurs with anxiety disorders among elderly patients. In addition, when depression is the primary diagnosis in elderly adults, anxiety frequently co-occurs. In general, the delineation between depression and anxiety is not clear, and research suggests that it might be even less clear for older adults.

Unlike anxiety or depression, the incidence of cognitive impairment increases with age. Research examining the relationship between anxiety and dementia in older adults typically finds that these syndromes often coexist. Symptoms of anxiety often occur in the

context of dementia as well as in nondemented older adults.

In summary, although the frequency with which anxiety disorders are present in older adults is lower than that in younger adults, these data may be confounded by methodological problems such as differences in diagnostic classification, commonalities between anxiety and depression in this population, and the frequent occurrence of anxiety symptoms accompanying medical conditions.

## 2. ASSESSMENT AND TREATMENT OF ANXIETY DISORDERS

### 2.1. Generalized Anxiety Disorder

#### 2.1.1. Assessment

Diagnostic criteria for GAD have undergone considerable revision over the past 20 years or so. The most recent version of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) emphasizes chronic uncontrollable worry that causes significant distress or impairment as the hallmark symptom of the disorder. In adults of all ages, GAD is most often comorbid with other affective and anxiety disorders and is sometimes viewed as a vulnerability factor for the development of additional psychiatric problems. GAD is often secondary to depressive disorders in older adults.

As noted by Carmin and colleagues in 1999, the only clinician-rated diagnostic interviews that have been found to yield reliable diagnoses with elderly patients are the Structured Clinical Interview Diagnostic (SCID) and the Anxiety Disorders Interview Schedule (ADIS-IV). Low interrater reliability estimates were reported during the 1980s for GAD; however, better estimates have been found using DSM-IV criteria. A recent investigation using the SCID and the ADIS yielded substantial diagnostic agreement for all of the anxiety disorders in an older treatment-seeking sample. Interrater reliability for GAD with the ADIS was similarly found to be excellent.

Two additional clinician-rated measures have been used to determine anxiety severity in the elderly. The Hamilton Anxiety Rating Scale (HARS) is a 14-item scale tapping anxiety symptoms in somatic, psychic, and affective domains. In 1997, Beck and Stanley found that the HARS distinguished older adults with GAD from normal controls, providing some preliminary support for the measure. The FEAR is a relatively

new instrument developed to measure GAD in older adults in the primary care setting. The FEAR, a 4-item version of the 11-item Anxiety Disorder Scale, is meant to be administered verbally during routine medical exams or in the waiting room. The FEAR demonstrated very good sensitivity and specificity in an initial administration to 88 older medical patients, 27% of whom were diagnosed with GAD. Additional psychometric properties of the measure await investigation.

There are only a few self-report measures that are recommended for use with late-life GAD patients. The Worry Scale (WS) is a 35-item, self-report questionnaire tapping three subscales of worry domains of importance to older adults—finances, health, social issues—using a 5-point Likert scale. The primary utility of the measure is the assessment of GAD. However, the use of the measure with a sample of older adults with GAD yielded poor convergent validity, with low correlations with other measures of anxiety. The internal reliability estimate was .93, but test–retest reliability was not reported. In a second study by Beck and Stanley in 1997, the measure showed adequate internal reliability and convergent validity. Currently, the properties of this measure seem to be better established in healthy controls than in patient samples.

The Penn State Worry Questionnaire (PSWQ) consists of 16 items on a 4-point Likert scale designed to tap the generality, intensity, and uncontrollability of worry. The measure has demonstrated good psychometric properties (e.g., internal consistency, convergent validity) in several studies of older adults with GAD. However, the PSWQ failed to discriminate between older adults with GAD and PD and has shown inadequate test–retest reliability in older GAD patients over a variable period with a mean of 70 days. Stanley and colleagues found that the PSWQ showed good divergent validity with measures of depression. Because late-life GAD is also comorbid with depressive disorders, this is one appealing feature of the measure.

The trait scale of the State–Trait Anxiety Inventory (STAI) assesses the tendency to experience frequent anxiety and nervousness. The measure has 20 items on a 4-point Likert scale and is one of the most frequently used self-report questionnaires in studies of anxiety. In heterogeneous older adult samples, the trait scale has shown adequate psychometric properties, and there are ample normative data. In 1996, Stanley, Beck, and Zebb found low to moderate correlations between the STAI and other anxiety and worry measures in older GAD patients but found good internal consistency of .88 for the trait scale. Test–retest

reliability estimates were not given for the GAD sample but were good in a sample of healthy older adults. A more recent investigation showed good internal consistency but low test–retest reliability (.58) in a sample of 57 older adults with GAD. Mohlman, deJesus, and colleagues found that scores on the measure were very similar in older adults with GAD and PD, suggesting that it taps a nonspecific construct in older samples.

### **2.1.2. Psychosocial Treatment of Late-Life Generalized Anxiety Disorder**

Two early studies by Stanley and colleagues indicated that supportive group therapy was beneficial for late-life GAD treatment. The 14-week intervention focused on the discussion of symptoms and experiences and providing support for group members.

Investigations of individual format cognitive behavior therapy (CBT) have suggested that this treatment is effective in older adults with GAD. However, it is notable that most of these studies have included nonstandard augmentations to the therapy such as concurrent medication and weekly meetings with a physician, treatment conducted in primary care or in patients' own homes, and the use of learning aids. The only investigation of standard individual CBT delivered in a mental health clinic indicated very modest efficacy as compared with a wait list condition. Currently, the efficacy of individual CBT is not well supported.

Several studies assessing the efficacy of group CBT for late-life GAD indicate that the treatment is typically more effective than wait list control conditions; however, it is not significantly better than other control conditions. Thus, it is possible that nonspecific elements of the group format (e.g., increased social interaction, mitigation of loneliness), rather than elements specific to CBT, led to improvement. Studies comparing group format CBT with individual format CBT should help to clarify this issue.

## **2.2. Panic Disorder and Agoraphobia**

### **2.2.1. Assessment**

Notably, very few studies have addressed the psychometric properties of measures of panic and related symptoms in older patient samples. Two studies, one using a nonclinical community sample and the other using patients drawn from a medical clinic, studied the properties of the Beck Anxiety Inventory (BAI), a well-known

measure of anxiety and panic symptoms and found that the scale had good discriminant validity and internal consistency. The latter study also demonstrated the BAI four-factor solution with autonomic, neuromotor, cognitive, and panic subscales, suggesting that anxiety symptom clusters are slightly different from those found among younger adults.

The Anxiety Sensitivity Index (ASI) is a 16-item measure tapping the fear of anxiety sensations, which is known to be a risk factor for the development of panic. In 2000, Mohlman and Zinbarg tested the structure and validity of the ASI in 322 healthy older adults (mean age 75 years). The ASI showed strong internal consistency and moderate correlations with measures of related constructs. Confirmatory factor analysis indicated a hierarchical structure with three group factors—physical concerns, mental incapacitation concerns, and social concerns—as well as a general factor, consistent with previous investigations of the ASI in younger adults. In 1998, Deer and Calamari found that 49% of their older sample (mean age 81 years) reported panic symptoms and that 27% reported a panic attack during the past year. Anxiety sensations predicted unique variance in panic symptomatology and may function as a risk factor for the development of late-life panic.

### **2.2.2. Psychosocial Treatment of Late-Life Panic Disorder**

Trials of psychosocial treatments in older samples with PD, panic disorder with agoraphobia (PDA), or agoraphobia without history of panic disorder (AWOHPD) are limited to case studies and three small pilot studies. One study found that principles of reality therapy, which focuses on an individual's situation and worldview, were effective when used by a neighbor to mitigate an older adult's paranoia and agoraphobia.

Early investigations of behavioral treatments conducted during the 1970s included relaxation, imagery, and exposure. In 1996, Rathus and Sanderson used CBT with two older panic patients: one 70-year-old male and one 69-year-old female. Treatment components were education, cognitive restructuring, interoceptive and situational exposure, and diaphragmatic breathing. Both participants achieved panic-free status and decreased depression following 4 to 5 months of therapy.

In 1991, King and Barrowclough tested CBT for panic and anxiety in a small sample of adults ages 66 to 78 years. Of the 10 participants, 8 had primary diagnoses of PDA. After treatment, 7 were free of

panic and 2 showed decreased symptom severity. Six months later, 8 of the remaining 9 participants were panic free and 6 showed improvement on depression.

In 1996, Swales and colleagues tested 10 90-minute sessions of CBT in 15 adults ages 55 to 80 years. Participants experienced decreased severity and frequency of panic attacks, depression, avoidance, and role impairment, and this was apparent at both post-treatment and 3-month follow-up. A reanalysis of Gorenstein and colleagues by Mohlman indicated that CBT plus medication management ( $n = 5$ ) was somewhat more effective than medication management alone ( $n = 5$ ) in assisting older adults with PD to decrease anxiety while tapering off anxiolytic medication. The use of interoceptive exposure was believed to facilitate habituation to sensations related to PD and medication withdrawal simultaneously.

### 2.3. Social Anxiety Disorder

Although the social anxiety literature has grown tremendously during the past several years, there is still a relative dearth of empirical research examining the assessment and treatment of social anxiety in older adults. This may be due in part to the fact that in epidemiological studies, social phobia is a relatively rare disorder among the elderly population. In addition, social anxiety as a distinct diagnostic category did not appear until the third edition of the DSM was published.

#### 2.3.1. Assessment

As noted previously, the SCID and the ADIS-IV are the only two clinician-administered diagnostic instruments with published data on older adults. In 1993, Segal and colleagues reported interrater reliabilities ranging from good to excellent across anxiety disorder diagnoses with use of the SCID.

The ADIS-IV is considered the gold standard for diagnosing anxiety disorders in adults. In 2001, Brown and colleagues found the interrater reliability of the lifetime version of this instrument to be excellent for social anxiety on a population of younger adults. Its reliability for diagnosing social phobia in older adults has not yet been established.

There are currently a number of self-report measures available for assessing social anxiety in younger adults. Unfortunately, none of the measures that has been used with younger cohorts has norms or other psychometric data supporting its use with older adults.

#### 2.3.2. Treatment

In contrast to the abundant literature examining psychological and pharmacological treatment of social anxiety disorder in adults, there are no published studies examining the efficacy or effectiveness of treatments for socially anxious older adults. In 2001, Fedoroff and Taylor conducted a meta-analysis of 108 studies. These researchers compared pharmacological treatments, including selective serotonin reuptake inhibitors (SSRIs), benzodiazepines (BZDs), and monoamine oxidase inhibitors (MAOIs), with components of CBT, including exposure and cognitive restructuring. Studies were included in the analysis if they consisted of at least four patients diagnosed with social anxiety using clinical interviews and consisted of standard treatments of social anxiety. The ages of the patients included in the meta-analysis were not provided, although all studies consisted of adults. Results of the analysis suggested that pharmacotherapies were superior to CBT at posttreatment and that both were superior to controls. BZDs were significantly more effective than both CBT and controls but did not significantly differ from SSRIs. Long-term maintenance of gains was not as well established because many of the studies did not contain sufficient information for estimating this variable. Although these results may have applicability to older adults, the use of BZDs is ventured very cautiously given that this class of medication can cause difficulties with dizziness, respiration, and cognitive functioning.

In summary, the literature is scarce regarding effective treatments for social anxiety in older adults. Although both CBT and pharmacotherapy have been proven to be effective treatments, extrapolation to older adults is premature. More research is needed in this area before conclusions can be made concerning best practice treatments of social anxiety treatment in older adults.

### 2.4. Assessment and Treatment of Obsessive–Compulsive Disorder

The observation made by McCarthy and colleagues more than a decade ago, that there is a paucity of research evaluating the efficacy of treatments for older adults diagnosed with OCD, still remains true today. The limited literature that exists has focused primarily either on a variation of CBT called exposure and response/ritual prevention (ERP) or on pharmacological treatment.

### **2.4.1. Assessment**

The assessment literature, with regard to OCD in older adults, offers few recommendations other than the Padua Inventory, which was used with older adults diagnosed with GAD. Some authors have suggested that certain presentations of OCD, such as obsessions and compulsions related to fear of forgetting names and pronounced ego-syntonic scrupulosity, are more likely to occur in the elderly. Besides these case reports, there is little evidence that particular constellations of obsessions and compulsions are unique to older adults. Typical presentations of OCD, such as contamination fears with washing rituals and fears of harming others accompanied by checking compulsions, are also commonly found in the elderly. DSM-IV criteria appear to be appropriate for use with older adults, but more research is needed using structured diagnostic instruments such as the ADIS-IV to better characterize the presentation of OCD in the elderly.

### **2.4.2. Treatment**

Two early uncontrolled case studies of elderly OCD patients reported significant reductions in OCD symptoms following ERP, whereas one anecdotal report documented unsuccessful combined ERP and medication treatments in a 74-year-old woman with OCD and a learning disability. Several controlled outcome studies of ERP using single-case designs yielded similarly successful results. The one study that directly compared treatment responses of older patients with those of younger patients examined the effectiveness of inpatient ERP administered to 10 severely impaired OCD patients age 60 years or over and 10 younger OCD patients matched for gender and clinical severity. No significant differences in response to treatment were found between the older adults and their younger counterparts, with the majorities of both groups being classified as treatment responders at posttreatment. This finding is particularly noteworthy given that the older patients reported having been symptomatic for more than twice as long as the younger adult cohort.

In a recent controlled case comparison, Carmin and Wiegartz described two older men with OCD. One experienced a successful outcome and the other experienced an unsuccessful outcome when intensive inpatient ERP was the treatment modality. These authors concluded that the duration of the illness, comorbidity of other psychiatric disorders as well as medical conditions, and the availability of social support can have an effect on treatment outcome.

Cognitive decline can exacerbate or mimic symptoms of OCD, and medical difficulties, such as cerebrovascular accidents (e.g., basal ganglia infarcts), are more prevalent in the elderly and have been noted to produce OCD symptoms in previously healthy patients. Such observations raise the question of whether individuals who have experienced neurological insults that result in OCD can benefit from psychological treatment. Of considerable importance is that ERP and pharmacological treatment were found to be effective in treating a 65-year-old man whose OCD was related to recent basal ganglia infarcts.

In comparison with the previously noted studies that used ERP, a follow-up study of medication and supportive psychotherapy that was offered to residents of an old-age home suggested that these methods can have a positive effect on treatment of anxiety disorders, including OCD and panic. One limitation of this study was that this was a diagnostically heterogeneous group, and no details were provided about what medications were used or what supportive psychotherapy entailed.

Studies focusing on CBT for late-life OCD have consisted of relatively small patient samples, thereby limiting generalizability. Preliminary findings, however, suggest that late-life OCD is treatable. Even if subsequent research finds lower success rates for the elderly than rates typically reported for the general adult population, there is sufficient evidence to conclude that at least some, if not most, older adults respond to ERP.

## **2.5. Assessment and Treatment of Posttraumatic Stress Disorder**

Much of the literature pertaining to posttraumatic stress disorder (PTSD) in the elderly focuses on holocaust and natural disaster victims and combat veterans. Given the vulnerability of older adults to physical violence, there is limited research that examines symptom presentations of elderly crime victims and treatment.

Typically, PTSD symptoms in older adults reflect a chronic waxing and waning of symptoms, with exacerbations linked to the expected stressors of advancing age. Despite early studies suggesting a level of resiliency in older adult disaster victims, this resiliency may reflect an underreporting of symptoms by PTSD sufferers, an underdiagnosis by clinicians, or an attribution of anxiety-related somatic symptoms to the normal frailties associated with old age by clinicians, thereby making the accurate diagnosis of this disorder difficult.

### 2.5.1. Assessment

There are several measures that have been used to assess PTSD symptoms in the elderly. The Clinician-Administered PTSD Scale (CAPS) is a clinician-administered semistructured interview that has been used extensively with older individuals who have typically been exposed to war-related trauma. In addition, the CAPS has been used as a process and outcome measure for those experiencing non-combat-related PTSD. These studies suggest that the CAPS is recommended for use with older PTSD sufferers.

Although a number of self-report measures have been used to assess PTSD in older adult samples, these studies typically provide descriptive rather than psychometric data. Although the Impact of Events Scale (IES) has provided psychometric data, the findings are equivocal with respect to its use with elders, suggesting that the type of traumatic event may be more important than symptomatology in older samples than in younger samples. The combat and civilian forms of the Mississippi PTSD Scales (MISS) have been significantly correlated with diagnostic measures of PTSD; however, the combat version appears to be the most accurate measure of PTSD severity in a small sample of elderly former prisoners of war.

### 2.5.2. Treatment

There are two reports of PTSD treatment in older adults. In 1998, Bonwick described a 16-week group day hospital treatment for veterans. The program included elements of psychoeducation, symptom management, relaxation, group therapy, and physical exercise. No outcome data were reported, but the author noted that those receiving treatment reported a greater understanding of PTSD, improved coping skills, and an enhanced quality of life. A recent conceptual review of PTSD in older adults indicates that a better understanding of the issues related to risk and vulnerability to trauma, such as the availability of social support networks, the use of coping strategies, and perceptions of the meaning related to the traumatic event, may allow for a better understanding of how to construct better treatment interventions.

## 2.6. Assessment and Treatment of Specific Phobias

### 2.6.1. Assessment

There is surprisingly little information with respect to the assessment and treatment of specific fears in older

adults given that specific phobias is the most prevalent anxiety disorder in this age group. Of the measures available that assess for the gamut of phobias, the Fear Survey Schedules (FSS-II and FSS-III) have been used with older adults and are promising screening measures for identifying specific fears. Kogan and Edelstein revised this measure specifically for use with older adults (FSS-OA). Their preliminary results are encouraging but not conclusive. One measure, the Falls Efficacy Scale (FES), appears to have good reliability and validity. However, additional psychometric evaluations (e.g., convergent and discriminatory validity) still need to be done before this self-report measure can be used independently of a comprehensive anxiety assessment battery in older persons with somatic symptoms of dizziness or balance disturbances.

### 2.6.2. Treatment

There is one randomized controlled study comparing CBT with an educational control group in more than 400 individuals (average age of 77 years) experiencing fears of falling. The CBT patients demonstrated posttreatment gains in mobility control and increased activity, but these gains were lost over a 6-month follow-up period. At 1 year follow-up, the CBT patients showed increased improvement in different areas (e.g., mobility range, social functioning). The authors of this study noted that only 63.4% of their patients attended more than five of the eight offered treatment sessions, thereby highlighting problems with compliance and attrition.

## 2.7. Issues Related to Anxiety Secondary to a Medical Condition

There is considerable overlap between many symptoms diagnostic of an anxiety disorder and symptoms that can be attributed to a medical illness. Medical illnesses such as cardiovascular disease, pulmonary dysfunction, stroke, hyperthyroidism, sensory impairments, and dementia can mimic, exacerbate, antedate, and/or accompany symptoms of anxiety. For example, symptoms of panic may overlap with certain symptoms related to angina, congestive heart failure, or emphysema, causing the diagnosis of panic to be overlooked. Alternatively, the normal developmental changes associated with aging can be mistaken for an anxiety disorder. Individuals who have sensory or mobility impairments may repeatedly check for where a hearing

aid or walker is located or ask for frequent reassurance, resulting in clinicians mistakenly suggesting a diagnosis of OCD. Clearly, differentiating between anxiety disorders and medical illness in older adults is a complicated task. There are no studies that address the treatment of comorbid anxiety and medical illness.

### 3. CONCLUSIONS

Enormous strides have been made in the area of the psychopathology, assessment, and treatment of anxiety disorders in adults. Unfortunately, in the area of geriatric anxiety, a tremendous amount of research still needs to be done.

There are significant questions as to whether the fundamental nature of anxiety disorders in adults is the same as that during their later years. A further complication arises in that neurobiological changes across the life span were found to cause an age-related decreased cortisol response to an experimental stressor. These findings suggest a decrease in reactivity to stress with advancing years. It has likewise been hypothesized that age-related changes in hypothalamic–pituitary–adrenal (HPA) activity, as indicated by cortisol level, are markers for central nervous system dysfunction. If this is indeed true, anxiety may be a link between central nervous system instability and the increases in cognitive impairment that are often found in aging and may explain the decreased prevalence of anxiety disorders in older adults.

As noted previously, epidemiological studies are confounded by their method of sampling. Cohort effects relevant to the stigma attached to mental illness have a greater influence on older adults. Thus, fewer elderly individuals may be willing to endorse anxiety symptoms in the course of an epidemiological study. Likewise, where samples are drawn from may influence prevalence data.

Finally, it would appear that both psychosocial and pharmacological treatments appear to aid in the reduction of anxiety symptoms in elderly samples. The prevailing form of psychotherapy that has been studied has been CBT. However, there is not firmly conclusive data that would allow one to unequivocally endorse CBT or a particular medication for use in treating a given anxiety disorder. Clearly, far more research is needed. One optimistic note is that the participants in existing longitudinal studies, such as the Harvard/Brown Anxiety Research Program (HARP), are aging. Data such as those generated by this study will allow

for the close examination of how anxiety disorders progress over adulthood and into later life and, hopefully, will provide answers to many of the questions that remain regarding geriatric anxiety.

### See Also the Following Articles

Depression in Late Life ■ Panic ■ Personality and Emotion in Late Life ■ Posttraumatic Disorders ■ Psychotherapy in Older Adults ■ Stress

### Further Reading

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Averill, P. M., & Beck, J. G. (2000). Posttraumatic stress disorder in older adults: A conceptual review. *Journal of Anxiety Disorders, 14*, 133–156.
- Ballenger, J. C., Davidson, J. R., Lecrubier, Y., Nutt, D. J., Borkovec, T. D., Rickels, K., Stein, D. J., & Wittchen, H. U. (2001). Consensus statement on generalized anxiety disorder from the International Consensus Group on Depression and Anxiety. *Journal of Clinical Psychiatry, 62*(Suppl. 11), 53–58.
- Barrowclough, C., King, P., Colville, J., Russell, E., Burns, A., & Tarrrier, N. (2001). A randomized trial of the effectiveness of cognitive–behavioral therapy and supportive counseling for anxiety symptoms in older adults. *Journal of Consulting and Clinical Psychology, 69*, 756–762.
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893–897.
- Beck, J. G., & Stanley, M. A. (1997). Anxiety disorders in the elderly: The emerging role of behavior therapy. *Behavior Therapy, 28*, 83–100.
- Beck, J. G., Stanley, M. A., & Zebb, B. J. (1996). Characteristics of generalized anxiety disorder in older adults: A descriptive study. *Behaviour Research and Therapy, 34*, 225–235.
- Bonwick, R. (1998). Group treatment programme for elderly war veterans with PTSD [letter]. *International Journal of Geriatric Psychiatry, 13*, 64–65.
- Brown, T. A., DiNardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of DSM-IV anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology, 110*, 49–58.
- Carmin, C. N., Pollard, C. A., & Gillock, K. L. (1999). Assessment of anxiety disorders in the elderly. In P. A. Lichtenberg (Ed.), *Handbook of assessment in clinical gerontology* (pp. 59–90). New York: John Wiley.
- Carmin, C. N., Pollard, C. A., & Ownby, R. L. (1998). Obsessive–compulsive disorder: Cognitive behavioral

- treatment of older versus younger adults. *The Clinical Gerontologist*, 19(2), 77–81.
- Carmin, C. N., & Wiegartz, P. S. (2000). Successful and unsuccessful treatment of obsessive-compulsive disorder in older adults. *Journal of Contemporary Psychotherapy*, 30, 181–193.
- Carmin, C. N., Wiegartz, P. S., & Scher, C. (2000). Anxiety disorders in the elderly. *Current Psychiatry Reports*, 2, 13–19.
- Carmin, C. N., Wiegartz, P. S., Yunus, U., & Gillock, K. L. (2002). Treatment of late-onset OCD following basal ganglia infarct. *Depression & Anxiety*, 15(2), 87–90.
- Deer, T. M., & Calamari, J. E. (1998). Panic symptomatology and anxiety sensitivity in older adults. *Journal of Behavior Therapy and Experimental Psychiatry*, 29, 303–316.
- Devanand, D. P. (2002). Comorbid psychiatric disorders in late life depression. *Biological Psychiatry*, 52, 236–242.
- Fedoroff, I. C., & Taylor, S. (2001). Psychological and pharmacological treatments of social phobia: A meta-analysis. *Journal of Clinical Psychopharmacology*, 21, 311–324.
- First, M., Spitzer, R., Gibbon, M., et al. (1995). *The Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I, version 2.0, final version)*. New York: New York State Psychiatric Institute.
- Flint, A. (1994). Epidemiology and comorbidity of anxiety disorders in the elderly. *American Journal of Psychiatry*, 151, 640–649.
- Flint, A. J., Koszycki, D., Vaccarino, F. J., Cadieux, A., Boulenger, J.-P., & Bradwejn, J. (1998). Effect of aging on cholecystokinin-induced panic. *American Journal of Psychiatry*, 155, 283–285.
- Forsell, Y., & Winblad, B. (1998). Feelings of anxiety and associated variables in a very elderly population. *International Journal of Geriatric Psychiatry*, 13, 454–458.
- Gray, J. J., & Acerno, R. (2002). Symptom presentations of older adult crime victims: Description of a clinical sample. *Journal of Anxiety Disorders*, 16, 299–309.
- Kabacoff, R. I., Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of Anxiety Disorders*, 11, 33–47.
- King, P., & Barrowclough, C. (1991). A clinical pilot study of cognitive-behavioral therapy for anxiety disorders in the elderly. *Behavioural Psychotherapy*, 19, 337–345.
- Kogan, J. N., & Edelstein, B. A. (in press). Modification and psychometric examination of a self-report measure of fear in older adults. *Journal of Anxiety Disorders*.
- Kogan, J. N., Edelstein, B. A., & McKee, D. R. (2000). Assessment of anxiety in older adults: Current status. *Journal of Anxiety Disorders*, 14, 109–132.
- Krasucki, C., Ryan, P., Ertan, T., Howard, R., Lindsay, J., & Mann, A. (1999). The FEAR: A rapid screening instrument for generalized anxiety in elderly primary care attenders. *International Journal of Geriatric Psychiatry*, 14, 60–68.
- Lindsay, J. (1996). Elderly people and crime. *Reviews in Clinical Gerontology*, 6, 199–204.
- Manela, M., Katona, C., & Livingston, G. (1996). How common are the anxiety disorders in old age? *International Journal of Geriatric Psychiatry*, 11, 65–70.
- McCarthy, P. R., Katz, I. R., & Foa, E. B. (1991). Cognitive-behavioral treatment of anxiety in the elderly: A proposed model. In C. Salzman, & B. D. Lebowitz (Eds.), *Anxiety in the elderly: Treatment and research* (pp. 197–214). New York: Springer.
- Meeks, S., Woodruff-Borden, J., & Depp, C. A. (2003). Structural differentiation of self-reported depression and anxiety in late life. *Journal of Anxiety Disorders*, 17, 627–646.
- Mohlman, J., Gorenstein, E. E., Kleber, M., deJesus, M. B., Gorman, J. M., & Papp, L. A. (2003). Standard and enhanced cognitive behavior therapy for late life generalized anxiety disorder: Two pilot investigations. *American Journal of Geriatric Psychiatry*, 11, 24–32.
- Mohlman, J., deJesus, M., Gorenstein, E. E., Kleber, M., Gorman, J. M., & Papp, L. A. (in press). Distinguishing generalized anxiety disorder, panic disorder, and mixed anxiety states in older treatment-seeking adults. *Journal of Anxiety Disorders*.
- Mohlman, J., & Zinbarg, R. E. (2000). The structure and correlates of anxiety sensitivity in older adults. *Psychological Assessment*, 12, 440–446.
- Ownby, R. L., Harwood, G. D., Barker, W. W., & Duara, R. (2000). Predictors of anxiety in patients with Alzheimer's disease. *Depression and Anxiety*, 11, 38–42.
- Radley, M., Redston, C., Bates, F., & Pontefract, M. (1997). Effectiveness of group anxiety management with elderly clients of a community psychogeriatric team. *International Journal of Geriatric Psychiatry*, 12, 79–84.
- Rathus, J. H., & Sanderson, W. C. (1996). Cognitive behavioral treatment of panic disorder in elderly adults: Two case studies. *Journal of Cognitive Psychotherapy*, 10, 271–280.
- Schaub, R. T., & Linden, M. (2000). Anxiety and anxiety disorders in the old and very old: Results from the Berlin Aging Study (BASE). *Comprehensive Psychiatry*, 41(2), 48–54.
- Segal, D. L., Hersen, M., VanHasselt, V. B., Kabacoff, R. L., & Roth, L. (1993). Reliability of diagnosis in older psychiatric patients using the Structured Clinical Interview for DSM-III-R. *Journal of Psychopathology and Behavioral Assessment*, 15, 347–356.
- Sheikh, J. I., & Cassidy, E. L. (2000). Treatment of anxiety disorders in the elderly: Issues and strategies. *Journal of Anxiety Disorders*, 14, 173–190.
- Snyder, A. G., Stanley, M. A., Novy, D. M., Averill, P. M., & Beck, J. G. (2000). Measures of depression in older adults with generalized anxiety disorder: A psychometric evaluation. *Depression and Anxiety*, 11, 114–120.
- Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). *State-Trait Anxiety Inventory: Test manual*. Palo Alto, CA: Consulting Psychologists Press.

- Stanley, M. A., & Beck, J. G. (2000). Anxiety disorders. *Clinical Psychology Review, 20*, 731–754.
- Stanley, M. A., Beck, J. G., Barron, M., & Wagner, A. L. (1994, November). *Treatment of generalized anxiety disorder in older adults: Examination of outcome predictors*. Paper presented at the annual meeting of the Association for the Advancement of Behavior Therapy, San Diego.
- Stanley, M. A., Beck, J. G., & Glassco, J. D. (1996). Treatment of generalized anxiety in older adults: A preliminary comparison of cognitive-behavioral and supportive approaches. *Behavior Therapy, 27*, 565–581.
- Stanley, M. A., Beck, J. G., Novy, D. M., Averill, P. M., Swann, A. C., Diefenbach, G. J., & Hopko, D. (2003). Cognitive behavioral treatment of late-life generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 71*, 309–319.
- Stanley, M. A., Beck, J. G., & Zebb, B. J. (1996). Psychometric properties of four anxiety measures in older adults. *Behaviour Research and Therapy, 34*, 827–838.
- Stanley, M. A., Novy, D. M., Bourland, S. L., Beck, J. G., & Averill, P. M. (2001). Assessing older adults with generalized anxiety: A replication and extension. *Behaviour Research and Therapy, 39*, 221–235.
- Swales, P., Solfrin, J., & Sheikh, J. (1996). Cognitive behaviour therapy in older panic disorder patients. *American Journal of Geriatric Psychiatry, 4*, 46–60.
- Swartz, M., Martin, T., Martin, M., Eizur, A., & Barak, Y. (1999). Outcome of a psychogeriatric intervention in an old-age home: A 3 year follow-up study. *Annals of Clinical Psychiatry, 11*, 109–112.
- Van Balkom, A. J., Beekman, A. T., De Beurs, E., Deeg, D. J., Van Dyck, R., & Van Tilburg, W. (2000). Comorbidity of the anxiety disorders in a community-based older population in The Netherlands. *Acta Psychiatrica Scandinavica, 101*, 37–45.
- Wetherell, J. L., Gatz, M., & Craske, M. G. (2003). Treatment of generalized anxiety disorder in older adults. *Journal of Consulting and Clinical Psychology, 71*, 31–40.



# Arousal in Sport

Leonard D. Zaichkowsky and Adam H. Naylor

*Boston University, Boston, Massachusetts, USA*

1. Introduction
  2. Conceptual Confusion Surrounding Arousal
  3. Neuropsychology of Arousal
  4. Arousal Theories
  5. Effects of Arousal on Athletic Performance
  6. Managing Arousal Levels
- Further Reading

## GLOSSARY

**anxiety** A negative emotional state generally expressed as worry and apprehension and accompanied by heightened arousal.

**arousal** The body's degree of activation from low to high; it is reflected in hormone levels and other physiological reactivity that is mediated by thoughts and feelings.

**stress** The body's reaction to perceived threat by an individual; the reaction is typically increased anxiety and arousal.

Arousal in sport and exercise is a human condition that ranges along a continuum from sleep to high excitation and is expressed physiologically, cognitively, and behaviorally. Three prominent theories of arousal in sport are drive theory, the inverted-U hypothesis, and reversal theory. Each has unique insights for the discussion of an athlete's ability to perform at optimal levels on the playing field or court and can guide an astute coach, counselor, or psychologist in his or her work. Lastly, the application of cognitive, behavioral,

and cognitive-behavioral interventions for arousal regulation are discussed.

## 1. INTRODUCTION

The coach gathers the team moments before the championship game for a pep talk laden with fire, brimstone, and proclamations of victory. The professional basketball player sits on the training room table, tattoos lining his arms, bobbing his head rhythmically to the music on his headphones. The sprinter takes a few quick strides, bounds up and down a couple of times, and takes two good deep breaths before assuming her position in the starting blocks. The stock car racer says a short prayer and finds a brief moment of silence before sliding into his car and heading out to the starting line. Excellent athletic accomplishments require the athlete's body to be appropriately energized, with physiological and psychological resources prepared for the stresses and physical demands of competition.

The preceding instances describe the concept of arousal and its regulation in sport. The term "arousal" has been a part of the English language for more than 400 years. The *Oxford English Dictionary* defines "arouse" as "to raise up from sleep or inactivity, to excite." James, in his classical work *Principles of Psychology*, used the term "arousal" to describe "activation." Arousal, as it is used in sport, describes a condition of controlled or uncontrolled preparedness and

manifests itself in varying levels of mental activity, emotional display, physical energy, and concomitant physiological reactivity.

## 2. CONCEPTUAL CONFUSION SURROUNDING AROUSAL

Besides the historical reference of exciting one to action, being aroused typically refers to physiological responses. That is, aroused individuals demonstrate high heart rates, increased sweat response, increased brain activity, and (typically) increased intensity of physical effort.

A perusal of the general psychology and sport psychology research over the past five decades shows that arousal is often used interchangeably with terms such as anxiety, activation, emotion, motivation, and psychic energy. This understandably has resulted in conceptual confusion. The early researchers of arousal, such as Cannon, Duffy, Hebb, and Malmo, treated arousal as a unidimensional physiological construct. Today, arousal is viewed as a multidimensional construct that includes a physiological dimension paired/grouped with cognitive, affective, and/or behavioral dimensions. For example, an ice hockey player who is optimally aroused would have an elevated heart rate, increased beta waves, increased respiration, and increased adrenalin. In addition, thought processes would be alert and focused, such that the player is able to read and react to on-ice situations quickly and accurately. The optimally aroused athlete would also be in control of essential affective states and would demonstrate “passion” or intensity in play.

The physiological component of arousal can be represented in a number of ways such as through muscle tension, cortical activity, electrodermal activity, respiration, and biochemical markers (e.g., epinephrine, cortisol). The cognitive interpretation appraisal component can influence the physiological component. Furthermore, the affective component can also interact

with thoughts to influence physiological responses and athletic performance (Fig. 1).

## 3. NEUROPSYCHOLOGY OF AROUSAL

During recent years, much has been learned about the neurological and biochemical mechanisms involved in human arousal. Preparation for competition, as well as competition itself, sets off a chain of events in the central nervous system and autonomic nervous system that results in arousal. Table I provides a summary of nervous system structures involved in arousal as well as the locations and functions of the structures.

## 4. AROUSAL THEORIES

A number of theories have been proposed to explain the relationship between arousal and athletic performance. Following is an analysis of the major perspectives.

### 4.1. Drive Theory

The drive theory of arousal states that arousal has a positive correlation with athletic performance, particularly in cases of skills that are well learned (Fig. 2). The theory is based on the classical work of Hull, who stated that drive (or arousal) multiplied by habit strength (amount of prior learning) determines an individual's level of performance:

$$\text{Performance} = \text{Drive} \times \text{Habit Strength.}$$

As emotions and excitement increase, so does an athlete's performance.

Mediating the role of anxiety on an athlete's performance is habit strength. Habit strength speaks to how well an individual has learned particular skills. The better learned a skill (or the more skilled the athlete), the more likely arousal is to enhance athletic performance. Some researchers have even gone so far as to suggest that the play of novice athletes (low in habit strength) is inhibited by increasing arousal levels.

Although this theory has many merits and has been closely associated with the concept of social facilitation, some have questioned whether the arousal-performance relationship depicted exists in all motor activities required in sport. The hypothesis is

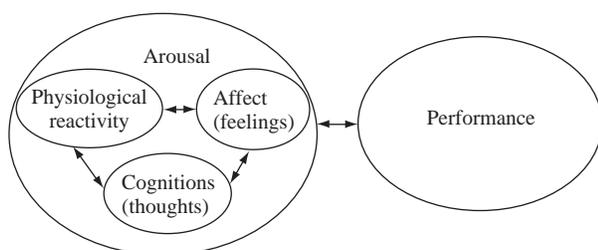
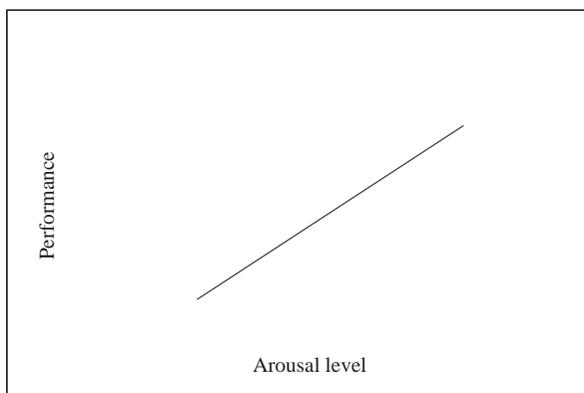


FIGURE 1 Arousal and performance.

**TABLE I**  
**Nervous System Structures Involved in Arousal**

<i>Structure</i>	<i>Location</i>	<i>Function</i>
Reticular formation	Part of brain stem that continues out of medulla and pons	Involved in sleep, waking, alertness, and optimal brain arousal
Hypothalamus	Between the thalamus and midbrain	Involved in appetitive and sexual behavior, regulation of sympathetic nervous system and pituitary gland
Limbic system	Series of structures, including the hippocampus and amygdala, located near the border between the cerebral hemisphere and the brain stem; the hypothalamus is sometimes considered part of the limbic system	Primarily concerned with emotional behavior; amygdala appears to be involved in aggression
Cerebral cortex	Convolutated outer layer of the human brain	Highest center involved in learning, remembering, planning, and performing motor acts
Sympathetic nervous system	A branch of the autonomic nervous system with nerve fibers originating in thoracic and lumbar regions of the spinal cord	Activates glands and smooth muscles during arousal
Parasympathetic nervous system	A branch of the autonomic nervous system with nerve fibers originating in the brainstem and sacral regions of the spinal cord	Maintains appropriate internal states during times of relaxation
Adrenal cortex	Outer layer of two small endocrine glands just above the kidneys	Secretes cortisol that regulates metabolism and stress response
Adrenal medulla	Inner layer of two small endocrine glands just above the kidneys	Secretes adrenaline (epinephrine) and noradrenalin (norepinephrine); both are involved in increased activation of the body
Pituitary gland	Deep inside brain, just below hypothalamus	Releases adrenocorticotrophic hormone that stimulates adrenal cortex to release cortisol



**FIGURE 2** Drive theory.

well accepted for gross motor tasks requiring strength and speed, but it is less well accepted for tasks requiring balance, accuracy, and fine motor skills (e.g., putting a golf ball).

Drive theory is most beneficial to practitioners in understanding the link between skill level and optimal arousal. It is unlikely that an excellent athlete facing an easy task will garner sufficient emotional and physiological motivation to perform at his or her best. In such a situation, it benefits that athlete to set challenging process goals for the task at hand, in essence increasing the perceived challenge of the task. Conversely, regardless of the potential benefits of arousal to explosive movements and strength, a novice athlete will not reap great rewards from achieving exceptional arousal levels. Such an athlete should be counseled to find a state of calm that allows him or her to feel comfortable on the playing field or court.

#### **4.2. Inverted-U Hypothesis**

The inverted-U hypothesis states that moderate levels of arousal are ideal for optimal athletic performance

(Fig. 3). Performance gradually improves as a competitor goes from underarousal to the alert state of moderate arousal, after which point performance declines as the individual becomes overaroused.

Unlike drive theory, the inverted-U hypothesis does state that there is a point of arousal that exceeds what is needed for excellence on the playing field. Yet since the original model posed in 1908 by Yerkes and Dodson, it has been suggested that one must account for task difficulty, unique sporting demands, ability level of athletes, and individual differences when employing the inverted-U hypothesis. For example, it has been suggested that the peak of a golfer's inverted U ought to be at relatively lower levels of arousal, whereas a sprinter's inverted U will exist at high levels of arousal. In both instances, too little arousal or too much arousal is debilitating, yet the midranges of arousal differ due to sporting demands. Therefore, it is suggested that although the arousal-performance relationship is that of an inverted U, the placement of this inverted U on the arousal continuum ought to vary due to sport demands and individual differences.

The inverted-U hypothesis is perhaps most readily embraced by those in sport psychology practice. This is due in part to the fact that its simplicity (a unidimensional model as opposed to a theory that considers arousal and anxiety or multiple forms of anxiety) makes it easily understood by athletes and coaches. An athlete's ability to move into and out of ideal performance states is clear by drawing an upside down U on a piece of paper. This theory is also appreciated by practitioners because it recognizes the possibility of overarousal. This further illuminates the challenges posed by a "one size fits all" fire and brimstone pep talk.

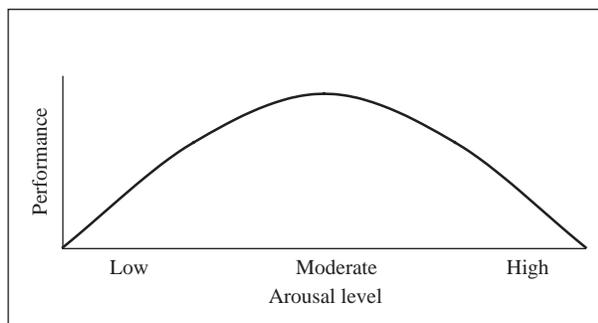


FIGURE 3 Inverted-U hypothesis.

### 4.3. Reversal Theory

Unlike other theories, the foundation of reversal theory is an athlete's interpretation of arousal rather than how high or low the energizing state is. Individual differences in perception determine how facilitative or debilitating arousal levels are. Depending on the situation that the athlete faces, arousal levels might be interpreted as "anxiety" or "excitement." For example, high levels of arousal might be interpreted by one athlete as lack of preparedness, whereas the same elevated arousal could be perceived as the appropriate energy for athletic success by another. Furthermore, at the root of this theory is the ability to change (i.e., reverse) one's interpretation of arousal. The reversal of interpretation might lie in improved skill levels (through practice), realization that tasks are not as challenging as thought previously, or support of coaches or teammates. Ultimately, according to this model, optimal performance is achieved when a player's preferred arousal level is in harmony with his or her actual arousal levels.

Sport psychology coaches who understand the reversal theory often see the importance of cognitive restructuring. The fact that perceptions of threats and opportunities on the playing field can change the facilitative nature of arousal makes an athlete's beliefs and thought processes essential for success. Regardless of the situation, an athlete nearly always has an opportunity to find his or her ideal arousal level. For example, a loud and hostile crowd could be perceived as an "extra player" for the home team, reducing the visiting team members' good feelings and confidence (i.e., debilitating perceptions). Conversely, the visiting team members could see and hear the hostile crowd and think, "We must be good! This crowd is worried and needs to help their team" (i.e., a facilitative perception). Drawing from the teachings of cognitive and/or rational-emotive therapies can be very useful to a practitioner who appreciates the reversible nature of arousal's effects on athletic performance.

### 4.4. Newer Theories

Other theories related to arousal in sport performance are the Individual Zones of Optimal Functioning (IZOF), the multidimensional anxiety theory, and the Cusp Catastrophe Model. Hanin suggested that in the IZOF, each individual has a certain level of emotion where he or she competes optimally. This "zone" of emotion can be high, low, or somewhere in between, depending on each person's unique

cognitive, physiological, and technical resources. The multidimensional anxiety theory suggests that cognitive anxiety (i.e., thoughts/worries) has an inverse linear relationship with performance, whereas somatic anxiety (i.e., physiological) displays an inverted-U relationship with performance. Lastly, the Cusp Catastrophe Model states that one must consider cognitive anxiety, physiological arousal, and performance demands simultaneously. In this theory, the relationship between anxiety/arousal and performance is not a neat linear one but rather one that can be filled with great peaks and valleys. In particular, one can achieve the greatest performances under high cognitive stress and arousal but, at the same time, can experience the greatest performance decline when things go awry.

## 5. EFFECTS OF AROUSAL ON ATHLETIC PERFORMANCE

Arousal levels directly influence physiological, cognitive, and motor modalities necessary for excellence on the playing field (Table II). Understanding these modalities

can help in the identification and management of underarousal and overarousal. Physiologically, modalities measurable through psychophysiological monitoring, such as brain waves, skin conductance, heart rate, blood pressure, and muscle tension, all are elevated when an individual is aroused. Although these modalities are directly related to somatic arousal, each must be carefully measured because most are automatically increased during physical exertion. Perhaps the speed of recovery to "normal" levels between points and after play is when the goodness of one's arousal level is most apparent.

Arousal level also influences one's cognitive functioning. It is suggested that an athlete beyond his or her optimal arousal level is likely to suffer from poor decision making, impaired memory, and rushed thinking. Nideffer posited that in sport, heightened arousal levels also impair an athlete's ability to focus on the task at hand. Contrary to the lay assumption that overaroused athletes think too much, it is suggested that being outside one's optimal arousal zone leads to thinking about the wrong things at the wrong time. In particular, Nideffer suggested that one begins to concentrate on thoughts, emotions, and feelings (specifically fear and worry). There is little to suggest that underarousal significantly influences one's cognitive

TABLE II  
Effects of Arousal on Athletic Performance

Modality	Low arousal	High arousal
Brain wave (electroencephalography)	Increased alpha waves or theta Good performance on fine motor skills	Increased beta waves Good performance on gross motor skills
Skin conductance	Lack of perspiration Good for a sport like golf	Increased perspiration, increased conductance Good for high-energy sports
Heart rate	Decreased heart rate Ability to compete over long periods of time	Increased heart rate Ability to perform skills with speed and power
Blood pressure	Decreased blood pressure Effects uncertain	Increased blood pressure Effects uncertain
Muscle tension (electromyography)	Muscles relaxed Fluid technique, potential poor form	Increased muscle tension Good strength and stability, potential for rigid movements
Focus	May be inattentive	Attention narrows, ultimately leading to focus on fears and worries
Memory	No deficit	Difficulty in remembering large chunks of information and complex tasks
Anaerobic power	Impaired strength and speed	Increased strength and speed
Fine motor control	Improved dexterity	Impaired

abilities, yet at low arousal levels few athletes are able to garner the necessary energy and alertness to succeed at the physical tasks of sport.

## 6. MANAGING AROUSAL LEVELS

There are many psychological techniques that allow one to effectively manage arousal levels. Understanding that arousal, anxiety, concentration, confidence, and physiological modalities all are inherently linked, it is clear that there are many approaches to help athletes find and maintain facilitative levels of arousal. Interventions can be broken down into three categories: cognitive, behavioral, and cognitive-behavioral. (Although dividing interventions into these categories is useful in understanding the practice of sport psychology, it is important to recognize that the mind and body are inherently linked. This reminds one that these categories are not mutually exclusive.)

### 6.1. Cognitive Interventions

Increasing an athlete's awareness about his or her optimal arousal levels is the foundation of an intervention. Developing awareness might be achieved through successive postperformance reflections that examine levels of excitement, concentration, emotions, and confidence during a competition. Similarly, a retrospective comparison of "best performances" versus "worst performances" can paint a powerful picture of an athlete's optimal arousal state.

Setting realistic and motivating goals gives an athlete the ability to energize while maintaining appropriate focus. It is appropriate to have a "mission" or goal for one's career and season as well as for each practice and competition. Goals direct an athlete (i.e., energizing) and clarify purpose (i.e., minimizing stress and relaxing).

Cognitive restructuring and self-talk can appropriately energize an athlete and calm him or her down when necessary. Cognitive restructuring fits nicely with the reversal theory of arousal, suggesting that modifying one's perspective and thoughts will help the athlete to find appropriate activation levels. Self-talk or the use of cue words can energize (i.e., "quick feet") or calm down (i.e., "slow and smooth") an athlete when necessary. Thought stoppage allows an athlete to catch himself or herself when he or she begins to focus on task-irrelevant cues.

Mental imagery, another cognitive intervention, is popular in the field of sport psychology. A good

imagery session in which the body is relaxed and all of the senses are incorporated can be used to relax or energize an individual. Accessing one's memory bank of sporting highlights can serve as very motivational imagery, hence increasing arousal levels. Conversely, imagery can be used to escape from the stresses of competition, consequently reducing arousal.

### 6.2. Behavioral Interventions

There are a variety of relaxation techniques that can benefit an athlete who is overaroused. The following are a few techniques that are used regularly by sport psychology practitioners: controlled breathing, progressive muscle relaxation, meditation, relaxation response, and biofeedback. A few of these interventions can also be used by an athlete to energize, specifically quick shallow breathing and gaining understanding of excited physiological levels through biofeedback. Actions such as quick sprints, jumping rope, and dynamic stretching also serve to energize the underaroused athlete.

### 6.3. Cognitive-Behavioral Interventions

Both cognitive and behavioral techniques are combined during preperformance preparation and competitive routines. Prior to a competition, successful athletes commonly engage in activities such as setting goals, listening to music, using mental imagery, and physically warming up so as to have optimal arousal at the start of the competition. Similarly, while in the midst of play, athletes often use competitive routines to relax and refocus. Using a physical cue to put a prior play in the past, engaging in mental planning, and taking a good breath or two to help focus attention are often elements of such a routine. Routines are commonly used in sports where the flow of play is intermittent such as tennis, baseball, and golf.

### See Also the Following Articles

Anxiety and Optimal Athletic Performance ■ Stress

### Further Reading

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review*, 96, 84-100.

Hanin, Y. (2000). *Emotions in sport*. Champaign, IL: Human Kinetics.

- Kerr, J. H. (1985). The experience of arousal: A new basis for studying arousal effects in sport. *Journal of Sports Sciences*, 3, 169–179.
- Nideffer, R. M. (1989). *Attention control training for sport*. Los Gatos, CA: Enhanced Performance Services.
- Spence, J. T., & Spence, K. W. (1956). The motivational components of manifest anxiety: Drive and drive stimuli. In C. D. Spielberger (Ed.), *Anxiety and behavior* (pp. 291–326). New York: Academic Press.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology of Psychology*, 18, 459–482.
- Zaichkowsky, L., & Baltzell, A. (2001). Arousal and performance. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (pp. 319–339). New York: John Wiley.



# Assessment and Evaluation, Overview

Rocío Fernández-Ballesteros

Autonoma University of Madrid, Madrid, Spain

1. Introduction
  2. Theoretical Perspectives
  3. Methods and Techniques
  4. The Process of Assessment, Treatment, and Evaluation
  5. Conclusions
- Further Reading

## GLOSSARY

**assessment contexts** The fields in which psychological assessment is applied.

**assessment process** Series of tasks ordered in a specific sequence conducted by the assessor for solving the client/subject demand.

**assessment targets** Specific psychological characteristics (intelligence, cognitive skills, personality dimensions, behavioral responses, defense mechanisms, environmental conditions, etc.).

**interview** The interaction between two people in which one person (the interviewer) asks for and collects information with a given purpose or demand, and the other person (the interviewee, client, or subject) responds or answers questions.

**observational method** A set of procedures for collecting data, including a protocol with a pre-defined set of behavioral categories in which the subject's behavior is coded or registered.

**projective methods** Instruments containing relatively structure-free stimuli, considered especially sensitive to covert or unconscious aspects of behavior, that encourage a wide variety of idiographic subject responses with a

minimum of subject awareness concerning the purpose of the test.

**psychological assessment** The discipline of scientific psychology devoted to the study of a given human subject (or group of subjects in a scientific applied field—clinical, educational, work, etc.), by means of scientific tools (tests and other measurement instruments), with the purpose of answering clients' demands that require scientific operations such as describing, diagnosing, predicting, explaining or changing the behavior of that subject.

**psychological testing** The process of administering, scoring, and evaluating tests or standard methods.

**psychometrics** The principles, theories, and techniques of test construction.

**self-report** A method for collecting data whose source is the subject's verbal message about him- or herself.

**standard methods or tests** Tests that have a standard set of stimuli, standard administration procedures, and a standard form of measuring subjects' responses, which can be transformed into standard scores.

## 1. INTRODUCTION

### 1.1. Concept

Psychological assessment is the discipline of scientific psychology devoted to the study of a given human subject (or group of subjects) in a scientific applied field—clinical, educational, work, etc., by means of scientific tools (tests and other measurement instruments), with

the purpose of answering client's demands that require scientific operations such as describing, diagnosing, predicting, explaining or changing the behavior of that subject. —Fernández-Ballesteros *et al.*, 2001.

Throughout the history of applied psychology, psychological assessment has been a key discipline, often to the extent of being confused with psychology itself and of constituting the public face of psychology. As several empirical studies have pointed out, assessment tasks are an ever-present element in all psychological work, from the scientific laboratory to applied settings. Thus, all types of psychological work involve assessment at some stage and to some extent, whether it be through the use of sophisticated data collection equipment, structured interviews, or qualitative observations, or by means of complex measures obtained through the administration of tests.

### 1.2. Brief History

Psychological assessment in some form or another can be found throughout the history of human thinking, in the Bible, in ancient China, and even in the works of Socrates, Plato, and Aristotle or those of the ancient physicians Hippocrates and Galen. Nevertheless, it is not until the late 19th century that we find scientific events related to psychological assessment. [Table I](#) offers an overview of the most important contributions to psychological assessment from the starting point of scientific psychology—the founding of Wundt's Psychology Laboratory in Leipzig in 1878—until 1950.

As can be seen in [Table I](#), psychological assessment is a sub-discipline developed at the same time as experimental psychology. While psychology is devoted to the study of the principles of mind, consciousness, and behavior, psychological assessment is devoted to the study of how, or to what extent, these principles are fulfilled in a specific subject or group of subjects. For the study of a particular subject (or group of subjects), psychological assessment requires methods for data collection. Thus, psychological assessment is closely related to psychometrics and to psychological testing. Methodological advances in psychometric theory are important because all measurement devices are developed and evaluated using psychometric criteria; psychological testing and psychological assessment have even been confused. But psychological assessment cannot be reduced to either psychometrics or the processes of administering tests. Rather, as referred to in the above definition, psychological assessment always

involves a process of problem-solving and decision-making in which measurement devices and other procedures for collecting data are administered for testing hypotheses about a subject (or group of subjects), with the specific purpose of answering relevant questions asked by the client or the subject. The client and subject may be the same person or unit, or the client may be another professional or relative (e.g., a psychiatrist, an educator or a judge) who asks for the assessment of a third person or group of persons with a given purpose (description, classification, diagnosis, prediction, prognosis, selection, change, treatment, or evaluation).

### 1.3. Objectives and Contexts

But these applied questions require scientific objectives, of which the following are the most common: description, classification, prediction, explanation, control, and evaluation. By way of example: when a psychiatrist asks a psychologist for a diagnosis of a given patient, the psychologist should proceed to the description and classification of the patient; when a client asks a psychologist who is the best candidate for a job, the psychologist must predict the candidate who is likely to perform best; when a couple or a family asks for counseling, the psychologist should analyze the situation before intervening; when a pupil has a learning problem, the psychologist must analyze the case and hypothesize about the circumstances that explain or cause the problem, manipulate its cause(s), and evaluate the extent to which it has been solved.

As can be seen from the examples, these various objectives emerge from different applied settings, the three most common assessment contexts being clinical psychology, educational psychology, and work and industrial psychology.

Finally, the relevant events listed in [Table I](#) emerge from different models of scientific psychology, are related to specific psychological content or targets (intelligence, aptitudes, personality characteristics, etc.), and imply a variety of methods (interview, self-reports, projective techniques, etc.).

## 2. THEORETICAL PERSPECTIVES

Like psychology in general, psychological assessment developed in association with different models or theoretical frameworks, which have given rise to different theoretical formulations, targets, methods, and objectives. [Table II](#) shows the psychological models that have had

**TABLE I**  
**Principal Contributions in the History of Psychological Assessment from 1880 to 1950**

Year	Author(s)	Contribution
1884	Francis Galton (1822–1911)	At London's South Kensington Museum, British scientist Galton set up the "Anthropometric Laboratory" for the assessment of physical and psychological characteristics (reaction time, visual acuity, discriminative speed and other sensory, perceptive, and motor measures).
1890	McKeen Cattell (1860–1944)	Working on his doctoral dissertation at Wundt's laboratory in Leipzig, the American Cattell developed the first definition of "mental test" as the measurement tool for higher mental processes.
1894	Emil Kraepelin (1856–1926)	One of Wundt's first pupils, Kraepelin proposed a comprehensive system for classifying psychopathology, taking into consideration mental states, memory, etc.
1896	Hermann Ebbinghaus (1850–1909)	Ebbinghaus defined intelligence as the ability to interrelate and combine correctly. He developed the first group intelligence test, the Completion test, consisting of a series of texts with words and parts of words omitted. Subjects (in a group) had to fill in as many blanks as possible in 5 minutes.
1896	Karl Pearson (1857–1936)	The Briton Pearson developed the mathematical method of correlation, a fundamental tool for the development of tests.
1898	E. L. Thorndike (1874–1949)	Thorndike reported the use of mazes and puzzles to measure the intelligence of cats. In 1904, he wrote <i>An Introduction to the Theory of Mental and Social Measurements</i> .
1904	Charles Spearman (1863–1945)	Two important developments can be attributed to this British scientist (who worked for several years in Germany): the development of reliability ( <i>The Proof and Measurement of Association between Two Things</i> ), and a theory of intelligence that led to the development of methods for locating a general factor underlying a group of tests.
1905	Alfred Binet (1857–1911)	Working on intelligence in an attempt to select students for compulsory education in France, Binet and his colleague Simon developed a set of situations that constituted the first test of intelligence.
1906	Karl Jung (1875–1961)	Jung introduced the first word association method for diagnosing psychopathology, using a list of 100 words, to each of which the subject was asked to respond as quickly as possible with the first word that came to mind.
1914	Robert S. Woodworth (1869–1962)	The first personality questionnaire was developed by Woodworth, intended as a screening process to identify those unfit for active service in World War I. The personal data sheet consisted of a set of self-reported items selected from psychiatric interviews.
1917	Arthur S. Otis (1886–1964) and Robert M. Yerkes (1876–1956)	Developed the "Army Alpha and Beta," general intelligence tests for the assessment and selection of American recruits for service in World War I.
1920	Kurt Goldstein (1878–1965)	Scientific neuropsychological assessment emerged in Germany after World War I from research on the concomitants of brain damage. After concluding that individuals with critical injuries display loss of ability for abstract thinking and that their behavior involves less generalization, Goldstein (in collaboration with Scheerer) developed several tests of abstract and concrete thinking.

*Continues*

Continued

Year	Author(s)	Contribution
1921	Hermann Rorschach (1884–1922)	The first projective assessment methods were developed by the Swiss Rorschach and published in his <i>Psychodiagnostik</i> . The method used 10 inkblots, to which subjects have to respond by saying what they “could be, or looked like.” This method was designed to assess personality and psychopathology.
1934	Lev Vygotsky (1896–1934)	From the perspective of “Soviet psychology,” the Russian Vygostky, in his <i>Thinking and Language</i> , developed the concept of the assessment of proximal development, which constitutes the roots of cognitive strategies of assessment.
1938	Henry A. Murray (1893–1988)	In the book <i>Exploration in Personality</i> , Murray (with Morgan) presented the thematic apperception test. Consisting of a set of ambiguous pictures with human figures, this test requires subjects to make up a story. The main objective of this method is to assess personality.
1943	Office of Strategic Services (United States)	Using methods first employed in Germany, but later developed in Britain, an important program for the global assessment of individuals, “Assessment of Man,” was consolidated. As Dubois (1970) stated, “Instead of attempting to measure rather limited aspects of behavior as had been the case with the most psychometric and clinical devices up to that time, a program was designed to describe the way the individual was able to act in a wide variety of situations” (p. 110).
1950	Monty B. Shapiro (1951–)	At a meeting of the British Society of Psychology, Shapiro presented an important work developed in the Maudsley Hospital, “Experimental Approach to Diagnosis and Psychological Testing,” which would serve as the basis for behavioral assessment and other idiographic developments.

most repercussion in psychological assessment: trait, medical, dynamic, behavioral, cognitive, and constructivist.

### 2.1. Trait

As human beings display wide variability in the way they think, feel, and behave, one of the objects of inquiry in psychology has been individual differences in human characteristics related to intelligence or mental aptitudes, affect and emotions (aggressiveness, hostility, etc.), personality (extraversion, flexibility, etc.), psychopathology (depression, anxiety, etc.). All of these have constituted important targets for assessment.

The trait approach uses several data collection techniques, from questionnaires or standardized tasks to rating scales. It is, therefore, the trait approach that has inspired the most well known psychometric tests. The main objective of this approach is to find the true score of a given subject in a specific trait or his or her true position in relation to a normative or standard group. In practical terms, these comparisons allow psychologists to describe and classify subjects with regard to a set of

traits or medical conditions and to predict their future behaviors. The explanation objective would not be applicable to the trait model, for reasons of tautology.

### 2.2. Medical

Historically, the development of psychological assessment has been closely related to clinical settings. Therefore, from the outset a medical model emerged, focusing principally on mental disorders. Several classification systems have been developed, the most well known being the American Psychiatric Association’s *Diagnostic and Statistical Manual of Mental Disorders* (currently the *DSM IV-R*) and the World Health Organization’s *Classification of Mental and Behavioral Disorders* (currently the *ICD-10*), both of which constitute descriptive category systems with no etiological bases (neuropsychological illness and assessment are not included in this medical approach).

The purpose of these classification systems is to provide clear descriptions of diagnostic categories of mental disorders in order to aid clinicians in diagnosing

**TABLE II**  
**Theoretical Approaches in Assessment: Model, Targets, Methods, Contexts, and Goals**

<i>Theoretical approach</i>	<i>Targets</i>	<i>Methods</i>	<i>Goals</i>
Trait	Trait	Questionnaires	Description
	Factor	Intelligence tests	Classification
	Dimension	Rating scales	Prediction, prognosis, or selection
	Characteristic Syndrome		
Medical	Mental and behavioral disorders	Classification systems Psychopathological tests	Description Classification and diagnosis
Psychodynamic	Unconscious mental process	Projective techniques	Description
	Conflicts	Completion methods (words, phrases, stories, etc.)	Classification Diagnosis
	Defense mechanisms		Prediction Explanation
Behavioral	Behaviors:	Observation	Description
	Motor	Self-observations	Prediction
	Physiological	Self-reports	Functional
	Cognitive	Physiological measures	explanation
	Environment:		
Cognitive	Physical		
	Social		
	Cognitive processes	Performance on cognitive tasks and cognitive equipment	Description Prediction
	Cognitive strategies	Thinking aloud	Explanation
	Mental representations	Self-reports	Evaluation
Constructivism	Other cognitive constructs		
	Emotions verbally expressed		
	Constructed knowledge	Narrative assessment	Description
	Personal constructs, personal plans, meanings, etc.	Autobiography REP-Grid Personal documents Content analysis	

and referring to these disorders. From this perspective, and with the aim of establishing objective procedures for assessing mental and behavioral disorders, several tests have been developed (e.g., the Multiphasic Minnesota Personality Inventory). In sum, description and classification are the two main objectives of the medical model.

### 2.3. Psychodynamic

Included under the title of psychodynamic are a series of relevant targets: unconscious mental processes, conflicting attitudes generating anxiety, defensive mechanisms, and interpersonal patterns governed by developmental experience. As stated by Weiner (2003),

the psychodynamic approach serves the purpose of elucidating “aspects of personality structure and personality dynamics in way that clarify the role of drives, conflict and defense, and object representations in shaping people are likely to think, feel and act” (p. 1012).

Throughout the history of psychological assessment, several instruments have been developed from a psychodynamic perspective. Perhaps the most well known are projective techniques using semi-structured visual materials (such as the Rorschach tests or the thematic apperception test), drawing a specific object (draw-a-person, draw-a-tree, draw-a-family), and finally, approaches using verbal material (words, sentences, stories, etc.). The goals of the dynamic model

include description and diagnosis as well as prediction and explanation.

#### 2.4. Behavioral

The behavioral assessment approach has been defined in several ways throughout the 20th century, but three fundamental pillars of these definitions can be identified and reduced to the following: (1) human behavior can be explained by the interaction between the person (from a biological and psychological point of view) and the environment; (2) there are three different modes or systems: motor, cognitive, and psychophysiological; and (3) psychological problems have multiple causes, essentially environmental. Thus, as Table II shows, targets in behavioral assessment are motor, cognitive, and physiological responses, in addition to environmental conditions.

A methodological condition of behavioral assessment is that all targets should be assessed through the multimethod approach. Thus, assessment of behavioral and environmental conditions and their respective interactions takes place through observation (direct observation as well as observations by relatives), self-observation, self-report, and psychophysiological techniques. Finally, because behavioral assessment emerges from a clinical perspective, its objectives are description, prediction, functional explanation (need for treatment), and treatment evaluation. The main goals of the behavioral psychologist are related to the environmental or personal conditions that can be functionally linked to target behaviors and provide indications for the planning of treatment.

#### 2.5. Cognitive

The cognitive model refers to two theoretical perspectives, the first related to the study of cognitive processes and the second associated mainly with therapy and treatment, under the assumption that cognitive processes or structures (or verbal expression of emotions) are causal aspects to be taken into account if the intention is to change behavior.

Table II shows the most important targets for the cognitive model: cognitive processes (such as attention, perception, learning and memory, language, and imagery), mental representations (e.g., visuospatial representations, automatic thinking), cognitive structures (cognitive abilities), and other cognitive constructs or forms of emotional expression. These processes, strategies or structures are assessed mainly through subjects' performance in cognitive tasks. For example, attention

can be measured by asking the subject to find a specific letter (e.g., "m") in a text, or to press a button when a specific cue appears on a screen (testing reaction time).

Cognitive processes or strategies can also be measured using meta-components. Thus, the subject may be asked to "think aloud" while solving a given task. The cognitive model also considers the meta-component of emotions, so that subjects may be asked to self-report what they feel in a particular situation. Finally, the objectives of the cognitive model—in the same line as those of the behavioral model—are the following: description (how the subject can be described on the basis of these processes and structures), prediction (what the subject can be expected to do on the basis of these processes), explanation (which processes may be controlling targets, so that they can be manipulated in order to change targets), and evaluation (to what extent treatment has produced the expected changes).

#### 2.6. Constructivist

The constructivist theoretical approach to psychology stresses conceptual, epistemological, methodological, and ontological characteristics, as opposed to an objective and realistic view of the science. From this approach, all knowledge is considered as a social product, so that there cannot be any universal law or principles: subjective aspects and local custom or practices dictate. Therefore, targets must be selected for each individual and his or her specific way of constructing social reality, the meaning of life, or personal plans, or of producing personal constructs. Methods are idiographic rather than standard. Finally, utility is the most important characteristic of any assessment, as opposed to validity. Table II shows these idiographic targets making up the basic personal knowledge developed across the subject's life span.

Several methods have been developed within the constructivist approach, all of which attempt to capture the essence of the person in question. Perhaps the most well known is the repertoire grid technique. Originally developed by Kelly in the 1950s, several different versions have appeared over the last half-century, all with the basic purpose of assessing the subject's personal constructs. Other methods involve the use of narratives developed by different authors, through which a sense of identity, coherence, meaning of life, and other personal phenomena can be assessed. With these narrative-based methods, the use of autobiography is an essential element for assessing these constructs. Finally, the data from these methods is analyzed through content analysis, which is

essential to constructivist methodology, because hermeneutics is, basically, a system for discovering targets.

### 3. METHODS AND TECHNIQUES

Marx and Hillix made a distinction between scientific method, methods, and techniques (methodic techniques or instruments). While the scientific method is the basic strategy for the advancement of science (e.g., the hypothetico-deductive method), methods are general data collection procedures, and techniques can be defined as those specific instruments or tests developed for collecting data in a particular field.

An important body of knowledge within psychological assessment is the set of methodic techniques for data collection. However, before going on to describe these methods, it is necessary to first identify some general conditions of psychological assessment methods.

1. Methods, techniques, and instruments should not be confused with psychological assessment. As stated previously, methods and tests constitute the tools for data collection and testing hypotheses, while psychological assessment is a lengthy process of problem-solving and decision-making.

2. There are standard and non-standard methods. All techniques and instruments have a set of stimuli (questions, tasks, etc.), a set of administration procedures, and a set of responses; finally, subjects' responses can be transformed into scores. Standard methods or tests are those that have a standard set of stimuli, standard administration procedures, and standard forms of obtaining subjects' responses—responses that can be transformed into standard scores. Standard methods usually have sound scientific properties and present psychometric guarantees. These standard methods, which allow between-subject comparisons, are known as psychometric tests. But not all data collected in psychological assessment derive from standard methods.

3. As described above, assessment targets are specific psychological characteristics (intelligence, cognitive skills, personality dimensions, behavioral responses, defense mechanisms, environmental conditions, etc.) with several degrees of inference. That is, the majority of these targets are psychological concepts that are operationalized by means of a given method, technique, or measurement device. Thus, all methods in psychology can be considered procedures for operationalizing a given psychological construct or target for assessment. Because each method has its own source

of error, the best way to conduct a rigorous assessment is to use several methodic techniques, tests, or measures or, in other words, to triangulate a given concept through multimethod. For example, if in assessing a particular personality construct (e.g., extroversion) the subject can be administered a standardized self-report, he or she can be given a list of adjectives, and his or her behavior can be observed in an analogue situation or by administering a projective technique.

With these three general conditions in mind, consider the broad categories in which psychological assessment methods can be classified: interview, observation, self-reports, and projective techniques. It should be borne in mind, however, that these four categories can themselves be divided into more specific sub-categories.

#### 3.1. Interview

Interviewing is the most commonly used method in psychological assessment in all basic and applied fields: clinical, educational, work and organizations, forensic, laboratory, and so on. It has the broadest scope, so it can be used for assessing any psychological event (subjective or objective, motor, cognitive or physiological).

For assessment purposes, an interview is basically an interactional process between two people in which one person (the interviewer) elicits and collects information with a given purpose or demand, and the other person (the interviewee, client, or subject) responds or answers questions. The information collected can refer to the interviewee (motor, cognitive, or emotional responses) or to relevant others. Thus, the interview can be thought of as a self-report or as a report to or by another.

Because the interview is an interpersonal process, it has the following several phases:

1. Interview preparation: Before starting an interview, the psychological assessor should plan the interview. What are the purposes of the interview? What are the subject's characteristics? What format will the interview take? How long will the interview last?

2. Interview starting point: A brief presentation by the interviewer about the purpose of the interview, what the interviewer is expecting, and the plans for the assessment process.

3. Body of the interview: The interview is always led not only by the purpose of the assessment but also by the theoretical approach of the interviewer. These two principal conditions influence the content (what type of

question) and the format (amount of structure in questions and answers) of the interview. By way of example, the main common and shared areas of an interview are as follows: sociodemographics, interview purpose or demand, family and social condition, education, work, interest and hobbies, health, beliefs, and values.

4. Closing the interview: The interviewer closes the interview by briefly summarizing the data collected and describing the subsequent steps or actions.

There are several typologies of the interview depending on: (1) the structure of the questions and answers (structured, semi-structured, non-structured), (2) the context in which the interview takes place (clinical, personnel selection, counseling, forensic, etc.), (3) interviewer style (directive, non-directive), (4) theoretical perspective (dynamic or psychoanalytic, behavioral, phenomenological, etc.), and (5) number of subjects assessed (individual, group, family, team group, etc.).

The interview as an assessment method has several advantages and limitations, such as the following:

1. It offers a situation in which subjects can act and interact with several verbal stimuli (or questions) and can be observed. Also, in this situation several interpersonal emotional relationships (interviewer–interviewee) develop that are highly relevant to forthcoming psychological actions.

2. The interview is a broad technique for collecting data; that is, any condition or purpose (past or present; internal or external; descriptive, classificatory, or explanatory, etc.) can be assessed. Also, this flexibility of the interview allows psychologists to clarify a given situation or action of the subject.

3. However, information collected in interviews may present considerable bias and error deriving from the interviewee, the interviewer, and the situation. Inter-interviewer agreement and test-retest reliability is relatively low, and validity depends on the construct assessed.

4. Finally, as emphasized by Meyer et al., although most types of psychological testing—including the interview—show evidence of their predictive validity, the use of multimethods is essential in order to gain a complete understanding of each case.

### 3.2. Observational Methods

All psychological assessment contains observations. Indeed, as Anastasi (1988) stated: “a psychological test is essentially an objective and standardized measure of a sample of behavior” (p. 23). Thus,

observation is one of the broadest categories in psychological assessment. Common intelligence tests (such as the Wechsler Intelligence scales), as well as cognitive or psychophysiological measurement using special equipment, can be used as observational procedures. In brief, an observational method consists of a set of data collection procedures containing a protocol with a predefined set of behavioral categories in which subjects’ behaviors are coded or registered. These behaviors are external, motor events observed in natural, analogue, or test situations; observation may sometimes involve the use of sophisticated equipment such as tachistoscopes, polyreactographs, or psychophysiological instruments that allow response amplification.

Observational methods have been classified depending on the following: (1) the observer: participant versus non-participant observer; (2) the target or observed event: overt event (motor response or stimuli parameter) versus subject’s attribute or molar characteristic of the observed situation; (3) the situation in which observations take place: natural versus artificial or standardized situations (or even with equipment amplifying the target); and (4) the protocol: coding system versus open description of the continuum of behavior. Table III shows examples of this classification system.

Observer, subjects and targets, and observational protocol are sources of bias of the observational assessment method. Below is a summary of how to maximize the accuracy of observational methods:

1. *Observer.* Observational methods in psychology should maximize objectivity through the assessment of overt (observable) events, ongoing observer training, use of several observers, and checking of inter-observer agreement.

2. *Subject observed.* One of the most important biases of observational methods is subject reactivity. Reactivity can be minimized in various ways, such as use of natural settings, subjects being unaware, and long habituation periods.

3. *Observational systems.* Coding or registration systems require well-defined targets and procedures, a small number of observation categories, and standard data about psychometric soundness.

### 3.3. Self-reports

As defined elsewhere, the self-report is a method for collecting data whose source is the subject’s verbal message about him- or herself. Self-reports are the most

TABLE III  
Type of Observational Methods and Examples of Observational Techniques

Observer	Target	Situation	Protocol	Example
<b>Nonparticipant</b> Expert observes target behaviors or attributes	Motor behaviors Product of behavior Attributes	Natural situation registered through video, film, two-way mirror, etc.	Continuum of observed behavior Category system Behavior checklist Rating scales	Standardized observational code (SOC-III) (Cerezo <i>et al.</i> , 1991)
<b>Participant but expert</b>	Motor behavior Product of behavior (dimensions) Verbal behaviors Attributes	Standardized/Non-standardized situations	Category system Behavior checklist (occurrence, frequency accuracy, reaction time, duration, etc.) Rating Scales (Likert-type scales)	WAIS III (Wechsler, 2000)
<b>Participant and/or relatives</b>	Dimensions of motor and verbal behavior Attributes	Natural situation Non-standardized	Behavior checklist Rating scales	Eyberg behavioral inventory (Eyberg & Ross, 1978)

common assessment procedures for collecting data in psychology and psychological assessment. Self-reports provide information about thousands of targets—from historical events to motor behaviors in the present situation, even subjects' future plans. In fact, human beings can report on their private, unobservable, subjective world, their thoughts, and their emotions, as well as public observable events. This means that, in answering a verbal question, subjects may carry out certain mental operations (from remembering what they did a few minutes previously to inferring why they acted in a certain way a long time ago). Such operations interfere with the fidelity of self-reports.

Also, when a self-report or a set of self-reports is collected, psychologists can either take them as they are (from an isomorphic perspective) or make inferences from them after carrying out statistical manipulations (e.g., item aggregation or item anchoring in new scales). In other words, in order to test theoretical hypotheses, psychologists interpret the self-reports as indices of internal structures or processes, such as attitudes, personality characteristics, motives, and other intrapsychic concepts developed from trait theory, medical models, cognitive psychology, and/or dynamic theories. Thus, since the very beginning of scientific psychology, aggregates of self-reports assessed via questionnaires through inter-subject designs (subject aggregates) have been

used as one of the main assessment methods for describing, classifying, predicting, and even (tautologically) explaining human behavior.

In psychological assessment, clients' self-reports are collected in different formats. As already mentioned, interviews frequently represent the starting point and conclusion of a professional relationship in clinical, educational, and work settings. When self-reports are used as formal procedures for data collection, they imply a broad methodological category: interviews always involve self-reports, and self-observation and self-monitoring also require self-reports. Finally, self-reports can be obtained through questionnaires, inventories, or scales containing a set of relevant verbal statements and a variety of response formats (true or false options, checklists, scaled responses, etc.).

As emphasized elsewhere, self-reports have several sources of bias. Whatever the type of event reported—from objective events to subjective ones—several internal and external conditions, such as length of the question, formulation of the question, type of answer and response requested, and subject's characteristics, strongly affect the fidelity of self-reports. Table IV shows examples of general issues related to self-reports.

In summary, human beings are able to report on a myriad of events and are an important source of information; several assessment methods (such as interviews

TABLE IV  
Examples of General Issues in Self-Reports<sup>a</sup>

<i>Targets</i>	<i>Subject mental operations</i>	<i>Psychologist inferences</i>	<i>Response format</i>	<i>Type of self-report</i>
Motor response, external events	Accessibility to the target	Self-report as data vs	Open	Interview
Cognitive behavior	Level of inference of required transformation	Self-reports as sign	YES/NO, multiple choice, rating-scale	Questionnaire, inventory, scales
Subjective appraisal Psychophysiological responses	Time: Past, present, future		Latency, occurrence, frequency, amplitude, etc.	Self-observation, self-monitoring

<sup>a</sup>Modified from Fernández-Ballesteros & Márquez, 2003.

and self-observation, self-monitoring, questionnaires, inventories or scales) are based on self-reports. Even though self-reports have important sources of errors, so that self-report accuracy is threatened by several types of response distortions (positive and negative impression management, faking, acquiescence, and social desirability, among others), these can, at least partially, be measured and controlled. Finally, it is commonly accepted that self-report questionnaires can be optimized through various strategies, and that their accuracy can be maximized. When self-reports are taken as a source of behavioral information, the level of inference should be as low as possible, items should be as direct as possible, questions should be specific and with situational reference, the level of transformation of the information requested should be as low as possible, and finally, the information requested should refer, preferably, to the present.

### 3.4. Projective Techniques

According to Pervin (1975), a projective technique “is an instrument that is considered especially sensitive to covert or unconscious aspects of behavior, permits or encourages a wide variety of subject responses, is highly multidimensional, and evokes unusually rich or profuse response data with a minimum of subject awareness concerning the purpose of the test” (p. 33).

The basic assumptions of projective techniques are closely related to the psychodynamic model and psychoanalytic theory: (1) they involve the presentation of stimuli relatively free of structure or cultural meaning, consisting of inkblots, pictures, incomplete verbal sentences or stories, or performance tasks; (2) these

materials allow subjects to express idiosyncratic and holistic aspects of their personalities, and allow the assessment of dynamic constructs as well as defense mechanisms against anxiety, and conscious as well as unconscious processes and structures; and (3) subjects are not aware of the purpose of the test and the meaning of their responses to the test. Table V shows four main categories of projective techniques that can be identified depending on the material used and the task involved.

Although projective methods have been widely used, they have received considerable criticism, and continue to cause controversy. In fact, all the characteristics of projective methods—ill-defined stimuli, idiographic responses, non-standardized procedures, and subjective evaluation—have contributed to a lack of psychometric soundness. In sum, much more research is required for projective methods to be used as scientific tools. Nevertheless, there is an exception, as Weiner (2001) has emphasized: “the Rorschach Inkblot Method has been standardized, normed, made reliable, and validated in ways that exemplify sound scientific principles for developing an assessment instrument” (p. 423).

## 4. THE PROCESS OF ASSESSMENT, TREATMENT, AND EVALUATION

As mentioned previously, assessment should not be confused with its methods and tests, nor with a specific applied setting (such as the clinical one). It begins when a client makes a specific demand of a psychologist in a given context. This demand triggers a long process of problem-solving and decision-making that requires a series of tasks ordered in a specific sequence. This sequence

**TABLE V**  
**Types and Examples of Projective Techniques Depending on Materials Used and Task**

<i>Type of technique</i>	<i>Material</i>	<i>Task</i>	<i>Examples</i>
Structural	Inkblots	Describe what it looks like	Rorschach Inkblots test (Rorschach, 1929)
Thematic	Pictures	Tell a story	Thematic apperception test (TAT) (Murray, 1936)
Expressive	White paper and pencil	Draw a picture	Machover Draw-a-Person test (Machover, 1948)
Associative	Words, incomplete sentences or stories	Complete the sentences or stories	Free association test (Rapaport, 1965)
Performance	Constructional elements (houses, trees, fences, etc.)	Build a town	Town test (Arthur, 1949)

is dictated by the objectives the psychologist must fulfill: description, classification/diagnoses, prediction/prognosis, explanation, and control. When the demand is for control, implying the need for an intervention, evaluation is an operation to be performed for appraising the treatment, or testing the extent to which the targets have changed in the expected direction. Thus, when assessment is conducted in order to produce change and is followed by treatment, evaluation should be the final step in the assessment process.

After 30 years of research into the assessment process, guidelines for the assessment process (GAP) have been set up in order to formalize it and published so that they can be discussed by the scientific community. The GAPs have been developed with the following purposes: (1) to assist psychological assessors in their efforts to optimize the quality of their work; (2) to assist the client coming from outside psychology in evaluating assessment tasks by allowing quality control; and (3) to facilitate the teaching of assessment, the standardization of practical considerations, and the design of advanced professional training programs.

Table VI presents the general framework of the assessment process. As it can be seen, the assessment, treatment, and evaluation process has four main steps: (1) analyzing the case; (2) organizing and reporting results; (3) planning the intervention; and (4) evaluation and follow-up. The two first phases deal with case formulation as well as with integrating, reporting, and discussing the case, and therefore refer to all assessment contexts (clinical, educational, work and industry, etc.), because description, classification/diagnosis, and prediction/prognosis are the assessment objectives in all cases. The third and fourth phases refer to those cases

demanding control and change and requiring explanatory assumptions, experimental designs, and, finally, treatment evaluation. Each step has several substeps, and from the substeps a set of 96 guidelines has been developed. The GAPs are procedural suggestions and should be taken as recommendations for helping assessors to cope with the complexities and demands of assessment processes in various applied contexts.

## 5. CONCLUSIONS

Psychological assessment is a subdiscipline of scientific psychology devoted to the study of a given subject with different purposes (describing, diagnosing, predicting, etc.) and in different settings (clinical, educational, etc.). Psychological assessment should not be confused with its methods or tests, since it involves a lengthy and complex process of decision-making and problem-solving that begins with the client's demands. There are several theoretical models proposing specific targets, methods, and assessment objectives.

Psychological assessment uses a variety of methods: interview, observation, self-reports, and projective techniques. Some of these have standard sets of stimuli, administration procedures, and ways of measuring subjects' responses and transforming them into standard scores; these methods are the well-known psychometric tests, but they are not the only means of collecting data in assessment. All assessment methods have particular sources of error, and all can be considered as fallible. The use of multimethod is highly recommended. Through the complex process of decision-making and problem-solving involved in

**TABLE VI**  
**Assessment Process: Basic Framework and Steps**

---

**START OF PROCESS:** Process begins when a person/institution (client) asks an assessor to answer a question or give professional advice about a subject/single case.

**PRECONDITIONS:** Assessor examines whether he/she is qualified to satisfy the demand and whether the demand is in line with ethical criteria.

**0. GENERAL PRINCIPLES**

**1. ANALYZING THE CASE (DESCRIPTIVE ASSESSMENT)**

1.1 ANALYZING DEMANDS, COMPLAINTS, AND/OR GOALS: Gathering general information about client's and/or subject's referral question.

1.1.1 Investigating and evaluating client's and/or subject's questions.

1.1.2 Synthesizing client's demands and aspects of the general problem situation.

1.1.3 Formal agreement.

1.2 FORMULATING TESTABLE ASSESSMENT HYPOTHESES ABOUT THE CASE: Converting demands, complaints, and/or goals into testable hypotheses.

1.2.1 Formulating questions in technical terms, based on information gathered.

1.2.2 Operationalizing technical terms by means of assessment procedures.

1.3 COLLECTING INFORMATION: Gathering information relevant to the questions.

1.3.1 Planning administration of assessment procedures.

1.3.2 Applying assessment procedures.

1.3.3 Evaluating the application of assessment procedures.

1.4 INFORMATION PROCESSING, RELATING DATA COLLECTED TO THE ASSESSMENT QUESTIONS

1.4.1 Analyzing data.

1.4.2 Drawing assessment conclusions.

**2. ORGANIZING AND REPORTING RESULTS:** Technical preparation of results and reporting to client/subject.

2.1 INTEGRATING RESULTS: Answering the client's/subject's questions as completely as possible.

2.1.1 Combining results into a comprehensive case formulation.

2.1.2 Formulating conclusions with respect to client's/subject's demands.

2.2 REPORTING: Written and/or oral presentation of reports.

2.2.1 Requirements for report generation.

2.2.2 Including relevant information in the report

2.2.3 Making report understandable.

2.3 DISCUSSING AND DECIDING

2.3.1 Discussing report with client, subject, and/or relevant others.

2.3.2 Analyzing whether general circumstances warrant stop, re-start, or moving on to an intervention.

**3. PLANNING THE INTERVENTION:** If the assessor considers that an intervention is required, several assessment operations are necessary before treatment administration.

3.1 SELECTING AND TESTING SPECIFIC INTERVENTION HYPOTHESES:

3.1.1 Selecting and operationalizing intervention and outcomes variables.

3.1.2 Reviewing and deciding on intervention procedures that best fit the single case.

3.1.3 Selecting and assessing relevant variables for monitoring.

**4. EVALUATION AND FOLLOW-UP:** If an intervention has been carried out, several assessment operations should be conducted.

4.1 COLLECTING DATA ABOUT EFFECTS OF INTERVENTION:

4.1.1 Inspecting already-available data.

4.1.2 Collecting post-intervention data.

4.2 ANALYZING INTERVENTION OUTCOMES:

4.2.1 Drawing conclusions from data gathered on the effects of the intervention.

4.2.2 Reporting results to client, subject, and/or relevant others.

4.2.3 If necessary, written report to client, subject, and/or relevant others.

---

*Continues*

Continued

#### 4.3 FOLLOW-UP:

- 4.3.1 Planning follow-up in agreement with client and/or subject.
- 4.3.2 Assessing subject according to the established plan.
- 4.3.3 Analyzing results.
- 4.3.4 Discussing results with client, subject, and/or relevant others.
- 4.3.5 If necessary, written report to client, subject, or relevant other.

**END OF PROCESS:** The assessment process comes to an end, if the assessor terminates the professional relationship with the person/institution (client) and subject/single case with respect to the assessment tasks.

psychological assessment, psychologists test hypotheses about the case, achieve assessment objectives, and draw certain conclusions in an attempt to respond to the demands of the client.

### See Also the Following Articles

Clinical Assessment ■ Diagnostic and Statistical Manual of Mental Disorders ■ Intelligence Assessment ■ International Classification of Diseases (WHO) ■ Personality Assessment ■ Psychological Assessment, Standards and Guidelines for ■ Psychometric Tests ■ Psychophysiological Assessment ■ Traits

### Further Reading

- Arthur, H. (1949). *Le village. Test d'Activité Créative*. Paris: Pub.
- Baer, R. A., Rinaldo, J. C., & Berry, D. T. R. (2003). Self-report distortions. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (Vol. 2, pp. 861–866). London: Sage.
- Cook, T. D. (1985). Postpositivist critical multiplisms. In L. Shotland, & M. M. Marck (Eds.), *Social science and social policy*. Beverly Hills, CA: Sage.
- Cerezo, M. C., & Whaler, R. G., et al. (1991). *SOC-III Standardized Observational Code*. Madrid: MEPSA.
- Cronbach, J. L. (1975). Five decades of public controversy over mental testing. *American Psychologist*, 30, 1–4.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis. Verbal reports as data*. Cambridge, MA: MIT Press.
- Eyberg, S. M., & Ross, M. (1978). Assessment of child behavioral problems. *Journal of Clinical Psychology*, 3, 113–116.
- Fernández-Ballesteros, R. (2002). Behavioral assessment. In Smelser, N. J., & P.B. Baltes (Eds.), *International encyclopedia of social and behavioral sciences*. New York: Pergamon.
- Fernández-Ballesteros, R. (2003). Self-report questionnaire. In S. N. Haynes & E. Heiby (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 3, pp. 194–221). New York: John Wiley & Son.
- Fernández-Ballesteros, R., De Bruyn, E. E. J., Godoy, A., Hornke, L., Ter Laak, J., Vizcarro, C., Westhoff, K., Westmeyer, H., & Zaccagnini, J. L. (2001). Guidelines for the Assessment Process (GAP): A proposal for discussion. *European Journal of Psychological Assessment*, 17, 187–200.
- Fernández-Ballesteros, R., & Márquez, M. O. (2003). Self-reports. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 871–877). London: Sage.
- Freixas, G. (2003). Subjective methods. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment*. London: Sage.
- Goldfried, M. R., Stricker, G., & Weiner, I. (1971). *Rorschach handbook of clinical and research application*. Englewood Cliffs, NJ: Prentice Hall.
- Goldstein, G., & Hersen, M. (1990). Historical perspectives. In G. Goldstein, & M. Hersen (Eds.), *Handbook of psychological assessment* (pp. 3–21). New York: Pergamon Press.
- Goldstein, G., & Beers, S. R. (Eds.). (2003). *Intellectual and neuropsychological assessment*. New York: Wiley.
- Haynes, S. N., & O'Brien, W. H. (2000). *Principles and practice of behavioral assessment*. New York: Kluwer.
- Haynes, S. N., & Heiby, E. M. (Eds.). (2003). *Behavioral assessment*. New York: Wiley.
- Machover, K. (1948). *Personality projection in the drawing of the human figure*. New York: Ch. C. Thomas.
- Marx, M. H., & Hillix, W. A. (1967). *Sistemas y teorías psicológicas contemporáneas. (Contemporary psychological theories and systems)*. Buenos Aires: Paidós.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment. *American Psychologist*, 56: 128–165.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Murray, A. H. (1936). *Exploration in personality*. New York: Oxford Press.
- Neimeyer, G. J. (Ed.). (1993). *Constructivist assessment*. Newbury Park: Sage.
- Pervin, L. A. (1975). *Personality, assessment, and research*. New York: John Wiley & Sons.
- Rapaport, D. (1965). *Test de diagnóstico psicológico (Manual of Diagnostic testing)*. Buenos Aires: Paidós.
- Rorschach, H. (1921). *Psychodiagnostik*. Berne: Hans Huber.

- Segal, D. L., & Frederick, L. (2003). Objective assessment of personality and psychopathology: An overview. In Segal, D. L., & Hilsenroth, M. (Eds.), *Personality assessment*. New York: Wiley.
- Sundberg, N. (1976). *Assessment of persons*. New York: Pergamon.
- Schwartz, N., Park, D. C., Knauper, B., & Sudman, S. (1998). *Cognition, aging and self-reports*. Ann Arbor, MI: Psychology Press.
- Scriven, M. (1991). *Evaluation thesaurus*. London: Sage.
- Weiner, I. (2001). Advancing the science of psychological assessment: The Rorschach Inkblot method as exemplar. *Psychological Assessment*, 13, 423–433.
- Weiner, I. (2003). Theoretical perspective: Psychoanalytic. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment*. London: Sage.
- Wiggins, J. (1973). *Personality and prediction. Principle of personality assessment*. Reading, MA: Addison-Wesley.



# Assessment in Sport Psychology

Marc-Simon Sagal, Paul Thomas Sagal, and Geoffrey E. Miller

Winning Mind, LLC, San Diego, California, USA

1. Introduction
  2. Role of Assessment
  3. Methods of Assessment
  4. Common Distinctions
  5. History and Current Assessment Tools (North American)
  6. List and Description of Frequently Used Tests
  7. Formal Assessment Guidelines and Ethical Considerations
  8. Future Directions
- Further Reading

## GLOSSARY

**applied sport psychology** The application of psychology, typically toward the improvement of performance, gained through the study of the psychological factors that affect and are affected by participation and performance in sport.

**consensual validation** The process of confirming, through interview and behavioral observation, the existence of certain psychological qualities previously identified through formal assessment.

**response set** A situation-specific approach employed by an individual undergoing a psychological assessment to manipulate resulting assessment data (e.g., faking good).

**response style** An approach that reflects cross-situational, core personality characteristics and that can have an equally strong impact on assessment results (e.g., pessimism).

**sport psychological assessment** The systematic attempt to measure those psychological characteristics that influence performance and behavior in sport contexts.

**sport psychology professional (SPP)** An individual with appropriate academic credentials and experience who

applies this expertise responsibly in sport settings; SPPs are interested in helping individuals to improve performance, appreciate the overall sporting experience, and enhance personal development and well-being.

**sport-specific measures** Inventories that attempt to elicit responses directly relevant to performance in sport generally or in one sport specifically.

**trait** A relatively enduring characteristic or habit.

**state** An occurrent characteristic (i.e., situationally dependent).

Assessment enables sport psychology professionals (SPPs) to obtain crucial performance-relevant information. The principal aim is to help athletes perform at an optimal level. To this end, both formal and informal methods are employed. These include interviews, behavioral observation, psychological testing, surveys, and inventories. Effective assessment requires sensitivity to test validity and reliability, response sets and styles, consensual validation, and ethical requirements. This article includes a brief history and a discussion of future directions.

## 1. INTRODUCTION

At its most basic level, assessment involves information gathering, interpretation, and evaluation. We all constantly engage in one form of assessment or another. It is how we make sense of the world. Sport psychology assessment is the systematic attempt to measure

psychological characteristics with the aim of improving sport experience and performance. Within the field of applied sport psychology, the method of assessment typically depends on the context (e.g., purpose, clients, time constraints). Although most sport psychology professionals (SPPs) associate assessment with more formal approaches (typically paper-and-pencil questionnaires and now online questionnaires), there are a variety of ways in which information can be collected and distilled. This article focuses on the practical applications of assessment and its use in applied settings. It also provides general information on selected assessment methods and a number of widely used instruments and tests.

## 2. ROLE OF ASSESSMENT

SPPs use assessment to gain crucial information about the people with whom they work. The aim is to help each of these individuals reach the level of performance desired, whether this is the level of a professional or Olympic athlete, an amateur competing in organized club competition, or a purely recreational sport enthusiast. But what exactly does an SPP need to understand about an individual to help? The answer to this question is controversial. Some SPPs believe that a clinical understanding of the client is necessary. Most SPPs focus on gaining more direct performance-relevant information (e.g., how the athlete pays attention, under what situations performance breaks down). Even if there were agreement on exactly what should be assessed, it would be difficult to get agreement on assessment methods. This difficulty increases where there is no agreement on which factors are most important.

There are four primary applications of applied sport assessment:

- Performance enhancement (individual and team)
- Athlete selection and screening
- Injury recovery
- Sport enjoyment

Some SPPs believe that all applications can be subsumed under the primary category of performance enhancement. Others emphasize the importance of having a positive sporting experience regardless of performance issues. Most people choose to work with SPPs because they want to be better at their sport. The presenting issue may be “recovering quickly from an injury,” “resolving conflicts among team members,” or “identifying mental strengths and liabilities,” but the bottom-line concern is with performance (Table I).

TABLE I

### Common Reasons for the Use of Assessment

- 
1. To measure variables that are important to performance, sport enjoyment, and a variety of other factors that contribute to the quality of the sport experience.  
Examples include the following:
    - a. To determine current mental skill ability
    - b. To determine presence of factors associated with overtraining
    - c. To monitor anxiety
    - d. To assist in the presence of clinical issues
  2. To reduce the amount of time it would ordinarily take to collect this relevant information through other means (e.g., observation)
  3. To use a repeatable method for gaining critical information to help athletes
  4. To quantify relevant information for analysis over time or between populations
  5. To determine whether new factors or groups of factors are more or less important contributors to desired outcomes, states, or behaviors
- 

## 3. METHODS OF ASSESSMENT

SPPs apply a variety of assessment methods to understand, predict, and modify behavior, including the following:

- Interviews
- Behavioral observation (including video review)
- Psychological testing, surveys, and inventories
- Third-party anecdotal data collection (from parents, coaches, teammates, etc.)

Typically, context will help to determine the most appropriate methods to employ. In most situations, SPPs will gather information using more than one of these methods.

### 3.1. Interviews

The most common form of assessment takes place through the process of formal and informal interviewing. There are typically three purposes for conducting an interview: to identify performance issues, to determine ways in which to intervene, and to validate inventory and questionnaire data. The interview itself can be structured by a set of preestablished questions and topics, or it can be free-flowing with a more open and dynamic feel. Topics and lines of inquiry depend on many factors, including the type of SPP–athlete relationship, the phase of the intervention/program, the severity

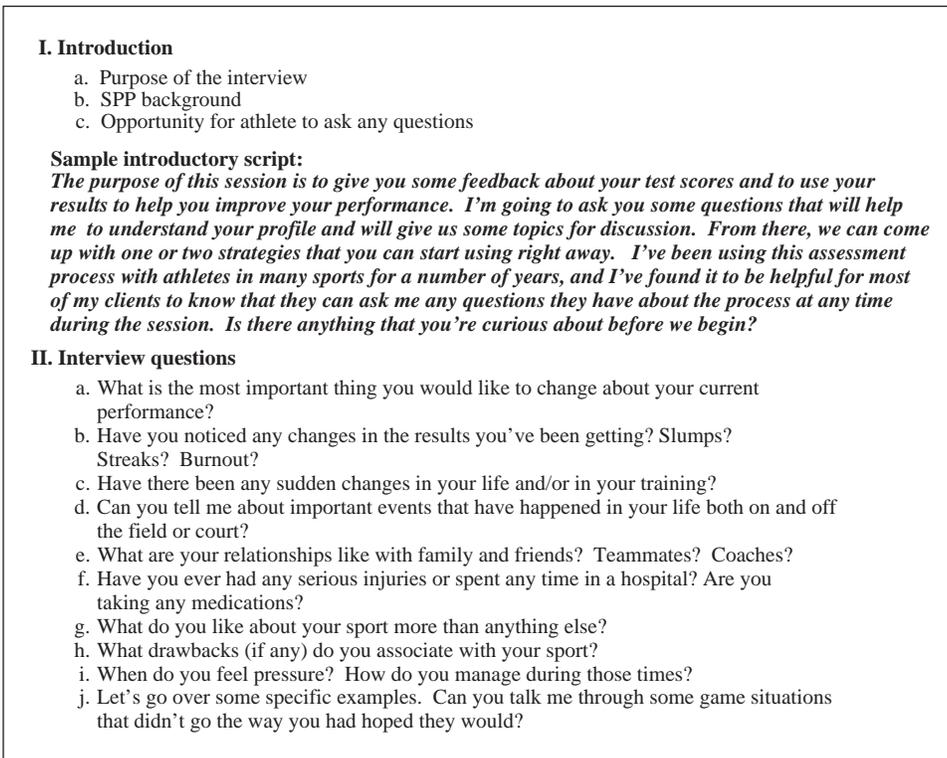


FIGURE 1 Sample interview outline.

of the issue(s), and the objective of the work. Many items on sport psychology inventories had their origins as interview questions. This should not be surprising. Like physicians who use intake forms and a wide range of tests to help learn about their patients, so too are SPPs interested in providing an efficient and consistent way in which to get to know their clients (Fig. 1).

### 3.2. Behavioral Observation

Behavioral observation is a straightforward process that involves watching live or recorded performance. It provides at least one distinct advantage over all of the other assessment methods in that SPPs can see firsthand how athletes behave in the most relevant settings (e.g., practice, competition). This information is invaluable because there are cues that may be identified in real-life settings that might not arise under less realistic conditions. Athletes might consistently make mistakes in situations but be unaware of the cause. The purpose of behavior observation is to watch athletes perform in those situations where insight into psychological performance factors can be ascertained. Both structured (i.e.,

controlled) and unstructured observation can be used to acquire this type of information. Trained SPPs can learn a great deal by attending to a variety of “targets” that include, but are not limited to, actual performance, verbal behavior, and body language. For example, slumped shoulders after a missed shot might reflect a slip in confidence. These kinds of observations can provide invaluable discussion points that can be critical in helping athletes to understand themselves and their performances (Table II).

TABLE II  
Behavioral Observation Checklist

<ul style="list-style-type: none"> <li>• <i>Verbalizations and utterances:</i> Listen for content, intensity, and timing.</li> <li>• <i>Attention and nonverbal expressions:</i> Look for facial reactions, eye movement, and frequency, intensity, and general indicators of stress.</li> <li>• <i>Posture and movement:</i> Look for bracing (stiffness), fluidity, limping, or other awkward positioning.</li> <li>• <i>Social:</i> Observe interactions with teammates, parents, coaches, and the like.</li> </ul>
--

### **3.3. Psychological Testing, Surveys, and Inventories**

Although the definition of assessment offered here is rather broad, most people identify assessment with the printed questionnaire. Ranging from several questions to several hundred questions, these instruments usually ask athletes to respond to items by selecting from some form of multiple-choice response list. It should be noted that many SPPs are careful not to use the label “test” and prefer instead to use a variety of more “acceptable” names (e.g., survey, inventory, assessment). There are many ways in which instruments may be categorized, with some being more theoretically sophisticated than others. The following labels are commonly applied:

- Clinical
- Cognitive
- Attentional
- Personality and emotional intelligence
- Sport-specific
- Team/Social interactive

Although these different kinds of information are not uniquely attainable through psychological testing, the advantage of such testing is that it provides a more objective, more scientific means of acquiring data. Presumably, SPPs want to be able to know as much as possible so as to improve performance. Testing can provide a scientific psychological portrait of the athletes with whom they work. Clinical inventories can inform SPPs about qualities such as depression, anxiety, bipolarity, and mania. Cognitive inventories measure things such as verbal intelligence and speed of learning. Attentional inventories identify preferred styles of concentration. Personality tests involve measures of confidence, self-esteem, social attitudes, and risk taking. Sport-specific tools provide information about performance-relevant variables in sport settings. Team and social interactive inventories can help to identify factors particularly relevant in group settings. Examples would include measures of cohesion and communication style.

### **3.4. Third-Party Anecdotal Data Collection**

Important information that is helpful to SPPs is often acquired through formal and informal communication with people in a special position to provide insight about clients. Third-party data collection may be as

simple as an interview conducted by someone other than the SPP (and, as such, may be viewed as a type of interview), but it can also be much less formally acquired information by parents, teammates, coaches, or other individuals who come in contact with the athlete. One formal version of this often-overlooked method of assessment is similar to the 360-degree review process widely used in business consulting, where an individual receives feedback on performance strengths and weaknesses from his or her superiors, peers, and subordinates so as to help formulate an individual action plan for career development. Informal versions include casual conversations and other means of information exchange.

## **4. COMMON DISTINCTIONS**

### **4.1. Formal Assessment Versus Informal Assessment**

Behavior can be studied formally or informally. An informal assessment is one that does not incorporate psychometrics or data collected for statistical analysis. Examples of informal assessment include reading non-verbal cues when speaking with someone, watching sport competitions, observing practice sessions without recording data, and interviewing athletes, coaches, and parents. Formal assessment involves data collection through psychometric tools, surveys, frequency distributions, and the like. Discussion of assessment in sport psychology is dominated by formal topics because it is relatively easy to examine the rationale, experimental design, reliability, validity, and motivation for construction of formal instruments. Informal assessment is, on the other hand, undervalued and less stable due to the difficulty in standardizing techniques. But veterans in the field develop a balance of formal assessment (which provides a scientific base for their work) and informal assessment (which allows them to anticipate client needs and come up with practical solutions to identified problems as they happen in real time). It should be noted that although this distinction is commonly accepted, it is not always so clear. For instance, controlled observations (not usually viewed as formal assessment) can incorporate observation schedules with preestablished behavior categories and, thus, provide data eligible for statistical analysis.

There is both an art and a science to assessment. It is certainly worth noting that there are a fair number of “gurus” in the field who work primarily by intuition

and feel. Many respected SPPs rely most heavily on this “sixth sense,” and it is likely that assessment by gut feelings will continue into the future.

#### **4.2. Subjective Assessment Versus Objective Assessment**

The distinction between subjective assessment and objective assessment is one commonly encountered in discussions of sport psychology. In general, subjective assessment encompasses various forms of interviews and observations on the part of SPPs. Objective assessment typically refers to psychological tools such as inventories and tests. Here, the distinction seems to hinge on the role of SPPs in the assessment process. In subjective assessment, SPPs are called on to interpret behavior directly. In objective assessment, SPPs use an appropriate instrument to provide additional information or clarification relevant to assisting athletes. What tends to confuse this distinction is that there are subjective components within objective assessment as well as objective components within subjective assessment. For example, formal scripted interviewing includes a measure of objectivity, and the interpretation of psychological inventories contains a large subjective component.

#### **4.3. State Versus Trait**

Because SPPs are ultimately interested in modifying behavior to improve performance, it is critical to understand whether a particular individual characteristic is likely to change over time. Traits are enduring characteristics that are likely present in a variety of situations. States are temporary conditions that are determined by situational factors. For example, even the most confident people can experience anxiety before an important speech or a big game. This could be called “state anxiety” and would be dependent on the situation in which an individual found himself or herself. However, an anxious person (i.e., one with “trait anxiety”) would likely become anxious before every speech or game. Gauging where certain characteristics fall on the state–trait continuum is helpful to SPPs as they use assessment information and apply methods of intervention.

#### **4.4. Sport-Specific Measures Versus General Measures**

Some assessment tools contain items that are framed in a sport-specific context to elicit responses more

relevant to the actual performance environment of the individual test-taker. Believing that sport-specific measures provide uniquely valuable information, many test developers have created sport-specific versions of existing inventories. This raises an interesting question: Is it more or less useful for items to focus on behavior within the situational context of primary concern? Not surprisingly, there is disagreement on the answer. Some believe that important information can be missed with sport-specific measures and that generally framed items will be more likely to pick up on those factors that are important in a variety of situations, including sport. This camp also believes that sport-specific items are more transparent and, thus, more easily faked or manipulated. Those who embrace sport-specific measures believe that the closer the context matches the particular performance environment of the athlete, the more validity the test will have and ultimately the more useful the results will be.

### **5. HISTORY AND CURRENT ASSESSMENT TOOLS (NORTH AMERICAN)**

Assessment has played an integral role in the history of sport psychology. Most historical reviews identify the Triplett social facilitation experiments in 1898 as the first recorded studies in the field. In showing that cyclists raced faster with a pacer than they did by themselves, Norman Triplett used a simple measurement tool still believed by many to be the most critical component in athletic competition—results.

Nearly 30 years passed before a definitive text on the subject of athletes and their thought processes would appear. In 1926, Griffith published the first of two books that would become the “big bang” in the creation of sport psychology. *Psychology of Coaching* and *Psychology and Athletics* were written by Griffith as collections of his findings from the first ever sport psychology laboratory on the campus of the University of Illinois. In *Psychology of Coaching*, Griffith attempted to identify character types of coaches and men. His typologies included the “crabber,” the “silent type,” the “talker,” the “loafer,” the “leader,” and the “follower.”

The 1960s were a watershed for assessment and sport psychology. The first modern inventory created specifically for use with athletes was created in 1969 by Tutko, Ogilvie, and their colleagues. The Athletic Motivation Inventory (AMI) was born out of Tutko and Ogilvie’s *Problem Athletes and How to Handle Them*. Until

this time, application of assessment in sport settings had meant adapting results of general personality and mental health diagnostics such as the Minnesota Multiphasic Personality Inventory (MMPI), 16 Personality Factor Questionnaire (16PF), and Thematic Apperception Tests. AMI attempted to measure traits categorized under the rubric of “Desire” factors and “Emotional” factors. Desire factors were drive, determination, aggression, leadership, and organization. Emotional factors were coachability, emotionality, self-confidence, mental toughness, responsibility, trust, and conscience development.

When international organizations began forming, communication increased between SPPs and assessment became a central focus. Could anxiety, personality, attention, moods, and motivation be tangibly measured, and could results be reliably reproduced, over time? If so, how would these tools contribute to the evolution of professional and amateur sport cultures in the real world? SPPs set to work on developing an arsenal of assessment tools for the sport environment.

One significant contributor, Rainer Martens, was a wrestling coach at the University of Montana who began to explore sport psychology after failing to help his talented team captain overcome prematch anxiety. In 1977, he published the Sport Competition Anxiety Test (SCAT) and, with help from colleagues in 1980, the Competitive State Anxiety Inventory (CSAI). Martens’ work was preceded by Spielberger’s theory that anxiety was present in both state and trait forms. Spielberger’s State–Trait Anxiety Inventory (STAI) was published in 1970.

Meanwhile, Robert Nideffer’s strong interest in aikido, as well as his expertise in MMPI research, led him to conclude that attention was the key psychological factor in sport performance. He developed The Attentional and Interpersonal Style (TAIS) Inventory while at the University of Rochester in 1976. His assessment tool directly measured the attentional preferences of athletes and helped practitioners to understand the conditions in which their clients would make concentration mistakes. In a survey taken 22 years later, only one other inventory, the Profile of Mood States (POMS), was being used more than TAIS by applied SPPs. POMS is a general inventory initially created in 1971 for use in psychiatry populations. Its popularity stems from its ability to assess both current moods and dominant mood traits.

The current state of assessment in sport psychology is one of many options. A directory of sport psychology assessments published by Ostrow in 1996 included 314 different inventories, questionnaires, and surveys. Sport psychology clients now include more than just elite

athletes; they include a population of serious recreational athletes and aspiring juniors in many sports. Assessment tools such as Jesse Llobet’s Winning Profile Athlete Inventory (WPAI) have been created to support developing athletes as well as professionals. Many assessments are now available on the Internet, where they can be used by athletes and practitioners from anywhere in the world in the privacy of their own homes. In professional ranks, player development and scouting departments count on psychological assessments and sport psychology consultants to help them in their coaching and drafting efforts. Examples of this application of psychological assessment include the use of AMI by the Major League Baseball Scouting Bureau and the use of Wonderlic by the National Football League (NFL) to assess intelligence levels of draft-eligible college players at the annual scouting combine. In addition, many professional teams rely on information from experts with self-designed assessments such as Robert Troutwine’s Troutwine Athlete Profile (TAP).

Biddle’s *European Perspectives on Exercise and Sport Psychology* provides a greater international perspective on history and current assessment tools.

## 6. LIST AND DESCRIPTION OF FREQUENTLY USED TESTS

This section contains descriptions of frequently used assessments employed in the field of sport psychology. The instruments are presented alphabetically by instrument name. Table III identifies the kinds of factors or variables measured by each assessment.

### 6.1. 16 Personality Factor Questionnaire (16PF)

*Year created:* 1950

*Author:* Cattell

Cattell believed that human personality consisted of 46 surface traits, from which 16 primary traits can be derived. Used primarily for sport psychological research, several studies showed meaningful differences on several scales for athletes and a significant correlation between extraversion and sport success. The 16 dimensions are Warmth (cool vs warm), Intelligence (concrete thinking vs abstract thinking), Emotional Stability (easily upset vs calm), Dominance (not assertive vs dominant), Impulsiveness (sober vs enthusiastic), Conformity (expedient vs conscientious), Boldness (shy vs venturesome),

**TABLE III**  
**Assessment Content Categories**

Assessment <sup>a</sup>	Attention	Cognitive	Personality	Sport-specific	Team/Social interactive
16PF			X		X
ACSI			X	X	
AMI		X	X	X	
Competition Reflections		X		X	X
CSAI-2			X	X	
MBTI			X		
MMPI		X	X		
POMS			X		
PSIS		X	X		X
SCAT			X	X	
STAI			X		
TAIS (AME)	X		X	X	
TAP			X	X	
TOPS	X	X			
WPAI			X	X	X
WPT		X			

<sup>a</sup>16PF, 16 Personality Factor Questionnaire; ACSI, Athletic Coping Skills Inventory; AMI, Athletic Motivation Inventory; CSAI-2, Competitive State Anxiety Inventory; MBTI, Myers–Briggs Type Indicator; MMPI, Minnesota Multiphasic Personality Inventory; POMS, Profile of Mood States; PSIS, Psychological Skills Inventory for Sports; SCAT, Sport Competition Anxiety Test; STAI, State–Trait Anxiety Inventories; TAIS (AME), The Attentional and Interpersonal Style Inventory (Athletes Mental Edge); TAP, Troutwine Athletic Profile; TOPS, Test of Performance Strategies; WPAI, Winning Profile Athletic Inventory; WPT, Wonderlic Personnel Test.

Sensitivity (toughminded vs sensitive), Suspiciousness (trusting vs suspicious), Imagination (practical vs imaginative), Shrewdness (forthright vs shrewd), Insecurity (self-assured vs self-doubting), Radicalism (conservative vs experimenting), Self-Sufficiency (group-oriented vs self-sufficient), Self-Discipline (undisciplined vs self-disciplined), and Tension (relaxed vs tense).

**6.2. Athletic Coping Skills Inventory (ACSI-28)**

*Year created:* 1995

*Authors:* Smith, Schutz, Smoll, & Ptacek

This is a 28-item inventory designed with the intent to understand how athletes deal with competitive stress. ACSI-28 measures seven sport-specific subscales and then provides an overall “power” ranking that is a composite of those scale scores. The power ranking is called the Personal Coping Resources score, and the subscales are Coping with Adversity, Peaking under Pressure, Goal Setting/Mental Preparation, Concentration, Freedom from Worry, Confidence and Achievement Motivation, and Coachability.

**6.3. Athletic Motivation Inventory (AMI)**

*Year created:* 1969

*Authors:* Tutko, Lyon, & Ogilvie

This is a 190-item, forced-choice inventory that is historically the most widely referenced in the field of sport psychology. It was developed specifically to measure traits related to elite athletic success. Traits measured are grouped as Desire factors or Emotional factors. Desire factors are drive, determination, aggression, leadership, and organization. Emotional factors are coachability, emotionality, self-confidence, mental toughness, responsibility, trust, and conscience development.

**6.4. Competition Reflections**

*Year created:* 1986

*Author:* Orlick

This is composed of 12 open-ended items that provide opportunities for athletes to reflect on their “best” and “worst” selves by recalling details about great and poor performances. Included within these items are four

10-point rating scales that provide some reference for perceived arousal and anxiety. Direction of the reflections leads to the athlete brainstorming changes in training, preparing for competition, and interacting with coaches.

### **6.5. Competitive State Anxiety Inventory (CSAI-2)**

*Year created:* 1990

*Authors:* Martens, Burton, Vealey, Bump, & Smith  
Originally born out of Martens' work with SCAT, this is a modification of Spielberger's State Anxiety Inventory (SAI). The original version, created in 1980, isolated 10 items from SAI that measured competitive anxiety. A multidimensional version of the original, CSAI-2, includes three 9-item subscales measuring cognitive state anxiety, somatic state anxiety, and state self-confidence. In applied settings, a "Mental Readiness" form has derived from CSAI-2 that allows athletes to put themselves through a more convenient assessment of their anxiety levels while on the field or court.

### **6.6. Myers–Briggs Type Indicator (MBTI)**

*Year created:* 1962

*Authors:* Myers & Briggs

Based on Jungian personality theory, this is the most widely used personal inventory in the world, with more than 2 million test takers annually. MBTI has a variety of applications in the worlds of personal development and management consulting. Its primary use in sport settings is for team cohesion and increasing awareness between teammates and coaches. Personality preference is measured along four dichotomies: Extraversion/Introversion, Sensing/Intuition, Thinking/Feeling, and Judging/Perceiving. Combinations of these scales produce a four-letter acronym that reflects the dominant score on each factor. These acronyms correspond to 16 different personality types that describe persons in detail according to behaviors associated with their personalities.

### **6.7. Minnesota Multiphasic Personality Inventory (MMPI-2)**

*Year created:* 1942

*Authors:* Hathaway & McKinley

This is one of the cornerstones of psychiatric and clinical psychology assessment. MMPI is used to identify personality and psychosocial disorders as well as

for general cognitive evaluation in adults and adolescents. A number of versions of MMPI exist, including the most recent version, MMPI-2, MMPI-A (an adolescent version), and short and long forms. Although MMPI is widely used in clinical settings, its use is limited in current applied sport psychology settings.

### **6.8. Profile of Mood States (POMS)**

*Year created:* 1971

*Authors:* McNair, Lorr, & Droppleman

This is a 65-item instrument assessing tension, depression, anger, vigor, fatigue, and confusion. A 5-point scale (from "extremely" to "not at all") is used to rate responses. This inventory can be used to measure both state- and traitlike factors, depending on the "set" provided to the test taker by the proctor.

### **6.9. Psychological Skills Inventory for Sports (PSIS)**

*Year created:* 1987

*Authors:* Mahoney, Gabriel, & Perkins

Consisting of 45 items with 5-point Likert response options, this instrument measures Anxiety Control, Concentration, Confidence, Mental Preparation, Motivation, and Team Emphasis. PSIS was created with input from 16 leading sport psychologists who answered the inventory questions as they thought the "ideal" athlete would. The original inventory asked true/false questions and contained 51 items. Test construction involved 713 athletes from 23 different sports.

### **6.10. Sport Competition Anxiety Test (SCAT)**

*Year created:* 1977

*Author:* Martens

This is a 15-item inventory that was created to measure "state" anxiety levels in people anticipating competition. Martens was very clear in the original publication of the assessment that SCAT was not to be used to predict performance outcomes even if there was a relation to anxiety levels. Therefore, functional use of SCAT is optimal when gauging manifestations of anxiety in athletes during pregame or prematch periods.

### **6.11. State–Trait Anxiety Inventories (STAI)**

*Year created:* 1970

*Authors:* Spielberger, Gorsuch, & Lushene

This was originally developed as a research tool for measuring anxiety in adults and has contributed mightily to sport anxiety research in a variety of populations. STAI is composed of two 20-item scales: one measuring State Anxiety and one measuring Trait Anxiety. The context for the A-State section is how the test taker is feeling at the current time, whereas the A-Trait section is based on general feelings of anxiety. Application of STAI has led to many significant findings in athlete work. For example, Yuri Hanin used STAI to create his Zones of Optimal Functioning (ZOF) model.

### **6.12. The Attentional and Interpersonal Style (TAIS) Inventory**

*Year created:* 1976

*Author:* Nideffer

This is a 144-item inventory with an emphasis on concentration and the critical interpersonal factors affecting performance under pressure. The latest version of TAIS is composed of 20 scales that range principally over attentional and interpersonal factors. The two newest scales are Focus over Time and Performance under Pressure. Athletes Mental Edge (AME) used a 50-item version of TAIS and was introduced in 1999. AME was the first widely used, Internet-based sport psychological assessment tool.

### **6.13. Troutwine Athletic Profile (TAP)**

*Year created:* 1988

*Author:* Troutwine

This is a 75-item questionnaire that receives considerable attention prior to the NFL draft. Troutwine does not publish his assessment for distribution, so there are few SPPs using TAP. However, this may be the most well-known, “self-written for self-use” assessment in the field. The questionnaire generates a Player Report for development and a Coach Report for recruiting and selection; this is why many NFL teams privately contract with Troutwine and value his data. The Player Report measures Competitive Nature, Coping Style, Work Style, and Mental Characteristics. The Coach Report topics are Mental Makeup, Social Style, Personal Characteristics, Coaching Suggestions, and Performance Tendencies.

### **6.14. Test of Performance Strategies (TOPS)**

*Year created:* 1999

*Authors:* Thomas, Murphy, & Hardy

This is a 64-item inventory intended for use when measuring psychological skills and strategies used by athletes both in competition and while at practice. TOPS results differentiate between practice and competitive skills and strategies on the following scales: Goal Setting, Use of Imagery, Ability to Relax, Ability to Activate/Energize, Use of Positive Self-Talk, Emotional Control, Attentional Control, Automaticity (i.e., performing without thinking), and Negative Thinking. Although TOPS is a fairly new inventory, its applied use was already seen with U.S. and Australian Olympic teams and coaches in preparation for the Sydney Games in 2000.

### **6.15. Winning Profile Athlete Inventory (WPAI)**

*Year created:* 1999

*Author:* Llobet

This is composed of 55 items with 5-point Likert response options that measure a wide variety of psychological skills in athletes. WPAI responses load onto 10 factors: Competitiveness, Dependability, Commitment, Positive Attitude, Self-Confidence, Planning, Aggressiveness, Team Orientation, Willingness to Sacrifice Injury, and Trust. Functional use is intended for athletes from high school to professional levels for recruiting and development purposes.

### **6.16. Wonderlic Personnel Test (WPT)**

*Year created:* 1937

*Author:* Wonderlic

This is a 50-item test of general intelligence that the user has 12 minutes to complete. Originally designed for use with job applicants in both blue- and white-collar settings, WPT has become a mainstay in the assessment of collegiate football players at the NFL's scouting combine. The test has 16 different versions with problems that include algebra, word analogies, mathematical word problems, and various other tests of verbal, computational, and analytic abilities. It measures an individual's ability to learn, adapt, solve problems, and understand instructions.

## 7. FORMAL ASSESSMENT GUIDELINES AND ETHICAL CONSIDERATIONS

When done well, psychological testing provides professionals with an efficient, objective, and repeatable approach to assessing behavior. Many factors affect the quality of the information gathered through the various assessment means. Sensitivity to issues around test validity, confidentiality, response sets, and response styles is critical for effective use of sport psychology assessment.

### 7.1. Test Validity

It is incumbent on SPPs to evaluate the validity of the assessment methods they use. A test's validity determines how accurately it measures what it purports to measure. There are three commonly referenced types of validity: face, construct, and predictive. Face validity is a surface measure that acts as a first step toward the integrity of an assessment. Face validity is attained when the parameters of a test seem like they would generate the results that SPPs hope to generate. Construct validity links theory with practice. It can be summed up as the ability that an assessment has to demonstrate, with results, the topic it is intended to measure. Predictive validity is simply the likelihood that the results of an assessment can predict the occurrence of a specified behavior in the future.

### 7.2. Confidentiality

According to the code of ethics set forth by the American Psychological Association, the confidentiality of the

psychologist–client relationship is protected. Because many nonlicensed individuals function as sport psychology consultants, the status of such a relationship is less clear and not likely afforded the same kind of protection. When it comes to the confidentiality of assessment results, SPPs should be careful to protect confidentiality explicitly and to explain clearly to clients who will have access to their assessment information. Because it can sometimes be difficult to identify exactly who the client is (e.g., the athlete, the coach, the organization), it is advisable to have a written policy that describes how assessment information will be kept and shared. Many SPPs use informed consent forms to establish clearly the confidentiality of assessment data (Fig. 2).

### 7.3. Response Sets and Styles

When asking athletes to answer any set of questions about themselves, it is important to take into account variables that will influence the accuracy of their responses. Response sets are situation-specific attitudes that individuals adopt when completing a psychological assessment. They are often employed when athletes are skeptical of how their assessment data may be used. Faking can occur in situations where athletes feel pressure to present themselves in an ideal light.

In contrast to response sets, response styles reflect cross-situational, core personality characteristics. Common response styles include overly conservative and extreme appraisals. People who adopt a conservative response style see the world in shades of gray and tend to respond with middle-range answers such as “maybe” and “sometimes” on inventories with Likert

Assessment Consent Form	
<p>I hereby grant permission to _____ for sport psychology assessment. I understand that any information gained from testing will be used for the purpose of performance enhancement and will be kept confidential. I further understand that none of my personal data will be released to anyone without my consent. I agree that although my individual data may be used as part of a database for research, my identity and other personal information will not be disclosed.</p>	
<p>I understand that my access to sport psychology services is not contingent on my granting of this permission. In addition, I understand that if at any time during the assessment phase I feel uncomfortable, I can end the assessment process.</p>	
<p>_____ Signature</p>	<p>_____ Date</p>
<p>_____ Printed name</p>	<p>_____ Sport psychology professional</p>

FIGURE 2 Sample informed consent form.

scale choices. Extreme response styles are often seen in individuals who have either an overly optimistic outlook or an overly critical outlook. Aside from the impact that response sets and styles have on individual scores, SPPs must be careful when comparing different athletes' assessment results. Although some degree of influence from both response sets and response styles is always likely to be present, their effects can be minimized by encouraging open and honest answering and by providing points of comparison for various response choices. Many SPPs prefer not to control for response sets because the sets and styles themselves provide a great deal of information about the athletes.

#### 7.4. Consensual Validation

To ensure that test information is accurate, it is important to have the opportunity to speak directly with the individuals who have completed the assessments to determine the impact of any existing response sets or response styles.

Consensual validation is the process of verifying the interpretations made from test information, which typically occurs through interview or direct observation. Effective consensual validation requires the ability to define operationally psychological constructs (e.g., confidence, control). By translating these concepts into specific behaviors, SPPs can determine their presence or absence more accurately. Imagine that a client completes a personality inventory and scores very high on control and self-esteem. A high score on these dimensions would lead the interpreting SPP to assume certain behaviors in various circumstances. More often than not, the SPP would expect the client to speak up, take charge, be reluctant to listen to contrary opinions, and oppose authority. The SPP would be looking to verify that these behaviors occur with the kind of frequency commensurate with the scores on the performance-relevant construct(s). *Table IV* offers a number of suggestions to increase effective consensual validation.

### 8. FUTURE DIRECTIONS

Just as athletes engage in the "pursuit of perfection," SPPs seek the most effective means of helping the athletes with whom they work. Technology is already having a pronounced effect on how assessment is carried out. Psychological testing can now take place over the Internet with the tester and the testee thousands of miles apart. More and more individuals will gain access to assessment information. Feedback

**TABLE IV**  
**Consensual Validation Techniques**

- 
- Select tests that measure constructs that are easily translatable into observable behaviors.
  - Select tests that measure a number of performance-relevant variables or choose several tests that measure a variety of constructs. The reason for this is that individual constructs are rarely able to explain performance across individuals and situations. A number of distinct cognitive and interpersonal characteristics must be examined to effectively account for the influences that individual and situational differences have on performance.
  - Translate each construct into expected behaviors (for the performance arena in which they operate).
  - Create conditions under which the individual's behavior can be observed under varying degrees of pressure. Remember that most people are able to control even dominant behaviors when they are not experiencing significant stress. The key is to see how much difference increasing the pressure has on the presence or absence of particular behaviors.
- 

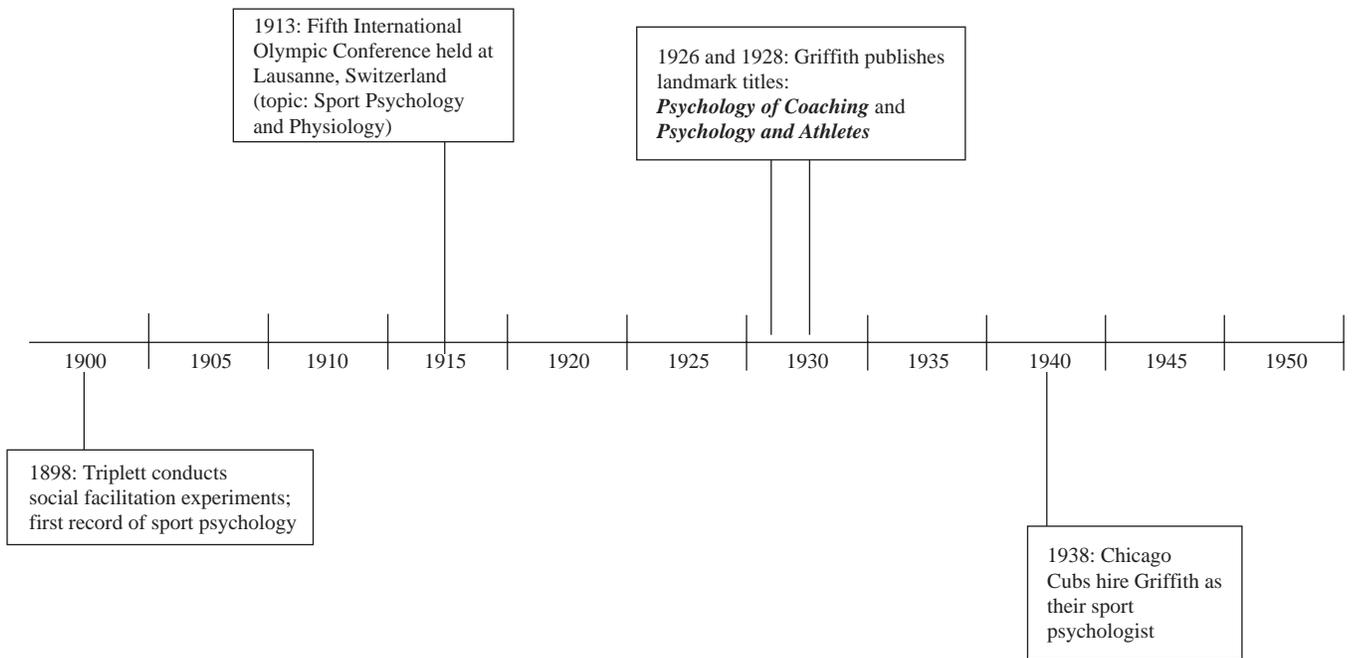
*Source.* Adapted from Nideffer and Sagal (2001).

reports delivered instantaneously, and sometimes without professional interpretation, will most certainly force the field of sport psychology to confront what is and is not acceptable practice.

As athletes are exposed to the benefits of working with SPPs, it is inevitable that more athletes themselves will develop a professional interest in the field. Consequences of this are likely to include the following:

- An increased respect for the field of sport psychology itself
- A greater emphasis on practical applications in specific sports
- Increased pressure on SPPs to have significant sport-specific athletic experience of their own

Although the focus of this article has been on assessment in applied sport psychology, it is important to note that it is nearly always the case that application follows research. Advancements in medicine and new insight into the mind-body connection will likely increase the use of neurological, chemical, and physiological testing to help predict and influence performance. The ongoing search for strong ties between emotions and body chemistry may bring new kinds of SPPs to the field. For example, heart rate monitors and other biofeedback mechanisms are already being used to help athletes achieve optimal levels of arousal and to help SPPs better understand the competitive experience (*Fig. 3*).



**FIGURE 3** Assessment timeline.

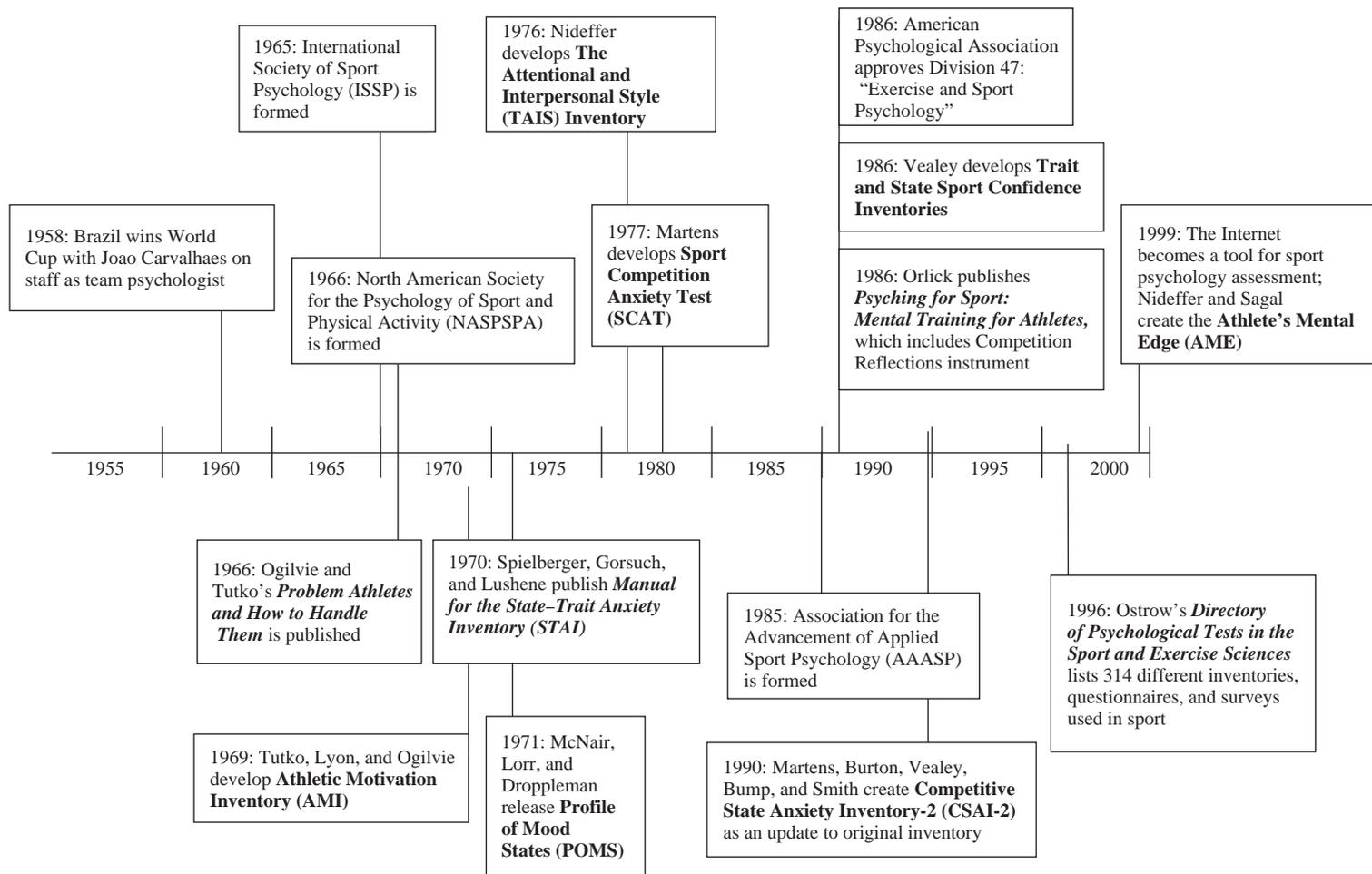


FIGURE 3 (Continued).

### See Also the Following Articles

Anxiety and Optimal Athletic Performance ■ Attention and Concentration Training in Sport ■ Goal Setting and Achievement Motivation in Sport ■ Intrinsic and Extrinsic Motivation in Sport ■ Performance Slumps in Sport: Prevention and Coping ■ Psychological Skills Training in Sport ■ Self-Confidence in Athletes

### Further Reading

Anastasi, A. (1988). *Psychological testing*. New York: Macmillan.

Biddle, S. J. (Ed.). (1995). *European perspectives on exercise and sport psychology*. Champaign, IL: Human Kinetics.

Duda, J. L. (Ed.). (1998). *Advances in sport and exercise psychology measurement*. Morgantown, WV: Fitness Information Technology.

Fisher, A. C. (1984). New directions in sport personality research. In J. M. Silva, & R. Weinberg (Eds.), *Psychological*

*foundations of sport* (pp. 70–80). Champaign, IL: Human Kinetics.

Heil, J., & Henschen, K. (1996). Assessment in sport and exercise psychology. In J. Van Raalte, & B. Brewer (Eds.), *Exploring sport and exercise psychology* (pp. 229–255). Washington, DC: American Psychological Association.

Nideffer, R. M., & Sagal, M. S. (2001). *Assessment in sport psychology*. Morgantown, WV: Fitness Information Technology.

Orlick, T. (1990). *In pursuit of excellence* (2nd ed.). Champaign, IL: Human Kinetics.

Ostrow, A. C. (1996). *Directory of psychological tests in the sport and exercise sciences*. Morgantown, WV: Fitness Information Technology.

Scott, S. (2002). *Fierce conversations*. New York: Penguin Putnam.

Silva, J. (1984). Personality and sport performance: Controversy and challenge. In J. Silva, & R. Weinberg (Eds.), *Psychological foundations of sport* (pp. 59–69), Champaign, IL: Human Kinetics.



# Attachment

Ronit Roth-Hanania and Maayan Davidov

University of Haifa, Haifa, Israel

1. Introduction
2. History
3. Measurement
4. The Antecedents of Attachment Security
5. Attachment Throughout the Life Span: Stability and Change
6. Attachment and Subsequent Child Outcomes
7. Applied Aspects of Attachment  
Further Reading

## GLOSSARY

***anxious–avoidant attachment*** An insecure form of attachment; during the Strange Situation, infants classified as avoidant appear to be relatively indifferent to separation from the caregiver even though physiological markers indicate that they are in fact distressed, and on reunion, they actively ignore and avoid the caregiver, diverting their attention to exploration.

***anxious–resistant (ambivalent) attachment*** An insecure form of attachment; during the Strange Situation, infants with this classification exhibit extreme distress on separation, and when the caregiver returns, they refuse to settle down, clinging while at the same time pushing the caregiver away in anger.

***attachment*** An emotional connection to another individual that develops between an individual who is seen as able to provide protection, comfort, and support in times of need; early attachment develops between an infant and a primary caregiver through patterns of caregiving interactions that evolve over the first year of life.

***disorganized attachment*** A Strange Situation classification given in addition to one of the other three classifications; infants classified as disorganized/disoriented show sequences of behavior that appear to lack a clear intent, purpose, or organization (e.g., freezing).

***internal working models*** Internal mental representations that are established during the formation of an attachment relationship; these mental representations refer to both self and the attachment figure, and later in life they guide a person's processing of relevant information and behavior in the context of interactions and relationships.

***secure attachment*** A positive and adaptive form of attachment between an infant and his or her caregiver; the infant uses the caregiver as a base for independent exploration of the environment and as a source of comfort and support during times of stress and distress.

***Strange Situation procedure*** A systematic laboratory observational method designed to assess the quality of attachment organization of a 12- to 18-month-old infant to a caregiver.

Attachment theory, put forth by John Bowlby, postulates that infants develop strong emotional ties to consistent caregivers during the first year of life and that these ties are crucial for the individual's normal social–emotional development and well-being. Bowlby, a British psychiatrist, combined ideas from psychoanalysis, ethology, and evolution to develop a theory regarding the attachment behavioral system—its nature, function, and development. Mary Ainsworth, Bowlby's collaborator, provided empirical support for many of Bowlby's ideas in her in-depth observations of

mother–infant interactions at home and in a laboratory environment. Ainsworth also developed a laboratory observation procedure, the Strange Situation, which enabled her to identify individual differences in attachment organization among 1-year-old infants. The Strange Situation procedure, currently the most widely used and validated measure of attachment security during infancy, is composed of a series of short separations and reunions between an infant and his or her caregiver, usually the mother. It enables psychologists to classify the quality of an infant's attachment to a caregiver into one of three major categories: secure attachment, anxious–avoidant attachment, or anxious–resistant attachment. A fourth category, disorganized/disoriented attachment, was added more recently. Ainsworth and colleagues also demonstrated that the caregiver's sensitive responsiveness to the infant's signals and cues is important for the formation of secure attachment. This finding has been replicated in many subsequent studies, although not always so strongly. Research on attachment yields some interesting findings regarding the inter-generational transmission of attachment, the stability of attachment throughout the life span, and the role of early attachment in later child outcomes, particularly in the areas of emotion regulation, cognition, social adjustment, and mental health. Because early attachment relationships are viewed as crucial to later development, there is a growing number of prevention and intervention programs designed to promote attachment security among populations at risk for insecure attachment relationships. The effectiveness of these programs has varied. Ever since it was first introduced, attachment theory has played a central role in shaping our understanding of child development and parent–child relations, as reflected in developmental theory and research, clinical and applied settings, and policymaking.

## 1. INTRODUCTION

Attachment theory centers on the notion that emotionally responsive care, including love and nurturance from a primary caregiver, is essential for healthy and normal development. John Bowlby, the founder of attachment theory, used the term “attachment” to describe the emotional connection that develops between an infant and a primary caregiver through patterns of interaction that evolve over time. According to attachment theory, during the first year of life, infants develop special ties to their primary caregivers (i.e., attachment figures). Every infant invariably becomes

attached unless no single caregiver is continuously present to care for the child (e.g., as in some institutions and orphanages). The particular nature or quality of the attachment relationship will vary, depending on the history of interaction patterns within the dyad. In particular, the consistency and appropriateness of the caregiver's response to the infant during times when the infant feels stressed or threatened, defines the pattern of interactions that develop during the first year. When the caregiver is consistently sensitive and responsive, a secure attachment typically ensues, whereas insensitive or inappropriate responding usually leads to an insecure attachment relationship.

After the attachment is formed, the infant shows a preference for the attachment figure. This is particularly evident during times of stress or emotional upset. When scared or distressed, the infant actively seeks comfort and reassurance from the caregiver and is not easily comforted by other individuals. The presence of the attachment figure is reassuring and enables more competent exploration of the environment. Mary Ainsworth, Bowlby's collaborator, noted that in a secure attachment relationship, the attachment figure provides both a “safe haven” during times of need and a “secure base” from which to explore.

Bowlby viewed the attachment system as a biobehavioral system that organizes behavior. The set goal of this system is to maintain proper access and proximity to the caregiver during times of need to ensure the protection and survival of the young. The attachment system operates in interaction and in collaboration with other behavioral systems (e.g., fear, exploration, sociability).

Attachment relationships are believed to play an important role throughout the life cycle. Although the specific behavioral markers of attachment change, their function remains the same. During times of need, infants, older children, and adults alike all seek comfort and support from significant others who are seen as more capable of coping with these situations. As individuals mature, new attachments are typically formed, but the early attachment relationships remain important because they are thought to exert an effect on subsequent behavior and close relationships.

## 2. HISTORY

### 2.1. John Bowlby (1907–1990)

Bowlby, a British child psychiatrist and psychoanalyst by training, was one of the first to systematically

observe the negative effect of impoverished family life and the lack of consistent care on the development of young children. He suggested, based on his work with hospitalized and homeless children during the 1940s, that a child needs to experience a close, warm, and consistent relationship with his or her mother or another permanent caregiver so as to develop normally.

Influenced by ethology and evolutionary theory, Bowlby rejected the then widely accepted notion that an infant's tie to his or her mother is rooted in the mother's role as a provider of food or oral gratification. Lorenz's observations of the "imprinting" phenomenon in geese during the 1930s demonstrated that an organism can become "attached" to its caregiver through a process unrelated to feeding. Harlow's laboratory studies of rhesus monkeys during the 1950s and 1960s also demonstrated that needs other than hunger are important in the formation of the infant-mother tie. For example, monkeys "raised" by two kinds of surrogate mothers—a wire mesh "mother" to which a bottle with milk was attached and a terrycloth surrogate "mother" with no milk bottle—preferred the soft cloth mother, spending most of their time clinging to her for comfort and rushing to her when frightened. They spent little time on the wire mother, approaching her only when wanting to be fed. Influenced by these and other studies, Bowlby posited that the need for proximity to a caregiver is a biologically based primary need that evolved to ensure the protection and survival of the young.

Bowlby also argued that it is children's actual experiences with caregivers (e.g., lengthy separations, lack of consistent care) that determine their well-being and adjustment. This view was at odds with the notions accepted by the psychoanalytic community, which placed greater emphasis on children's fantasy lives than on their actual experiences with caregivers.

These disagreements led Bowlby to establish his own research unit at the Tavistock Clinic in London, where he proceeded to develop his groundbreaking ideas. His work is summarized in the trilogy titled *Attachment and Loss*, published in 1969, 1973, and 1980, as well as in many other books and articles. The focus in Bowlby's writings was primarily on the evolutionary origin and function of the attachment system, how it organizes behavior, its developmental course, and the effects of early separations and loss of attachment figures on a person's well-being and adjustment. The work of Ainsworth provided empirical grounding for Bowlby's theoretical thinking, extended Bowlby's ideas on attachment, and brought to the forefront a focus on different patterns of attachment organization.

## 2.2. Mary Ainsworth (1913–1999)

Ainsworth met Bowlby during the early 1950s at the Tavistock Clinic. She subsequently carried out an observational study in Uganda designed to examine Bowlby's ideas empirically. After returning to the United States, she proceeded to conduct a second attachment study in Baltimore, Maryland, involving intensive home observations of 23 mother-infant dyads. As part of the Baltimore study, Ainsworth also developed a systematic laboratory observational method that enabled her to observe the reactions of infants to short sequences of separations from their mothers. This procedure, known as the Strange Situation, allowed her to identify individual differences in infants' attachment to their mothers, that is, their different strategies of using the mother when dealing with the stress brought about by brief separations. The Strange Situation procedure initially yielded three types or classifications of attachment: a secure pattern and two insecure classifications (anxious-avoidant and anxious-resistant attachment). The procedures and findings of the Baltimore study are summarized in the book *Patterns of Attachment*, published in 1978. Ainsworth's work generated a surge of empirical studies that employed the Strange Situation and focused on individual differences in attachment security.

## 3. MEASUREMENT

Over the years, researchers have developed a variety of measures to assess attachment patterns at various ages. The most widely used and best validated assessment procedures employed with infants, children, adolescents, and adults are described in what follows.

### 3.1. Assessing Infant Attachment

#### 3.1.1. The Strange Situation

**3.1.1.1. Procedure** For more than 30 years, quality of parent-infant attachment has been assessed using the Strange Situation procedure developed by Ainsworth in the Baltimore study. This procedure, which is appropriate to use with infants ages 12 to 18 months, involves a series of moderately stressful separations (3 minutes each) and reunions (3 minutes each) between the infant and a caregiver while in an unfamiliar playroom. Two systems are activated during the Strange Situation procedure: exploration (of a new environment, toys, and people) and attachment (through the repeated separations

and reunions). The Strange Situation procedure enables researchers to examine how the infant uses the caregiver to cope with stress. The infant's strategy is thought to reflect his or her accumulated experience as it pertains to the caregiver's availability during times of need.

**3.1.1.2. Attachment Patterns** Ainsworth identified three major attachment classifications. Secure infants use their caregivers as a "secure base" for exploration. These infants rely on occasional visual, verbal, or physical contact with their mothers as a basis for their own initiated exploration of the environment. When their mothers leave the room, secure infants may or may not cry. On reunion, these infants either greet their mothers positively or, if upset, go to them for comfort and then shortly after return to activities associated with exploration.

Infants who show patterns of insecure attachment of the avoidant type exhibit little or no reference to their mothers while exploring the room and show no overt signs of distress on their mothers' departure (although some studies have shown that these infants do show physiological markers of stress during separation). Most important, on reunion, these infants actively ignore and avoid their caregivers (e.g., by looking away).

Insecure-resistant (or ambivalent) infants are preoccupied with their mothers' presence, often unable to leave their sides even in light of curious attempts to explore their new environment. When their mothers leave the room, these infants become extremely distressed, and on reunion, the infants refuse to settle and to resume exploration, clinging to their mothers and at the same time expressing anger and dissatisfaction.

The large body of work using the Strange Situation shows that across cultures, the majority of all infants (approximately 65%) are securely attached to their mothers, whereas the remainder are classified as insecure. The distribution of insecure attachment classifications (i.e., frequencies of the avoidant and resistant patterns) appears to be affected by cultural context. In Western Europe and the United States, the avoidant type is more prevalent than the resistant pattern, whereas the opposite has been observed in some other cultures (e.g., in Israel).

A fourth category of attachment classification was added in 1986 by Main and Solomon because studies showed that there are infants whose behavior does not correspond with any of the three existing classifications. During the Strange Situation, these infants showed sequences of unusual behavior with no clear purpose or orientation (e.g., freezing, falling prone).

They seemed to lack an organized strategy for coping with the stress of the situation and looked "confused"; thus, they were classified as having disorganized/disoriented attachment to their caregivers. Disorganized classification is given in addition to one of the other classifications. In normative middle-class populations in North America, approximately 14% of infants are judged to be disorganized. The rate is somewhat higher (24%) in low-socioeconomic status samples.

**3.1.1.3. Validity** Evidence supporting the validity of the Strange Situation (i.e., its capacity to indeed capture the organization of the child's attachment system) comes from the correspondence between Strange Situation classification and Ainsworth's home observations of the same 23 infant-mother pairs. Ainsworth and colleagues found a strong association between infants' Strange Situation classification and their secure base behavior in the home. Secure infants were much more likely than insecure infants to exhibit a good balance of exploratory and proximity-seeking behavior, to transition smoothly between the two, and to enjoy and be comforted by physical contact. In contrast, insecurely attached infants showed disturbed balance between attachment and exploration behavior, had difficulty in transitioning from one to the other, and often expressed negative affect during physical contact with their mothers. Further supporting evidence comes from several studies that found links between the Strange Situation classification and infants' secure base behavior in the home as assessed by the Attachment Q-sort. Finally, additional less direct evidence for the validity of the Strange Situation comes from its links to prior maternal behavior and subsequent child outcomes.

### **3.1.2. The Attachment Q-Sort**

An alternative to the Strange Situation is the Attachment Q-sort developed by Waters and Deane in 1985. The AQS requires observers to describe a target child's home behavior by sorting a set of items according to how well they each describe the particular child (from most characteristic to least characteristic). Version 3.0 of the AQS consists of 90 items depicting a wide array of child home behaviors. A continuous measure of security is derived by comparing the sort describing the child of interest with a criterion sort developed to describe the ideal secure child. An advantage of the AQS is that it is suitable for use with a relatively wide age range (from 1

to 4 years). Both mothers and independent observers may use the AQS, although mothers have to become familiar with its items and observe their children for a few days in the home environment prior to their finally sorting all of the items from most characteristic to least characteristic of their children.

Findings regarding the association between the Strange Situation classifications and AQS security score have been mixed; some studies have found correspondence, whereas others have not. Overall, it appears that correspondence is better when independent observers, rather than mothers, are used. Summarizing more than 130 studies, a recent meta-analysis showed moderate convergence between observer-reported AQS security and Strange Situation security. Observer reports of AQS security were also found to be quite strongly linked to maternal sensitivity. This suggests that the AQS, when rated by observers, is a reasonable alternative to the Strange Situation. On the other hand, maternal-reported AQS security had weaker associations with both Strange Situation security and maternal sensitivity assessments, suggesting that mother-reported AQS scores are less adequate at assessing attachment security.

### 3.2. Measuring Attachment During Childhood

Beyond infancy, measurement of the attachment relationship relies heavily on observations of children's behavior on reunion following relatively long separations from parents (30–60 minutes). Children's emotional expression during separation and reunion, as well as their patterns of verbal and nonverbal communication with their caregivers, is examined and used to determine their attachment classification. The three main attachment patterns identified in children parallel those identified in the Strange Situation. A child can be described as secure (also referred to as balanced), as avoidant (also termed defended), or as dependent (parallel to the resistant classification during infancy and also referred to as coercive or involving). Children classified as disorganized during infancy typically exhibit controlling role-reversing behavior on reunion during childhood, taking either a caregiving or a punitive role toward their parents.

Narrative measures attempting to assess children's symbolic representations of attachment have also been developed. These measures use projective techniques. They ask children to respond to pictures or story-stems depicting relevant attachment scenes. The narratives constructed by the children are then coded and scored for attachment

style. Finally, the AQS can also be used with young children to obtain a measure of attachment security.

### 3.3. Assessing Adult Attachment Representations

During adolescence and adulthood, attachment is most often assessed through structured interviews or questionnaires. The central instrument is the Adult Attachment Interview (AAI) developed by Main and colleagues. Throughout the interview, the individual is asked to describe his or her relationships with caregivers during childhood and to illustrate these descriptions with specific examples. The individual is also asked to describe situations where he or she needed support from those attachment figures (e.g., when the individual was hurt or upset) as well as relevant experiences of loss or trauma. The interviews must be transcribed verbatim and scored by a trained coder.

The AAI is designed to assess the individual's current state of mind with respect to attachment rather than what his or her early attachment classification might have been. Thus, scoring of the AAI is based on the coherence of the narrative produced by the individual during the interview rather than on an assessment of his or her early life experiences.

The AAI yields adult attachment classifications that mirror the patterns identified in the Strange Situation. Securely attached adults are referred to as "autonomous" or "free to evaluate" because they appreciate close relationships and value their effect. They are reflective and open, and they speak about early experiences in a coherent and objective manner, even when such experiences were difficult or painful. Insecure adults could be classified as either "dismissing" (parallel to avoidant during infancy) or "preoccupied" (parallel to resistant attachment during infancy). Dismissing individuals minimize the effect of meaningful relationships in their personal history. They may idealize or derogate their attachment figures. When asked to support their descriptions of their relationships with specific examples, they are often unable to remember relevant incidents. Preoccupied (or "enmeshed") individuals are unable to detach themselves from irrelevant details associated with their past history and are so wrapped up in their "old" experiences and feelings that they fail to provide an overview. They appear to still be immersed in anger over old issues and unable to "let go."

Finally, the coding of the AAI also yields a score of the individual's unresolved loss or trauma (when such painful experiences had occurred). This score is based on indications of disorganization and disorientation in thinking and affect during the discussion of the loss/trauma (e.g., speaking of a dead person as if he or she were alive, going into an extreme amount of detail when describing the events surrounding the loss).

AAI classifications have been found to be stable over time and unrelated to intelligence, general memory skills, or the individual's narrative style about topics irrelevant to attachment. Some empirical evidence regarding the validity of the AAI comes from numerous studies that found links between mothers' AAI classification and the quality of their caregiving behavior to their own children at home as well as their own children's attachment classification. More direct evidence that the AAI indeed reflects how adults interact with attachment figures when coping with stress is still relatively scarce. In one relevant study, engaged couples who completed the AAI were observed during a standard marital interaction task. AAI security was associated with more adaptive use of the partner as a secure base to obtain support as well as with greater propensity to serve as a secure base for the partner and provide such support. More research linking the AAI to adult secure base behavior is still needed.

Other interview procedures have also been developed to assess adult attachment. For example, the Current Relationship Interview (CRI) was developed by Crowell and Owens to assess attachment within current close relationships. Finally, self-report questionnaires are used as well for the assessment of attachment styles, particularly in the context of adult romantic relationships. Of these, the most widely used are Hazan and Shaver's Attachment Style Questionnaire and Bartholomew and Horowitz's Relationship Questionnaire. Overall, studies examining the associations between questionnaire measures and the AAI have produced mixed findings; several studies find little or no relationship, whereas a few studies report moderate associations. More research is still needed concerning the relationship among various measures of adult attachment.

#### **4. THE ANTECEDENTS OF ATTACHMENT SECURITY**

Because attachment describes an aspect of the relationship that develops between an infant and a caregiver,

both members of the dyad influence the nature and quality of the attachment relationship that emerges. These two classes of antecedents are reviewed in what follows. Genetic influences on attachment are also considered.

### **4.1. Caregiver Influences on Attachment Security**

#### **4.1.1. Parental Caregiving Behavior**

In Ainsworth's studies, infant attachment classification was found to be strongly linked to prior maternal behavior in the home. The extensive home observations revealed that the mothers of secure, avoidant, and resistant infants differed qualitatively in their styles of caregiving. Specifically, mothers of infants later found to be secure during the Strange Situation at 12 months were highly sensitive, responsive, and accepting. They responded consistently and appropriately to their infants' signals (e.g., crying) and were affectionate when holding and taking care of their infants. In contrast, the mothers of avoidant infants were observed to be consistently insensitive and rejecting. They were interfering and showed an aversion to close physical contact with their infants as well as a relative lack of emotional expression. Finally, infants classified as resistant had mothers who were inconsistently responsive. They were generally insensitive to their infants' signals but were not highly rejecting. They were inept at holding their infants and responding to signals, but they showed no aversion to close physical contact with their infants. Therefore, the strategies employed by the insecure infants during the Strange Situation were seen as adaptations to the nonoptimal caregiving they have received and come to expect.

Overall, Ainsworth and colleagues found a very strong positive association between observer ratings of maternal sensitive responding and Strange Situation security. A considerable body of subsequent work employing the Strange Situation has generally replicated this link between early maternal sensitive and responsive caregiving and secure attachment in the Strange Situation, although the effects observed in those replication studies were considerably weaker than those reported by Ainsworth and colleagues. Three separate meta-analytic studies summarizing this body of work all concluded that the replication studies, taken together, showed a moderate linkage between maternal sensitive responding and infant security. Therefore, it appears that maternal sensitive and

responsive caregiving plays an important, but not an exclusive, role in the formation of infant attachment security as assessed by the Strange Situation.

In addition to attachment security, the antecedents of attachment disorganization have been the subject of much theorizing and research. Hesse and Main posited that disorganized attachment is the result of frightening or frightened behavior on the part of the caregiver. The underlying assumption is that these circumstances present a very difficult situation for the infant because the same person who is supposed to help alleviate the child's fear is also the person who causes it or is also afraid and unable to help the child. This leads to an approach–avoidance conflict and to behavioral disorganization. Support for this proposition is found in studies of children who are maltreated. These children show significantly higher rates of disorganized attachment than do children who were not abused or neglected.

#### **4.1.2. Parental Attachment Representations**

Another variable that has been systematically linked to infant attachment security in the Strange Situation is maternal attachment security as assessed by the AAI. Thus, secure autonomous mothers typically have secure infants, dismissing mothers tend to have avoidant babies, and preoccupied mothers often have resistant infants. Finally, maternal unresolved loss or trauma as assessed by the AAI has been found to predict disorganized/disoriented infant attachment. This correspondence in attachment patterns, referred to as the “intergenerational transmission of attachment,” has been found across many studies, even when the AAI was assessed prior to the infants' births. One study involving grandmothers as well as mothers and infants also found transmission of attachment across three generations. The correspondence between parental and child attachment is greater for mothers than for fathers, possibly because mothers typically serve as the primary caregivers.

How is attachment security passed on from parent to child? The main process suggested by attachment theory is through the quality of caregiving provided by the parent. Indeed, research has shown that a substantial part of the correspondence between parent AAI and infant Strange Situation classification (approximately 23–25%) is due to the level of maternal sensitive responsive caregiving. However, a large portion (75–77%) of the correspondence is not explained by quality of caregiving, at least as assessed in the relevant studies

(this discrepancy has been labeled the “transmission gap”). Thus, additional processes must play a role in the transmission of attachment security from generation to generation. Some of these processes may involve genetic transmission of attachment or child influences on attachment security.

## **4.2. Genetic Transmission of Attachment**

Beyond caregiver sensitive responding, another mechanism that might underlie the transmission of attachment from parent to child is through their shared genes. Several twin studies have explored the potential genetic component of attachment development. Overall, these studies show a negligible role for genes, and a paramount role for the environment, in the formation of attachment security during infancy. However, it is possible that genes might be implicated in attachment security beyond infancy; one twin study with preschoolers found evidence for a small genetic effect. With respect to attachment disorganization, the pattern of findings has been inconsistent. Some studies (including a small-scale molecular genetics study) found evidence for heritability, whereas a recent twin study found no evidence for genetic influence on attachment disorganization. Further research is needed to elucidate the role played by nature on the formation and maintenance of attachments throughout the life cycle.

## **4.3. Child Influences on Attachment Security**

### **4.3.1. Child Temperament**

Although attachment researchers have traditionally focused on caregiver contributions to attachment security, infants themselves also play an important role in the formation of parent–child interaction patterns. Recall that attachment to caregivers is established during the first year of life, but long before then infants already exhibit their individuality in their affect and behavior, referred to as temperament. Many different aspects and components of early temperament have been identified, and their effects on the nature of parent–child interactions have been examined. Some aspects of temperament seem particularly relevant to infants' expression and regulation of attachment behavior, for example, proneness to distress (e.g., difficulty, emotionality, negative reactivity) and ability

to regulate distress effectively. Therefore, the potential influence of these temperamental features on the development of attachment security has been the subject of much debate and investigation.

In the strongest form of critique, some temperament researchers have argued that the Strange Situation reflects infants' temperamental characteristics rather than their accumulated experience regarding caregivers' responsiveness to their expressed needs. However, this analysis seems too biased in the direction of temperament effects and overlooks some important evidence. It appears that attachment classifications cannot be reduced to temperamental characteristics.

Today, it is generally acknowledged that both parent and child characteristics play a role in the formation of a secure attachment. In general, small to moderate associations between aspects of temperament and attachment security have been found, particularly when the AQS, rather than the Strange Situation, was used to assess attachment security. These studies show that infants who are more prone to distress early in life (e.g., difficult, reactive) generally show less secure base behavior later on.

However, beyond ascertaining the relative contribution of parent and child effects, it seems that the greater challenge for researchers during the upcoming years is to uncover how the two classes of antecedents interact in the formation of attachment. It seems likely that infants' temperament affects parents' ability to provide sensitive and responsive caregiving. Keeping highly irritable infants soothed and content is more demanding than caring for "easy" babies. Thus, infants prone to distress may elicit a lower level of sensitive responding, compared with calmer infants, from the same parents. Moreover, aspects of the environment (e.g., SES, family constellation, culture, support systems) also exert their effects on the two partners and the dyadic process. For example, parents who are exposed to high levels of stress or have little social support may be less able to provide responsive caregiving, particularly when caring for highly irritable infants.

#### **4.3.2. Attachment in Children with Special Needs**

In addition to temperament, other infant characteristics can influence the development of attachment security. Of special interest to researchers is the potential effects of infants' special needs (e.g., a serious medical condition, premature birth, developmental delay) on the formation of attachment security. It has been speculated that infants' medical or developmental problems (and parental perceptions of these problems)

might disrupt parent–infant interactions and thus hamper the development of secure attachments. However, studies have generally found that this is not the case. A meta-analysis summarizing this work showed that the majority of medically compromised infants were securely attached to their mothers. Yet, some conditions can delay the time course of the formation of early attachments. For example, studies on infants with Down's syndrome generally show that the emergence of early attachment is slower in this group.

### **5. ATTACHMENT THROUGHOUT THE LIFE SPAN: STABILITY AND CHANGE**

Attachment theory posits that early attachment relationships serve, to a large extent, as a model for later close relationships. According to Bowlby, based on the early history of interactions with a caregiver, an infant develops internal representations or implicit ideas of both self and the attachment figure (e.g., a view of the self as worthy of love and of the mother as available and responsive). These internal representations, termed "internal working models," guide the individual's processing of relevant information and thus affect how the person approaches new situations. It is through these internal working models that early attachments are thought to influence later development and relationships. Because internal working models are believed to operate largely outside the realm of consciousness, they are considered quite resistant to change. However, such change is possible as a result of significant life events and new relationships. Thus, attachment theory hypothesizes that (a) there should be some degree of stability in individuals' attachment representations or internal working models throughout the life span and (b) change in the nature of these internal representations is not random but rather linked to important life events (negative or positive) or to new meaningful relationships (dysfunctional or constructive).

Findings from a number of independent long-term longitudinal studies generally reflect this complex mixture of stability and change. In these studies, attachment security was assessed during infancy using the Strange Situation and again during late adolescence or early adulthood using the AAI. Some of these studies found considerable stability in individuals' security status, whereas others found no such continuity. Importantly, stability of attachment appears to be linked to the

characteristics of the samples involved. Thus, stability of attachment from infancy to adulthood is more likely to be found in individuals brought up in stable and low-stressed families than for those raised in highly stressed and unstable family environments. Moreover, across various samples, change from early security to later insecurity was not arbitrary. Rather, such change was often linked to negative life events that likely affected the quality of caregiving experienced by the child (e.g., loss of a parent, divorce, serious physical or mental illness).

## 6. ATTACHMENT AND SUBSEQUENT CHILD OUTCOMES

With the gradual unfolding of attachment theory, and particularly since the Strange Situation was established as a valid measure of attachment during infancy, researchers began to examine the implications of early attachment relationships for later child outcomes. The underlying assumption in the majority of these studies is that positive attachment relationships (usually classified as secure) would be associated with positive outcomes and that negative attachment relationships (usually classified as insecure) would be associated with less positive or even negative outcomes. In other words, researchers hypothesized, and findings generally support the notion, that secure infants would subsequently show better adaptive skills in the areas of cognitive, emotional, and social development, whereas insecure infants' development may be associated with maladaptive characteristics in various domains.

### 6.1. Emotion Regulation

Emotional regulation refers to an individual's ability to manage and cope successfully with various levels of emotional arousal, including negative emotions, in response to external events and internal stimulation (e.g., anxiety, excitement). Current views suggest that the attachment system itself provides a basic mechanism for emotional regulation. Studies show that during the years when the attachment relationship is formed, children who develop secure attachment ties to their caregivers express a variety of emotions freely, whereas children who form insecure-avoidant attachment ties tend to inhibit emotions. Overall, studies following children from infancy to early childhood have found that secure children are spontaneous in their expression of a variety of emotions, read the emotional cues of

others better, and generally tend to be more emotionally positive. Insecure children, particularly those with an avoidant classification, tend to exhibit minimal emotional expressiveness overall, and particularly restrain the expression of negative emotions. Also, these children tend to misperceive others' emotional displays, and to exhibit inappropriate affect considering the emotional climate of the circumstances (e.g., show anger when expressing positive emotions would be more appropriate).

### 6.2. Prosocial Behavior and Interpersonal Relationships

Longitudinal studies have found that a history of responsive care resulting in secure attachment is associated with a number of positive social developmental outcomes. Yet it is difficult to draw certain conclusions as to the role of early attachment security in later social competence because the parent-child relationship involves many aspects in addition to attachment. In general, findings suggest that children who were classified as securely attached during the first year of their lives were judged during the preschool years to be more socially competent and empathic toward their peers, as well as to have higher self-esteem, than children with a history of insecure attachment. During middle childhood, children who were classified as secure in their early attachment relationships were also found to be more socially accepted by their peers and more adept at forming close friendships than children who were classified as insecure. Moreover, secure infants grow to be adolescents who are more capable of creating intimacy than do insecure infants. However, the correlational nature of this evidence precludes any firm conclusions regarding the causal role played by attachment security. The early attachment relationship is only one component within a larger developmental context predicting later competence in interpersonal domains.

### 6.3. Cognitive Development

Studies have found relations between classification of attachment during infancy and specific cognitive functions during childhood. The most consistent evidence indicates that secure infants grow to have some cognitive advantages over insecure infants. Caution should be exercised in interpreting results regarding the relations between early attachment classification and later cognitive performance given that findings generally

refer to the social aspects of cognition (social cognition), particularly problem solving, reasoning, and information processing within social contexts. However, there is some evidence suggesting that secure infants grow to have better cognitive functioning in the areas of attention, memory, and generalization during childhood than do insecure infants.

#### **6.4. Problem Behavior and Mental Health**

Research has shown that children with an insecure attachment history are more likely to develop behavioral and emotional problems than are children who had secure attachment relationships during infancy. Children with an insecure-avoidant history were found to exhibit more negative “acting out” behaviors, as well as to suffer from more depressed moods and anxiety during the preschool and early school years, compared with children who were securely attached, although the magnitude of these effects was generally small. Moreover, an avoidant-disorganized attachment classification during infancy was found to predict hostile and aggressive behavior during the school years. As with the other domains of development, no research provides evidence that insecure attachment is the primary factor, or even one of the more important factors, underlying any psychological disturbance.

It has also been argued that early insecure attachment contributes to psychopathology during adolescence and adulthood. Most notably, disorganized attachment during infancy has been linked to greater psychopathology in general during late adolescence, as well as to a specific type of psychological disturbance, namely dissociative symptoms. Infants identified as disorganized during the Strange Situation were found to report more dissociative states and experiences during late adolescence, reflecting disturbances of memory, identity, awareness, and cognition (e.g., lack of memory for significant past events). In addition, studies examining adult attachment have found a high proportion of insecure AAI classifications among psychiatric patients, although it might be that the psychiatric problems are affecting these individuals’ responses during the AAI (rather than these individuals’ state of mind with respect to attachment, or their attachment history, being the cause of the disturbance). A recent review suggested that dismissive adult attachment style is associated with acting out or externalizing disorders, whereas preoccupied attachment is associated with depression and anxiety or internalizing disorders. More research is still needed to

elucidate the role played by early attachment in psychological disturbance later in life.

### **7. APPLIED ASPECTS OF ATTACHMENT**

Since attachment theory was first articulated by Bowlby during the 1950s and 1960s, it has had a substantial impact on how early parent-child relationships are viewed. This, in turn, has shaped the thinking and practices of clinicians and policymakers. These influences are presented in this final section.

#### **7.1. Interventions with Infants and Young Children**

Although Bowlby’s ideas, when first introduced, were met with strong resistance on the part of the psychoanalytic community, over the years practitioners have come to incorporate many of these notions into their clinical work. Specifically, because early secure attachment is associated with positive child outcomes, interventions that focus on the enhancement of attachment security in populations at risk for developing insecure attachment have been developed. There are two main approaches that guide these intervention attempts. The first aims at improving the quality of parental caregiving behavior, particularly parental sensitivity. The second approach is typically more long term and focuses on altering the parent’s insecure mental representations or implicit ideas of attachment into more adaptive secure ones. Of course, the two approaches are not mutually exclusive, and both can be incorporated into the same intervention.

##### **7.1.1. Enhancing Parental Sensitivity**

Improvement of the quality of caregiving behavior is typically done by first observing how the mother interacts with her infant. Many programs, as well as individual clinicians, that practice mother-infant psychotherapy use videotaping of mother-infant interactions to learn through observation about the strengths and weaknesses of each mother’s caregiving style. Portions of the videotaped material are then shown to the parent to identify moments where she interpreted her child’s signals correctly and responded appropriately as well as moments where she ignored or misinterpreted her child’s cues. More general patterns of parental responding to the infant are also identified and discussed. Through this process,

the clinician improves the parent's ability to read the child's various cues and signals and to respond more sensitively and appropriately to them. Examples of the use of this process in interventions are the STEEP (Steps Toward Effective Enjoyable Parenting) program developed at the University of Minnesota, the Infant Parent Program at the University of Michigan, and a preventive intervention study conducted in The Netherlands by van den Boom. In addition, similar principles of enhancing responsive caregiving through the provision of information, feedback, modeling, and support often guide individual clinicians when working with mother–infant dyads.

### **7.1.2. Changing Internal Working Models**

Interventions and therapies that attempt to modify insecure internal working models of attachment typically involve parents discussing and reflecting on their experiences with their own early attachment figures (i.e., their parents). The goal is to gain insight into the effects of these early (often negative) childhood experiences on the parents' current caregiving practices. The underlying assumption is that working through these early anxiety-provoking experiences will lead to the adoption of more secure internal working models of attachment, thereby preventing the replication of the same maladaptive interaction patterns with their own children. Note that although this approach focuses on the parents' internal representational level, parenting behavior is also addressed, but without the intense behavioral focus involved in the first approach to intervention discussed in the previous subsection.

Overall, a meta-analytic review of intervention programs using attachment theory in individual, dyadic, or group settings suggested that short-term, behaviorally based interventions with a specific focus were more effective in improving mother–infant quality of interaction and attachment security than were interventions focused on parental internal representations. However, the magnitude of the effects was generally small. Moreover, enhancing maternal sensitivity was found to be a more consistent outcome of interventions than was the effect on infant attachment classification.

## **7.2. Policy**

Attachment theory has had an impact not only on the work of researchers and individual clinical practitioners

but also on policymakers. Early on in his career, Bowlby and his colleague, John Robertson, observed and documented the emotional difficulties that hospitalized children exhibited when forced to separate from their parents. At the time, parents were not allowed to stay with their sick children, and children were forced to spend days and weeks with only brief, highly controlled visits from their parents. A filmed documentation of one hospitalized child's grief and pain over her separation from her parents had a tremendous impact on hospitals' practices and policies. Today, in most hospitals, parents are allowed and encouraged to spend unlimited time with their hospitalized children.

Child care policies were also largely affected by Bowlby and his followers' views on the importance of high-quality caregiving. One important area is the care provided for orphans and children who need to be removed from their parents' care. In the past, these children were typically put in institutions such as residential nurseries and orphanages. This practice was criticized by attachment theorists because the emotionally detached style that characterized care in these institutions did not enable the children to form meaningful relationships with consistent caregivers. As a response to the criticism of institutionalization, there has been a growing tendency to rely on foster care in an attempt to promote more intimate personalized caregiving within a family context. However, this alternative often appears to be less successful than expected because many children are passed on from one foster family to another for short amounts of time.

Finally, another area in which attachment theory has had an influence is how nonmaternal care during the day, and particularly group day care for infants and young children, is perceived and evaluated. In general, research shows that nonmaternal care, including group day care, does not, in and of itself, disrupt children's attachment security to their parents. However, low-quality day care—marked by large group size, high caregiver–child ratios, high turnover of caregivers, and lack of personalized caregiving—can be detrimental to infant–mother attachment security, particularly when other risk factors are present. This highlights the importance of providing infants with good quality day care settings to enable them to form meaningful relationships with consistent caregivers.

## **See Also the Following Articles**

Child Development and Culture ■ Emotion ■ Prosocial Behavior, Development of

**Further Reading**

- Atkinson, L., & Zucker, K. (Eds.). (1997). *Attachment and psychopathology*. New York: Guilford Press.
- Bowlby, J. (1988). *A secure base*. New York: Basic Books.
- Bretherton, I., & Waters, E. (Eds.). (1985). *Growing points of attachment theory and research*. *Monographs of the Society for Research in Child Development*, 50(1-2, Serial No. 209).
- Cassidy, J., & Shaver, P. R. (Eds.). (1999). *Handbook of attachment: Theory, research, and clinical applications*. New York: Guilford.
- Goldberg, S. (2000). *Attachment and development*. London: Arnold.
- Goldberg, S., Muir, R., & Kerr, J. (Eds.). (1995). *Attachment theory: Social, developmental, and clinical perspectives*. Hillsdale, NJ: Analytic Press.
- Greenberg, M. T., Cicchetti, D., & Cummings, E. M. (Eds.). (1990). *Attachment in the preschool years*. Chicago: University of Chicago Press.
- Parks, C. M., Stevenson-Hinde, J., & Marris, P. (Eds.). (1991). *Attachment across the life cycle*. London: Routledge.



# Attention

Amir Raz

Columbia University and New York State Psychiatric Institute, New York, USA

1. Historical Context
  2. Gross Characteristics of Attention
  3. Attention and Perception
  4. Atypical Attention
  5. Substrates of Attentional Networks
- Further Reading

## GLOSSARY

**alerting** Achieving and maintaining a state of high sensitivity to incoming stimuli.

**attention** The mental ability to select stimuli, responses, memories, and thoughts that are behaviorally relevant among a host of others that are behaviorally irrelevant.

**attentional networks** Neural circuits subserving attentional processing that preserve a degree of anatomical and functional independence but interact in many practical situations.

**cognitive control** Processes such as conflict resolution, error correction, inhibitory control, planning, and resource allocation.

**executive** The mechanisms for monitoring and resolving conflict among thoughts, feelings, and responses (e.g., an attentional system concerned with such tasks as working memory, planning, switching, and inhibitory control).

**neuroimaging** Technological advances, often noninvasive, that permit tapping neurophysiological aspects of the behaving brain.

**orienting** The selection of information from sensory input.

**top-down modulation** A downstream (vs bottom-up) effect (e.g., cognitive control).

**self-regulation** Key mediator among genetic predisposition, early experience, and adult functioning (e.g., in controlling the reaction to stress, the capacity to maintain focused attention or the ability to interpret mental states both internally and in others).

Attention, endowed with a rich set of empirical findings, is a central theme in psychological research. Since the 1980s, human neuroimaging studies have allowed examination of the whole brain during tasks involving attention and, consequently, have provided much information on how the brain houses attentional processes. Brain networks subserving attention have been identified and described both anatomically and functionally. It is now possible to use these networks as model systems for the exploration of symptoms arising from various forms of pathology. This article outlines selective attention, its phenomenology, and its constituent attentional networks.

## 1. HISTORICAL CONTEXT

Attention, one of the oldest and most central issues in psychological science, is the problem of selecting for active processing certain aspects of the physical environment (e.g., objects) or ideas in a person's mind (which are stored in his or her memory). Many great minds have wrestled with the definition of attention. In 1890, William James wrote, "Everyone knows what attention

is. It is the taking possession of the mind in clear and vivid form of one out of what seem [to be] several simultaneous objects or trains of thought." James's account heavily joined attention with subjective experience. Moreover, James's effort to deal with both attention to objects and attention to "trains of thought" is important for understanding current approaches to sensory orienting and executive control. However, because attention in the sense of orienting to sensory objects can actually be involuntary and can occur unconsciously, attention is not, as the quote from James implies, the same as being aware.

Behavioral psychology postponed research into the internal mechanisms of attention after 1900. The quest for attentional mechanisms resumed following World War II, when Donald Broadbent proposed a filter, limited for the amount of information (in the formal sense of information theory), that was located between highly parallel sensory systems and a limited-capacity perceptual system. This view was of immense aid in making possible objective studies of limitations in the human ability to deal with more than one signal at a time in a variety of practical tasks.

As psychology moved toward the study of cognitive mechanisms, the new objective methods made it possible to ask about the processes of selection. It was found that words could activate their semantic associates even when there was no awareness of the words' identities (i.e., priming). The highly parallel organization found for sensory information extended to semantic processing. The act of selecting a word meaning for active attention appeared to suppress the availability of other meanings of the selected item and of competing items. More recently, it became increasingly clear that although attention can be contextualized as a form of "alertness" as well as an index to resource allocation/limitation, it has been viewed less as a filter or bottleneck and more as a mechanism for providing priority for motor acts, consciousness, and some kinds of memory.

Although the psychology of attention furnished a number of interesting results about the limits of performance and of unconscious processing, there was no agreement on whether attention involved separate mechanisms from those used to process data, let alone any analysis of what these mechanisms might be in neural terms. Advances in the understanding of neural systems underlying attention developed from experimental paradigms involving selection of sensory information and more recently from the technological innovation of noninvasive tools for imaging the living brain.

Brain imaging has forged an impressive link between psychology and neuroscience. As early as 1990, some

scholars suggested that the human brain entertains several attentional systems of different but interrelated functions (e.g., orienting, target detection, alerting). More recently, it has been demonstrated that distinct brain areas indeed mediate different attentional processes and that it is possible to examine selective attention as an organ system with its own functional anatomy, circuitry, and cellular structure. Although still incomplete, this information has illuminated multiple important questions in cognitive science and has provided insights into neurological and psychiatric disorders of both children and adults. Attention allows humans to exercise voluntary control over thoughts, feelings, and actions. Variations in the operational efficiency of these attentional systems serve as a basis for differences in self-regulation and emotional control, and they promise to help describe mechanisms of volition and sustained effort.

## 2. GROSS CHARACTERISTICS OF ATTENTION

Humans are largely visual animals, perhaps due to a common adaptation to visual predation generally found in primates. Researchers and clinicians have investigated the optics, anatomy, development, pathology, and underlying neural processes of the visual system, making it the most widely studied perceptual system. Therefore, it is didactically advantageous to discuss attention within the visual realm. Visual attention allows humans to move attention around to various areas of the visual field and to change the detail with which they look at any given area. For example, a reader can look at this page and pay attention to its setup as a whole, or he or she can zoom in on specific words and certain letters therein. If the reader is paying attention to single characters, he or she can glean a lot of information about punctuation marks, catch typos, and even spot minute imperfections on the physical sheet. But in that case, the reader might miss the idea that a paragraph is trying to convey. A person has the ability to change the location of attention and also to change the size of the attentional focus. This has been called the "zoom lens" idea, or the "attentional spotlight," and relates to humans' common experience concerning the kind of attention needed for reading versus proofreading.

Given a large visual array of individual features, a person can have a choice of examining it globally or instead examining its specific features. The person can shift back and forth between them by changing the

focus. Some patients have difficulty in examining the local features; these patients usually have damage to the left temporo-parietal lobe. Other patients may do well with the local features but fail to get the overall contour; they usually have damage to the right temporo-parietal lobe. The parietal lobe tends to emphasize the shifting between local and global, whereas the temporal lobe seems to determine whether one can actually examine a local feature or a global feature of the stimulus.

People usually foveate, or look at, exactly the thing in which they are interested, and that generally relates people's attention to where they fixate. However, it is easy to dissociate the two: to cue people to attend to some location in space other than the center of gaze and then to show that they are very sensitive (i.e., have a low threshold or fast response time) to information that occurs at the cued location and are relatively slow or insensitive to information that occurs at the fovea. These covert shifts are believed to be used to select the part of the visual field to which people usually want to move their eyes. In everyday life, people usually follow covert shifts of attention with eye movements. Attention to visual elements can also apply to other modalities (e.g., auditory).

When multiple people talk simultaneously, it is sometimes necessary to select one of these streams of conversation to follow it in detail. An individual usually does that based on the location of the other person by visually orienting toward the person and/or honing in on the frequency of his or her voice (e.g., it is easier to separate a male voice from a female voice than to separate two male voices) or on the content of the information (e.g., following by content). When an individual attends to one stream, the other information goes into the background; it is present but does not reach focal analysis.

There are data indicating that much of this unattended information is processed in subtle and complicated ways. Unattended information can suddenly get interesting because a person's name is mentioned or because something happens that is related to the conversation he or she is following and the person then finds himself or herself orienting to the new information. These phenomena have been studied experimentally in some detail.

### 3. ATTENTION AND PERCEPTION

The psychophysics literature provides good accounts of how visual thresholds correlate with attentional

investment. However, improvement in "visual acuity" is not synonymous with altered thresholds for detection, better performance, or faster reaction times. For example, acuity requires the resolution of detail, whereas detection thresholds and reaction time can involve the summation of luminance, and this might obscure detail.

There is evidence that attention improves performance in spatial resolution tasks. Cognitive scientists draw a distinction between how attention may be useful for simple detection of events and how performance can improve at those events. Although performance may improve on increased attentional investment, there has been great controversy over what orienting attention to a sensory (e.g., visual) stimulus actually does. There is general agreement that the attended stimulus receives priority, so that reaction time to it is usually faster. For example, in the visual modality, there is evidence of enhancement of brain electrical activity over extrastriate visual areas by approximately 90 milliseconds after visual presentation. On the other hand, attention is not a panacea to perception, and there is a great deal that attention cannot do. For example, it is clear that attention to a peripheral stimulus does not compensate for the lack of acuity that would be present for a foveal stimulus. Stimuli in the fovea always have an advantage in detail, although the priority for processing the input has been placed elsewhere. Thus, while orienting to a location, attention can give priority to that location (i.e., targets that appear there will be perceived more rapidly and with lower thresholds), but it cannot substitute for the acuity provided by the fovea. Although the fovea is critical for acuity, the costs in reaction time for an unexpected foveal stimulus are just as great as those for an unexpected peripheral event. Thus, visual attention influences priority or processing preference.

Whereas investing attention is frequently associated with looking directly at the scene of interest, covert attention is the ability to select visual information at a cued location, without eye movements, and to grant such information priority in processing. Researchers have shown that the performance improvement at attended locations results, to some extent, from an enhanced spatial resolution at the cued location. Findings from further psychophysical studies support the hypothesis that attention increases resolution at the attended location. Studies exploring the relationship between visual attention and contrast sensitivity show that covert attention not only improves discriminability in a wide variety of visual tasks but also can speed up the rate at which information is processed. There are findings indicating that a person's contrast

sensitivity is greater in the lower visual meridian (not the higher one). The bulk of the evidence sets limits to the effects of attention on spatial resolution and specifies that certain visual (not attentional) constraints determine aspects of spatial resolution.

#### 4. ATYPICAL ATTENTION

Attentional performance is affected by the biological rhythm. Diurnal reductions in attention normally occur during the hours of maximum sleepiness (2–7 AM) when body temperature is at a nadir, and enhanced performance is usually seen during the evening when body temperature peaks. During sleep, voluntary attention is often considered to be markedly attenuated or indeed absent. However, there is evidence to suggest that certain attentional, as well as preattentive, mechanisms remain intact. Dreaming is usually divorced from a sense of controlled awareness, but purported accounts of lucid dreaming (i.e., dreaming while knowing that one is dreaming) suggest that some control mechanisms may be available during sleep. Other common anecdotes include the incorporation of ambient sound into the dream content and the idea of sensitivity to a person's own name.

One of the ways in which to investigate information processing in sleep involves electrical recordings from the scalp (EEG). By averaging the brain's electrical response potentials to stimuli (ERP), it is possible to examine the processing capability of the sleeping brain. One such component to examine is the mismatch negativity (MMN). The MMN is an electrophysiological manifestation of involuntary preattentive processing in response to oddball stimuli. In a typical MMN paradigm, a "deviant" auditory stimulus is infrequently interspersed within a sequence of "standard" auditory stimuli. The MMN is evident in the difference waveform resulting from the subtraction of the ERP elicited by the standard stimulus from that elicited by the novel auditory stimuli (the deviants). The difference waveform, occurring even without attention, normally peaks between 100 and 250 milliseconds from the onset of the deviant event, depending on the dimension of deviance and its magnitude. The MMN is presumably associated with a mechanism that compares the current auditory input with the memory traces formed by previous auditory inputs and signals the occurrence of a mismatch.

In adults, MMN tends to decline during drowsiness. Whether it persists into adult human sleep is still debated, but other EEG components do reflect the brain's reaction to novelty. Although active midbrain inhibition blocks

cortical activity in the developed brain, there is reason to believe that the sleeping infant brain is not as capable of blocking and inhibiting information efficiently. Indeed, MMN is obtainable from newborns and young infants, and there are nascent experimental data showing that the brain can learn (e.g., certain language contrasts), even while asleep, during those early developmental stages.

Another special cognitive state sometimes confused with sleep is hypnosis. Some people (e.g., highly hypnotizable individuals) may experience attentional and perceptual changes, which might not typically occur during common awareness, following particular suggestions. Similar phenomena can occur in certain patient populations. For example, within vision, hypnotic suggestions have been demonstrated to induce tunnel vision, color-blindness, hallucinations, alexia, and agnosia. Such phenomena can manifest in other modalities as well (e.g., auditory).

Hypnosis has been used clinically for hundreds of years and is primarily a phenomenon involving attentive receptive concentration. Clinicians practicing hypnosis suggest that when a person is in a hypnotic state, attentional and perceptual changes may occur that would not have occurred if the person had been in a more usual state of awareness. In a responsive individual, hypnotic perceptual alteration is accompanied by reproducible changes in brain action. For example, the activity of the anterior cingulate to painful stimuli or conflict resolution can be modulated by hypnotic suggestion. Most children are highly hypnotizable and are more easily inducted into the hypnotic state than are adults.

#### 5. SUBSTRATES OF ATTENTIONAL NETWORKS

Attention can enhance neural processing at multiple levels. It is a selective aspect of information processing; some things are privileged and some things are ignored. The modulation of neural response by attention has been studied vigorously. For example, one recent hypothesis suggests that as processing becomes more complex, it becomes advantageous to spread processing across more neural areas (e.g., both hemispheres). Another experimental approach carefully crafted a behavioral task to probe and assess at least three distinct attentional networks: (a) achieving and maintaining the alert state, (b) orienting to sensory objects, and (c) selecting among conflicting actions. Behavioral and imaging data suggest that these networks

are independent. It is further possible to identify the neuroanatomy subserving these attentional networks.

Alerting involves a change in the internal state in preparation for perceiving a stimulus. The alert state is critical for optimal performance in tasks involving higher cognitive functions. Neuroimaging studies have shown activity in the frontal and parietal regions, particularly of the right hemisphere, when people are required to achieve and maintain the alert state even for a brief period. Lesions of these areas will reduce the ability to maintain alertness. Right frontal lesions have been shown to impair ability to voluntarily sustain attention, producing a larger number of errors over time than is found for left frontal patients in tasks involving continuous performance. Right parietal patients show deficits in maintaining the alert state and difficulty in attentional orienting that together produce a profound neglect in the visual field opposite the lesion. Alerting is thought to involve the cortical distribution of the brain's norepinephrine system arising in the locus coeruleus of the midbrain.

The orienting network involves the selection of information from sensory input. Cholinergic systems arising in the basal forebrain play an important role in orienting. The pulvinar, superior colliculus, superior parietal lobe, and frontal eye fields are often activated in studies of the orienting network. Orienting can be reflexive (e.g., when a sudden target event directs attention to its location), or it can be more voluntary (e.g., when a person searches the visual field looking for some target). Orienting typically involves head and/or eye movements toward the target, that is, overt orienting. However, it can also be covert. A few dorsal brain areas, including the superior parietal lobe and temporo-parietal junction, serve as a common source of attention to sensory stimuli. They produce effects within a network of areas that depend on modality (e.g., ventral visual areas in the case of visual input). The strongest evidence for localization of mental operations stems from the area of attentional orienting toward sensory stimuli, where a confluence of methods and experimental sophistication has demonstrated how separate brain areas can be invoked to organize a simple attentional shift. There is agreement that orienting of attention to a visual stimulus produces amplification in prestriate regions and that this affects processing in all subsequent regions and feeds back to influence processing in the primary visual cortex and perhaps in the lateral geniculate nucleus of the thalamus.

Executive control of attention involves more complex mental operations in monitoring and resolving conflict among computations occurring in different

brain areas. Executive control is most needed in situations that involve planning or decision making, error detection, novel (or not well-learned) responses, and conditions judged to be difficult or dangerous as well as in overcoming habitual actions. The anterior cingulate and lateral frontal cortex are target areas of the ventral tegmental dopamine system. Brain imaging data have shown repeatedly that the anterior cingulate cortex (ACC) is an important node in this network. Toward that end, a number of neuroimaging studies have shown activation of the dorsal anterior cingulate in tasks requiring people to respond either to a prepotent response or to a rather strong conflicting dimension. For example, in the classic Stroop task, experienced readers name the ink color of a displayed word. Readers are usually slower and less accurate in responding to the ink color of an incompatible color word (e.g., the word "red" displayed in blue ink) than in identifying the ink color of a control item (e.g., the word "lot" displayed in red ink). Another task involves participants responding to the direction of a central arrow when flanking arrows could point in either the same (congruent) direction or the opposite (incongruent) direction. The Attention Network Test uses this flanker task to measure conflict and shows strong activation in the dorsal ACC. Neuroimaging studies have shown that these conflict tasks activate midline frontal areas (e.g., ACC), lateral prefrontal cortex, and basal ganglia. These experimental tasks provide a means of fractionating the functional contributions of areas within the executive attention network. Chiefly, the ACC was more active on incongruent trials than on congruent trials. This result could reflect the general finding that lateral areas are involved in representing specific information over time (i.e., working memory), whereas medial areas are more related to the detection of conflict.

Patients with focal brain lesions of the ACC initially display deficits of voluntary behavior. The notion of ACC involvement in cognitive control and volition has been a topic of much interest recently. Based on behavioral, optical, and neuroimaging data, it was reported that effective suggestion (i.e., verbal exhortation) prevented word reading and modulated focal brain activity in highly suggestible individuals. This top-down influence was both potent and selective in that it annihilated the Stroop interference effect and reduced ACC activity, respectively. Interpretation of these data implies that, at least in highly suggestible individuals, attentional manipulations can influence aspects of self-regulation by affecting neural activity in specific brain areas. Although attention and self-regulation have arisen

within two different research traditions, there have been some recent efforts to integrate these two directions considering hypotheses about specific neural mechanisms involved in both attention and cognitive control. For example, these prospective approaches advocate a synthesis that is likely to be productive in linking brain systems to aspects of child socialization.

### See Also the Following Articles

Environmental Versus Individual Risk Taking: Perception, Decision, Behavior ■ Interpersonal Perception

### Further Reading

- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14, 340–347.
- Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Ann Neurol*, 10(4), 309–325.
- Posner, M. I., Petersen, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science*, 240, 1627–1631.
- Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. New York: Scientific American Books.
- Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation, and consciousness. *Philosophical Transactions of the Royal Society of London B*, 353, 1915–1927.
- Raz, A., Shapiro, T., Fan, J., & Posner, M. I. (2002). Hypnotic suggestion and the modulation of Stroop interference. *Archives of General Psychiatry*, 59, 1155–1161.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396, 72–75.



# Attention and Concentration Training in Sport

Aidan Moran

University College, Dublin, Ireland

1. "Attention" and "Concentration" in Sport
2. Training Attentional Skills in Athletes: Concentration Exercises and Techniques
3. Unresolved Issues in Training Attentional Skills in Sport
4. New Directions in Research on Concentration Skills Training  
Further Reading

**selective attention** The ability to "zoom in" on task-relevant information while blocking out distractions.

**simulation training** The use of simulated sport situations in practice in an effort to help athletes become accustomed to anticipated pressures and distractions.

**trigger words** Instructional cues that remind athletes what to focus on in competitive situations.

## GLOSSARY

**attention** People's ability to focus on information derived either from the external world or from internal sources such as memory and imagination.

**concentration** The ability to focus effectively on the task at hand while ignoring distractions.

**divided attention** The ability to perform two or more concurrent actions equally well.

**mental practice** The systematic use of mental imagery to rehearse actions covertly before they are executed physically.

**meta-attention** People's informal theories about how their attentional systems work.

**performance goals** Specific tasks or actions that lie within the control of the athlete.

**preperformance routines** Sequences of preparatory actions that athletes tend to engage in before they execute self-paced skills.

The term "attention" refers to people's ability to focus on information derived either from the external world or from internal sources such as memory and imagination. In sport psychology, attentional processes such as "concentration," or the ability to focus mental effort on the task at hand while ignoring distractions, are regarded as vital determinants of successful athletic performance. Given this importance of attention in sport, a variety of psychological exercises and techniques have been postulated to enhance athletes' concentration skills. Although none of these interventions has been validated adequately so far, some theoretical support exists for the use of the following strategies: simulation training, performance goal setting, preperformance routines, trigger words, and mental practice. Following an explanation of each of these concentration interventions, this article considers some unresolved issues along with potentially fruitful new directions for research in this field.

## 1. “ATTENTION” AND “CONCENTRATION” IN SPORT

Cognitive sport psychology is concerned with the scientific study of how the mind works in athletic settings. Within this field, the term “attention” refers to people’s ability to focus on information derived either from the external world or from internal sources such as memory and imagination. This ability is regarded as a multi-dimensional construct that has at least three key components that can be explained as follows. First, when an athlete is asked to “pay attention” to an instruction provided by a coach at a training session, the athlete is required to concentrate on, or to exert mental effort in absorbing, the information presented. The second component of attention involves selective perception and occurs when sport performers “zoom in” on task-relevant information while blocking out distractions. To illustrate, a soccer goalkeeper who is preparing to catch a corner kick must focus only on the flight of the ball while ignoring the potentially distracting movements of other players in the penalty area. Finally, attention involves “multitasking,” that is, the coordination of simultaneous skilled actions. For example, when a basketball player dribbles with the ball while scanning the court for a teammate to pass the ball to, the player is engaging in a form of mental time sharing called “divided attention.” In summary, the construct of attention refers to three distinct cognitive processes: concentration or effortful awareness, selectivity of perception, and the ability to coordinate two or more concurrent actions successfully.

Since the 1950s, two main theoretical models of selective attention have been postulated in cognitive psychology. First, the “spotlight” metaphor suggests that selective attention resembles a mental beam of light that illuminates targets located either in the external world around us (sensory information) or in the subjective domain of our own thoughts and feelings (mental events). Another popular theoretical approach is the “resource” model of divided attention. This metaphor suggests that attention is analogous to a pool of mental energy that is allocated to fulfill processing demands according to certain strategic principles. To illustrate, the principle of “automaticity” suggests that the more practiced a task is, the more automatic it becomes and the less attentional resources it requires. Based on this principle, it may be predicted that expert athletes are especially vulnerable to distractions because they have “spare” mental resources due to the automated

nature of many of their perceptual motor skills. The issue of how to refocus the minds of such performers is addressed later in this article.

Athletes, coaches, and sport psychologists regard attentional skills such as concentration, or the ability to focus effectively on the task at hand while ignoring distractions, as essential prerequisites of success in sport. This claim is supported by both descriptive and experimental evidence. First of all, surveys of expert sport performers such as U.S. Olympic athletes indicate that concentration is consistently rated among the most important mental skills required for sporting excellence. Next, qualitative studies suggest that total absorption in the task at hand is a key feature of “flow” or “peak performance” experiences in sport, that is, those coveted yet elusive moments during which athletes perform at the best of their abilities. Finally, there is growing experimental evidence that athletes (e.g., sprinters) who have been trained to use certain concentration strategies (e.g., deliberately focusing on task-relevant cues) tend to perform better than their counterparts who are assigned to nonfocusing control conditions. Taken together, these sources of evidence converge on the conclusion that the ability to pay attention to the task at hand is a vital determinant of success in sport.

Based on peak performance research, three key principles of effective concentration can be identified. First, a focused state of mind requires intentionality and a willingness to expend deliberate mental effort on the part of the athlete concerned. Put simply, one must prepare to concentrate effectively rather than merely hope that it will happen fortuitously. Second, although skilled athletes might be able to divide their attention efficiently between two or more concurrent actions, they can focus consciously on only one thought at a time. This principle stems from the fact that “working memory,” or the cognitive system that regulates people’s attentional deployment, is limited in its capacity and duration. Accordingly, it can be overloaded easily when several thoughts and/or behavioral intentions compete for conscious attention. Interestingly, this danger of cognitive overload is increased when athletes become preoccupied with winning a competitive encounter. In this situation, they may find themselves torn between trying to focus on performing an action and worrying about the likely outcome of the encounter in which they are participating. Finally, as mentioned earlier, research on flow states indicates that athletes’ minds are focused optimally when they are totally absorbed in the task at hand. In such a state of

mind, there is no difference between what they are thinking about and what they are doing. In summary, athletes tend to concentrate most effectively when they prepare to exert mental effort and can aim their attentional spotlight at actions that are specific, relevant, and under their own control.

## 2. TRAINING ATTENTIONAL SKILLS IN ATHLETES: CONCENTRATION EXERCISES AND TECHNIQUES

Given the importance of attentional processes in sport, a wide range of strategies has been promulgated by applied sport psychologists in an effort to enhance athletes' focusing skills. In general, these interventions fall into two types of activities: concentration training exercises and concentration techniques. Theoretically, the main distinction between these activities is that the former are designed for use during athletes' training sessions, whereas the latter tend to be practiced most frequently in actual competitive situations. Not surprisingly, however, this distinction is blurred by the fact that athletes are increasingly aware of the value of replicating competitive situations in their training regimes. Accordingly, there may be an overlap between the concentration exercises and techniques they employ in training and in competition.

### 2.1. Concentration Exercises

One of the most popular focusing exercises recommended by sport psychologists is a visual search task called the "concentration grid." In this task, participants are presented with a random array of two-digit numbers (from 00 to 99) on a page and are required to scan and "tick off" as many digits as possible in a given sequence (e.g., beginning with 00 and increasing consecutively) within the specified time limit of 1 minute. The number of digits marked in the correct sequence within this duration serves as an alleged index of the athletes' concentration skills. Unfortunately, although the concentration grid may have some intuitive plausibility, it has not been validated as a concentration exercise either theoretically or empirically. Therefore, its conceptual rationale and practical utility remain unresolved.

In contrast to the grid, a more recent concentration exercise known as "simulation training" seems to have some merit. Briefly, this exercise is based on the proposition that athletes can learn to improve their

concentration skills if they simulate real-life competitive situations in practice. Anecdotal testimonials to the value of this practice in various sports have emerged during recent years. For example, some leading soccer coaches have prepared their players for the pressure of penalty "shoot-outs" by requiring them to practice walking from the center circle of the pitch to the penalty area—as happens in actual match conditions. Unfortunately, despite its apparent face validity, simulation training has received little or no empirical scrutiny as a concentration strategy. Nonetheless, some support for its theoretical rationale may be found within the field of cognitive psychology. For example, research on the "encoding specificity" principle of learning shows that people's recall of information is facilitated by conditions that resemble those in which the original encoding occurred. Applying this principle to simulation training, it may be argued that replication of real-life conditions in practice inoculates athletes against anticipated pressures and distractions. By this reasoning, it is conceivable that simulation training could enhance athletes' concentration skills.

### 2.2. Concentration Techniques

The second category of concentration strategies recommended for use by athletes in competitive situations involves the use of psychological techniques such as (a) performance goal setting, (b) preperformance routines, (c) trigger words, and (d) mental practice. These techniques are explained in what follows.

#### 2.2.1. Performance Goal Setting

In psychology, a "goal" is a target or an objective that people strive to attain such as sinking a putt, winning a match, or being selected for a given team. Therefore, "goal setting" is the process by which people establish desirable objectives for their actions. Typically, sport psychologists distinguish between "result goals" (e.g., the outcome of a competition) and "performance goals" (i.e., specific actions that are under the athlete's control). Using this distinction, some psychologists have speculated that the process of setting performance goals can improve athletes' concentration skills. For example, a golfer could improve his or her concentration on the course by focusing on specific "controllable" goals such as keeping his or her head steady and/or maintaining a slow rhythmic swing. Empirical support for this conjecture springs from studies on the correlates of people's "best" and "worst" athletic performances.

In particular, research shows that collegiate athletes performed worst when they were preoccupied with result goals. Conversely, their best performances tended to occur when they adopted explicit performance goals. In summary, there is a legitimate theoretical rationale for the use of performance goals as a concentration technique by athletes.

### **2.2.2. Preperformance Routines**

It has long been apparent that many athletes display distinctive idiosyncratic sequences of preparatory actions before they perform key skills, especially in individual sports. For example, top tennis players tend to bounce the ball a consistent and characteristic number of times before their first serve and a different yet equally consistent number of times before their second serve. These bounce patterns help the player to focus on the first step in his or her pre-serve preparatory sequence. Similarly, expert golfers tend to adhere to a preferred number of practice swings and “waggles” of the club before they strike the ball. These distinctive action sequences and/or repetitive behaviors are called “preperformance routines” and are especially prevalent prior to the execution of “self-paced” skills (i.e., actions that are carried out largely without interference from other people).

Three main types of routines are evident in sport. First, “pre-event” routines are general patterns of preparation that athletes display on the days preceding competitive action. Included here are preferences with regard to food and sleeping arrangements. Second, “pre-performance” routines are characteristic sequences of thoughts and actions that athletes adhere to prior to skill execution. Finally, “postmistake” routines are action sequences that may help performers to put errors behind them so that they can refocus on the task at hand. For example, a tennis player may “shadow” the correct movement of a volley that led to an error.

Support for the value of preperformance routines as concentration techniques comes from both theoretical and empirical sources. Theoretically, preperformance routines may improve concentration for at least two reasons. First, they may enable athletes to concentrate on the current moment rather than on past events or on possible future outcomes. In addition, they may prevent athletes from devoting too much attention to the mechanics of their well-learned skills. This habit, which is also known as “paralysis by analysis,” can “unravel” automaticity and lead to a deterioration in performance. In other words, routines may help to

suppress the type of inappropriate conscious control that often occurs in pressure situations. Empirically, there is some evidence from case studies that routines can improve athletes’ concentration skills and performance. For example, research suggests that more proficient golfers benefited more from using routines than did less skilled golfers.

Unfortunately, advocacy of preperformance routines gives rise to two practical issues that need to be addressed. First, routines may lead to superstitious rituals on the part of the performer. Thus, certain athletes feel compelled to wear “lucky” clothes (e.g., Tiger Woods likes to wear something red on the final day of a golf tournament) or to listen to “lucky” songs before important competitive events. Indeed, athletes’ precompetitive behavior is often a mixture of routines and superstitions. A second problem with routines is that they need to be revised regularly as a precaution against the possibility that they will become automated through frequent practice. In other words, if athletes maintain the same preperformance behavior indefinitely, their minds may begin to wander as they proceed through the various steps of the routine. Because of this problem, applied sport psychologists face the challenge of helping athletes to develop consistent but not automated preparatory actions.

### **2.2.3. Trigger Words**

Most athletes talk to themselves covertly as they train or compete. For example, gymnasts may use words such as “forward” to remind themselves to push their bodies upward as they attempt to master a floor routine. Similarly, tennis players may try to trigger a smooth service action by telling themselves to “reach and hit.” In general, this internal speech (or “self-talk”) may involve praise (e.g., “well done—that’s good”), criticism (“you idiot—that’s a stupid mistake”), or instruction (“swing slowly”).

Typically, the purpose of instructional verbal triggers is to help athletes refocus their minds on task-relevant cues just before skill execution. For example, when standing over a putt, golfers may use phrases such as “steady head” to prevent themselves from looking up to see whether or not the ball has gone into the hole.

Theoretically, self-talk could enhance concentration skills by reminding athletes about the most important cues in any given situation. Interestingly, this strategy was used by Serena Williams during the 2002 Wimbledon women’s singles tennis final in which she defeated her sister Venus. During this match, Serena

read notes to herself during the “change-over” time between games. Afterward, she explained that these notes had contained trigger words to remind her to “hit in front of you” or to “stay low.”

Despite such anecdotal testimonials to the value of trigger words, few studies have tested the validity of self-talk as a concentration technique. However, it is generally believed that for optimal efficacy, instructional self-statements must be short, vivid, and positively phrased.

#### **2.2.4. Mental Practice**

The term “mental practice” or “visualization” refers to the systematic use of mental imagery to rehearse physical actions. Put simply, it involves “seeing” and “feeling” a skill in one’s imagination prior to its actual execution. This strategy is used by golfers such as Tiger Woods when preparing to play a shot. Although there is considerable experimental evidence that mental practice facilitates skill learning and performance, few studies have been conducted on imagery as a concentration technique. Anecdotally, however, many athletes report using imagery to prepare for anticipated scenarios, thereby reducing the likelihood of being distracted by unexpected events.

### **3. UNRESOLVED ISSUES IN TRAINING ATTENTIONAL SKILLS IN SPORT**

At least five issues need to be addressed before attentional and/or concentration training techniques can be used effectively with athletes in applied settings. First, before they attempt to train concentration skills, sport psychologists should try to find out why athletes lose their focus in the first place. Unfortunately, until recently, little research was available on the causes of distractibility in sport performers, especially distractibility that arises from “internal” sources such as athletes’ own thoughts and feelings. However, some recent theoretical models have examined the role of unconscious factors in this regard. Augmenting this research, self-report scales have been developed by sport psychologists to assess athletes’ susceptibility to self-generated cognitive interference. Second, sport psychologists have much to learn about the attentional demands of various sports. For example, whereas some activities (e.g., weightlifting) may require short periods of intense concentration, others (e.g., cycling) appear to demand

sustained alertness for longer periods of time. Intuitively, it seems unreasonable to expect that the same toolbox of concentration interventions will work equally well in all sports. Third, some researchers have raised the question of whether or not sport performers know exactly what they should be concentrating on in various athletic situations. For example, should a tennis player focus only on the ball during a rally, or should the player use cues from his or her opponent’s behavior (e.g., position of feet) in an effort to anticipate the likely target of a subsequent shot? Unfortunately, this question has been largely neglected by sport psychologists in their enthusiasm to provide practical mental skills training programs for athletes. Fourth, what is the best way in which to measure concentration skills in athletes? Until this question has been answered empirically, it is difficult to evaluate whether or not concentration skills interventions are effective. Finally, no explicit criteria currently exist for evaluating the maintenance of attentional skills improvements over time. In the absence of such criteria, researchers have not yet ruled out the possibility that athletes’ concentration patterns may return to preintervention levels despite athletes’ participation in mental skills training programs.

### **4. NEW DIRECTIONS IN RESEARCH ON CONCENTRATION SKILLS TRAINING**

At least four potentially fruitful new directions in research on concentration skills training in athletes may be identified. First, further research is required on the “meta-attentional” processes of athletes, that is, their theories on how their own concentration systems work. It may be argued that the entire program of concentration skills training in sport psychology is really an exercise in meta-attentional training, whereby athletes are empowered psychologically by gaining an understanding of, and control over, their own concentration processes. As yet, however, little is known about the nature, accuracy, and/or modifiability of athletes’ meta-cognitive theories. Second, additional research is required on the relationship between the structure of various athletic activities and their attentional demands. For example, do untimed activities such as golf place different cognitive demands on athletes’ concentration systems compared with those imposed by timed competitions such as soccer? If so, what theoretical mechanisms could account for such differences? Third, more research is needed to establish

how emotional factors (e.g., anxiety) affect athletes' concentration processes. This task could be achieved by exploring the effects of anxiety on the visual search behavior of athletes tackling laboratory simulations of sport-relevant tasks. Finally, research is required to evaluate the efficacy of simulation training and the various concentration techniques described earlier in improving athletes' focusing skills in competitive sport situations.

### See Also the Following Articles

Attention ■ Goal Setting and Achievement Motivation in Sport ■ Intrinsic and Extrinsic Motivation in Sport ■ Psychological Skills Training in Sport

### Further Reading

- Abernethy, B. (2001). Attention. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 53–85). New York: Macmillan.
- Boutcher, S. H. (2002). Attentional processes and sport performance. In T. Horn (Ed.), *Advances in sport psychology* (2nd ed., pp. 441–457). Morgantown, WV: Fitness Information Technology.
- Fernandez-Duque, D., & Johnson, M. L. (1999). Attention metaphors: How metaphors guide the cognitive psychology of attention. *Cognitive Science*, 23, 83–116.
- Hatzigeorgiadis, A., & Biddle, S. J. H. (2000). Assessing cognitive interference in sport: Development of the Thought Occurrence Questionnaire for Sport. *Anxiety, Stress, and Coping*, 13, 65–86.
- Jackson, R. C., & Baker, J. S. (2001). Routines, rituals, and rugby: Case study of a world class goal kicker. *The Sport Psychologist*, 15, 48–65.
- Moran, A. P. (1996). *The psychology of concentration in sport performers: A cognitive analysis*. Hove, UK: Psychology Press/Taylor & Francis.
- Moran, A. P. (2000). Improving sporting abilities: Training concentration skills. In J. Hartley, & A. Branthwaite (Eds.), *The applied psychologist* (2nd ed., pp. 92–110). Buckingham, UK: Open University Press.
- Moran, A. P. (2003). The state of concentration skills training in applied sport psychology. In I. Greenlees, & A. P. Moran (Eds.), *Concentration skills training in sport* (pp. 7–19). Leicester, UK: British Psychological Society.
- Moran, A. P. (2004). *Sport and exercise psychology: A critical introduction*. London: Psychology Press/Routledge.
- Schmid, A., & Peper, E. (1998). Strategies for training concentration. In J. M. Williams (Ed.), *Applied sport psychology: Personal growth to peak performance* (3rd ed., pp. 316–328). Mountain View, CA: Mayfield.
- Simons, J. (1999). Concentration. In M. A. Thompson, R. A. Vernacchia, & W. E. Moore (Eds.), *Case studies in applied sport psychology: An educational approach* (pp. 89–114). Dubuque, IA: Kendall/Hunt.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, 101, 34–52.
- Williams, J. M., & Leffingwell, T. R. (2002). Cognitive strategies in sport and exercise psychology. In J. Van Raalte, & B. W. Brewer (Eds.), *Exploring sport and exercise psychology* (2nd ed., pp. 75–98). Washington, DC: American Psychological Association.



# Attention Deficit Disorders: School-Based Interventions

George J. DuPaul and Katy E. Tresco

*Lehigh University, Bethlehem, Pennsylvania, USA*

1. Description and Definition of Attention Deficit/  
Hyperactivity Disorder
2. Conceptual Foundations of School-Based Interventions
3. Interventions for Elementary School-Aged Students
4. Conclusions and Future Considerations  
Further Reading

## GLOSSARY

**antecedent-based intervention** Any treatment strategy that involves manipulating environmental events prior to a target behavior occurring so as to increase or reduce the probability that the target behavior will occur; these interventions are considered to be proactive preventive strategies.

**consequent-based intervention** Any treatment strategy that involves manipulating environmental events after a target behavior occurs so as to increase or reduce the probability that the target behavior will occur; these interventions are considered to be reactive strategies.

**curriculum-based measurement (CBM)** Assessment of a student's math or reading skills using brief (2- to 3-minute) probes derived directly from the curriculum being taught in the classroom; CBM provides a cost-efficient and sensitive method of documenting a child's instructional level and changes in skills over time.

**functional assessment** Collection of interview and/or observation data to establish the environmental events that increase the probability that a target behavior will occur.

**point of performance** The exact time that a target behavior is most likely to occur and the setting where a behavior is likely to be exhibited.

Attention deficit/hyperactivity disorder (AD/HD) is associated with academic and social impairments that significantly compromise school performance. Effective school-based interventions include antecedent-based and consequent-based strategies that are implemented by teachers, peers, parents, computers, and/or the target students themselves.

## 1. DESCRIPTION AND DEFINITION OF ATTENTION DEFICIT/ HYPERACTIVITY DISORDER

Attention deficit/hyperactivity disorder (AD/HD) is a psychiatric disorder characterized by developmentally inappropriate levels of inattention, impulsivity, and/or overactivity. There are three subtypes of AD/HD: predominantly inattentive, predominantly hyperactive-impulsive, and combined type, with the latter being most common among clinic referrals and research participants. AD/HD affects approximately 3 to 5% of the school-aged population in the United States, with boys outnumbering girls by two to one in epidemiological

samples and by six to one in clinic samples. The symptoms of AD/HD frequently are associated with impairment in academic, social, and behavioral functioning. The disorder tends to be chronic and often is a precursor to the development of other disruptive behavior disorders, including oppositional defiant disorder and conduct disorder. Among the greatest long-term risks associated with AD/HD are scholastic underachievement, grade retention, and school dropout. The most effective treatments for this disorder include psychotropic medication, especially psychostimulants (e.g., methylphenidate), behavioral strategies applied at home and school, and (when necessary) academic interventions. It is assumed that most children with AD/HD will require a combination of these treatments to successfully ameliorate symptomatic behavior over the long term.

The most widely studied and used medications for AD/HD are central nervous system (CNS) stimulants such as methylphenidate (Ritalin, Concerta, and Metadate), combined stimulant compounds (Adderall), and dextroamphetamine (Dexedrine). CNS stimulants are associated with improved attention and impulse control in the majority of treated children with AD/HD. Side effects are benign in most cases, with insomnia and appetite reduction being the most common adverse effects. Behavioral effects vary considerably across individuals as well as across doses within individuals. Furthermore, stimulant medication effects are observed primarily during the school day because their duration of action ranges from 4 to 8 hours. For these reasons, it is important that information regarding behavior changes at home and school, as well as possible improvement or deterioration in academic performance, is taken into account when making medication decisions. For those individuals who do not respond to a CNS stimulant, alternative medications that may improve AD/HD symptoms include antidepressants (e.g., bupropion), antihypertensives (e.g., clonidine), and selective norepinephrine reuptake inhibitors (e.g., atomoxetine).

## **2. CONCEPTUAL FOUNDATIONS OF SCHOOL-BASED INTERVENTIONS**

There are several principles that are critical to the design of effective school-based interventions for students with AD/HD: gathering assessment data that directly inform intervention planning, implementing interventions at

the point of performance, individualizing intervention strategies for each student, using a balanced treatment plan composed of both antecedent-based and consequent-based interventions, using strategies to address both academic and behavioral difficulties, and employing multiple individuals to implement treatment components.

### **2.1. Linking Assessment Data to Intervention**

The assessment of children suspected of having AD/HD not only should be directed at establishing a diagnosis but also should provide data that can directly inform treatment planning. Thus, although norm-referenced measures such as behavior rating scales are helpful for making diagnostic decisions, additional measures should be included to provide information about possible directions for treatment. Specifically, functional assessment procedures, such as observations of target behaviors (e.g., calling out in class without permission) in the context of environmental events (e.g., classmates laughing), can be used. Data from a functional assessment can help to identify antecedents and consequences that may be maintaining a target behavior and that can be manipulated to change the frequency of a behavior. For example, disruptive behavior that consistently leads to peer attention as a consequence can be reduced by designing an intervention that provides the student with peer attention for engaging in appropriate behavior (e.g., peer tutoring).

### **2.2. Intervening at the Point of Performance**

Because children with AD/HD typically exhibit poor impulse control, their behavior is more likely to be controlled by immediate environmental events than by long-term contingencies. For this reason, the most effective intervention strategies will be those that are implemented as close to the “point of performance” as possible. For example, if an intervention is necessary to reduce disruptive behavior occurring in math class and the latter is conducted from 9:00 to 9:45 AM each school day, the most effective interventions will be those that are implemented in math class from 9:00 to 9:45 AM. Treatment procedures that are removed in time and place from the point of performance (e.g., weekly counseling sessions in the school psychologist’s office) will be less effective.

### **2.3. Individualizing Intervention Strategies**

Children with AD/HD exhibit a wide range of possible symptoms, disruptive behaviors, and academic difficulties. Thus, a “one size fits all” approach to treatment, wherein it is assumed that all students with AD/HD respond the same way to the same interventions, is likely to fail. The selection of intervention components should be made on the basis of individual differences in symptom severity, presence of comorbid conditions (e.g., oppositional defiant disorder), possible functions of the target behaviors, teacher acceptability of prescribed interventions, and empirical evidence supporting specific treatments for the target behaviors.

### **2.4. Combining Antecedent and Consequent Interventions**

The application of both antecedent and consequent intervention strategies allows for students to have clear expectations of what is expected of them in advance (antecedent) as well as what will occur if problem behavior arises (consequent). Combining antecedent and consequent interventions also provides for more positive interactions and experiences, thereby potentially increasing the chances that the student will behave appropriately. Interventions based solely on aversive consequences or punishments often result in increased negative interactions between the student and the teacher and minimize the opportunity for positive reinforcement.

Antecedent or proactive intervention strategies are those that are implemented to prevent or preempt academic and/or behavioral difficulties from occurring. These could involve implementing classroom strategies and rule systems to provide structure, making environmental modifications (e.g., placing a child with academic or behavioral difficulties near the teacher or near children who are less likely to interact negatively with the child), having a child repeat instructions before beginning an assignment or activity, and/or providing limited task choices to a child prior to beginning seatwork.

Alternatively, consequent or reactive strategies are those implemented to respond to both appropriate and problem behavior when it does occur. An example of a consequent intervention might be a response cost system in which the student is “fined” for exhibiting problem behavior (e.g., loses points or reinforcement)

yet also has the ability to earn and receive positive feedback when appropriate behavior is exhibited.

### **2.5. Combining Behavioral and Academic Interventions**

Children with AD/HD often exhibit both academic and behavioral difficulties in the classroom. It is important to address both areas because a student who is frustrated academically may be more likely to exhibit negative behavior. Based on the results of a 1997 meta-analysis performed by DuPaul and Eckert on studies of school-based interventions from 1971 to 1995, improving a child’s academic skills may potentially improve his or her behavior in the classroom as well. In addition, targeting the attention difficulties associated with AD/HD and improving on-task behavior through behavioral intervention may potentially have a positive impact on academic skills. By including strategies targeting both academic skill areas as well as relevant behavior problems, teachers may be able to maximize the effectiveness of the interventions implemented. Collecting curriculum-based measurement and functional assessment data can lead to a better understanding of the factors contributing to academic and behavioral problems. Identifying the specific instructional needs and environmental variables that reliably proceed or follow inappropriate behavior can be a key component in the formulation of effective interventions

### **2.6. Using Multiple Intervention Agents**

Effective intervention strategies can be implemented by a variety of mediating or intervention “agents.” Interventions can be mediated by teachers, peers, computers, parents, and/or the students themselves. By using multiple agents, sole responsibility for addressing all academic and behavioral difficulties associated with AD/HD does not fall on the teacher. Instead, multiple resources can be used to provide a more comprehensive, and potentially more cost-effective, method to support teacher instruction as well as to deliver interventions. [Table I](#) provides some possible intervention strategies listed by mediating agent. Specific intervention strategies organized by mediating agent and setting (elementary school vs secondary school) are discussed in more detail in later sections.

**TABLE 1**  
**Intervention Strategies by Mediating Agent**

<i>Mediating agent</i>	<i>Elementary school</i>	<i>Secondary school</i>
Teacher	Classroom rules Response cost systems Environmental manipulations Instructional modifications Instructional strategies	Contingency contract Direct instruction in study/organizational skills
Peer	Peer tutoring Group contingencies	Peer coaching
Computer	Instruction Drill and practice	
Parent	Goal setting Contingency contracting Home-based reinforcement Parent tutoring	Home-based reinforcement Homework program
Self	Self-monitoring	Self-monitoring Self-evaluation/ Reinforcement

Source: From DuPaul, and Power (2000). Copyright 2000 by the American Psychiatric Association. Adapted by permission.

### 3. INTERVENTIONS FOR ELEMENTARY SCHOOL-AGED STUDENTS

As described previously, combinations of interventions appear to be most effective in improving and maintaining appropriate classroom behavior. This section discusses intervention strategies for improving academic and behavior performance using both antecedent and consequent intervention strategies across mediating agents. Typically, the most effective interventions include the use of multiple strategies and intervention agents.

#### 3.1. Teacher-Mediated Strategies

Classroom strategies that have previously been found to be effective in improving the academic and behavior

performance of children with AD/HD include teaching classroom rules, using response cost systems, altering the structure of the classroom environment, adjusting academic expectations to match students' current skills or instructional level, modifying instructions, and arranging for choices in academic tasks.

Most classrooms have rules that are typically reviewed with students at the beginning of the academic school year. Periodic review and reteaching of these rules throughout the year can be an effective support system for maintaining appropriate behavior. Teachers should develop four or five positively phrased classroom rules. By stating them in a positive manner, the rules themselves help to inform students as to how to behave instead of providing an example of what "not" to do. Initial teaching should include both examples and nonexamples of following the rules, and examples should be elicited from the students to ensure their understanding. Periodic review should include a few minutes spent each week reviewing one or two of the rules and their importance and again eliciting examples from students to confirm their understanding. This frequent review and discussion of rules enables students to have a clear understanding of what is expected of them. In addition, pairing classroom rules with a response cost system can be an effective combination of antecedent- and consequent-based strategies.

A response cost system requires a clearly defined rule system that is reviewed on a frequent basis. Within a response cost system, students may earn "tokens" (e.g., points, chips) for following rules and can lose tokens when rules are broken. The tokens can then be applied toward privileges such as taking the attendance to the office, passing out papers, and purchasing items from a "class store." For a response cost system to be effective, students must have clear expectations of when the rules are in effect and what behavior will result in both the earning and losing of tokens. Examples of both situations should be given to confirm students' understanding while the rules are being reviewed. Students should also understand what privileges are available, when they may exchange their tokens for privileges, and how many tokens are necessary for each privilege. Student motivation is a key component of an effective response cost system. Therefore, it may be beneficial to have students participate in a discussion of which privileges will be available and at what cost. It may also be helpful to determine ahead of time what privileges students are working for and to remind them periodically.

It is important to note that although classwide systems are widely effective for most students in the classroom,

individual modifications may be necessary for those students whose behavior problems are more severe. If a student typically breaks 20 rules in a 30-minute class period, asking the student to break only 2 rules may be an unrealistic expectation and the child may be unmotivated by the system. Therefore, the amount of tokens necessary to earn a privilege might need to be altered on an individual basis. Similarly, if students can earn points for following rules for a certain amount of time, it might be necessary to decrease the required duration for specific students to allow them to earn points. Adjusting the time period also provides for more immediate feedback to those students who might need more frequent reinforcement.

In addition to classroom rules and response cost systems, children with AD/HD may benefit from alterations to the classroom environment to allow for increased monitoring of classroom behavior. Children with AD/HD may benefit from being placed near the teacher or with peers who are less likely to interact negatively with them. In addition, although it may seem advantageous to place children with severe attention and behavior difficulties in isolated areas of the classroom, it is important that they remain a part of important classroom activities and have access to teacher instruction and positive peer interactions.

Behavior and academic difficulties often associated with AD/HD may be exacerbated by instructional demands that are too high for students' current level of functioning. Frustration with tasks that are too difficult may increase the likelihood that some children will exhibit negative behaviors. Thus, it is important to assess and adjust academic expectations to match a student's current skills or instructional level. Curriculum-based measurement (CBM) data may be a helpful tool for teachers to determine the appropriate level of instruction that a child should receive. CBM can also provide information as to areas in which a child might need additional instruction and academic intervention. Instructional modifications and interventions for children may include direct instruction in particular skills that a child is lacking as well as drill and practice activities such as flashcard drills and structured worksheets.

Students with attention or behavior difficulties may also benefit from being provided with choices in academic tasks within a particular academic area. For example, students may be provided with a menu of acceptable tasks for a particular subject from which to choose. For this strategy, the teacher offers a menu of tasks that he or she finds acceptable, and the student chooses one to complete during the time allotted. The menu and choice component allows for both the teacher and the student to

maintain control over the work completed. Thus, providing choices in academic tasks may increase the student's work productivity and on-task behavior.

### 3.2. Peer-Mediated Interventions

Peers working together on an instructional activity can also have a positive impact on the academic engagement and performance of students with AD/HD. In addition, peer tutoring strategies, by targeting academic instruction through peer interaction, are able to address both academic and social skills simultaneously. Because children with AD/HD have often been found to have difficulty in both areas, peer-mediated strategies can be a particularly useful tool.

Although different peer-mediated strategies vary regarding their foci of instruction and the structure and procedures of the tutoring teams, Barkley's 1998 handbook for diagnosis and treatment of AD/HD listed the following four common characteristics that have been shown to enhance task-related behaviors of children with AD/HD: (a) working one-on-one with another individual, (b) letting the learner determine the instructional pace, (c) continually prompting academic responses; and (d) providing frequent immediate feedback about quality of performance.

In 2002, Greenwood and colleagues described peer tutoring as an instructional strategy in which classroom teachers train and supervise students to teach their peers. Peer-directed instruction allows for high levels of student engagement, sufficient practice, immediate error correction and feedback, and the integration of students with varying abilities. In their review of peer tutoring strategies, Greenwood and colleagues described four programs that have been found to have well-defined procedures and supporting evidence in the research for use with students with disabilities: Classwide Peer Tutoring (CWPT), Peer-Assisted Learning Strategies (PALS), Classwide Student Tutoring Teams (CSTT), and Reciprocal Peer Tutoring (RPT).

### 3.3. Computer-Assisted Interventions

Computer-assisted interventions (CAI) can be an effective tool for promoting skill acquisition through the use of instructional technology as well as for providing additional drills and practice to promote mastery of previously acquired skills. Features of computer programs (e.g., provision of immediate performance

feedback) may be able to positively affect on-task behaviors and work production of children with AD/HD. In 2003, DuPaul and Stoner listed the following important design features of CAI that may be beneficial for children with AD/HD: specific instructional objectives presented along with activities, use of print or color to highlight important material, use of multiple sensory modalities, content divided into manageable sized chunks of information, the ability to limit nonessential distracting features (e.g., sound effects, animation), and immediate feedback on response accuracy.

### 3.4. Parent-Mediated Interventions

Parents should also be considered a valuable resource in promoting the academic success of their children. Frequent and clear communication between teachers and parents provides parents with knowledge about their children's academic and classroom-related behaviors and also provides teachers with information regarding events at home that may affect classroom behavior and performance. Parents can then support and reinforce instruction in the classroom as well as help to promote appropriate classroom behavior.

Parents, in providing instructional support to their children, can engage their children in flashcard drills for additional practice in an area, act as tutors to aid in skill acquisition, and make themselves available for homework support. Many parents may already participate in similar activities or would like to participate but are unsure as to how they can be most helpful. A key component to effective instructional support at home is parents' knowledge regarding instruction that their children are receiving at school, areas in which their children are struggling and could use additional support, and areas in which their children are succeeding.

Communication can take place through multiple methods. Brief notes can be sent home on a daily or weekly basis, weekly phone calls or e-mails can be scheduled, brief conversations can occur when picking up children, or another system that is mutually agreeable between families and the school can be arranged. One example of school-home communication that supports positive behavior in the classroom is a school-home contingent reinforcement system (i.e., daily report card). This type of system involves setting goals and contingencies for students. When appropriate, parents, teachers, and students should collaboratively create a template of student behaviors or goals. Approximately three to five target areas or goals should be chosen and can be rated by all teachers who are participating. Goals are individualized for each

student, should generally consist of areas of primary concern for the student, and should be stated in a positive manner. Some examples of classroom goals may include participation in class discussions, appropriate interactions with peers, work completion, and accuracy. Goals can be rated on a yes/no scale (e.g., appropriately participates in class discussion two times, completes work with 85% accuracy) or on a range of scores (e.g., a 5-point scale ranging from "excellent" to "terrible"). The report card ratings should be filled in by each teacher (in ink to prevent alterations by students) and carried from class to class by students. At home, scores on the report card can then be used in a token economy to be exchanged for privileges or points to be applied to later privileges (e.g., television time, computer time, later bedtime). Table II lists components of effective school-home communication programs.

As with classroom-based response cost systems, students should be aware of specific criteria necessary to earn privileges, which privileges they can work toward, and when they may exchange tokens for them. Exchanges for privileges might need to occur on a more frequent basis during the initial stages of a school-home contingency system and should be adapted based on the age and developmental level of each student. Motivation of the students is vital to the success of the school-home contingency system. Therefore, it may be beneficial for students to have input into goals and privileges. In addition, a space for teacher comments and parent

**TABLE II**  
**Components of Effective School-Home Communication Programs**

1. Daily and/or weekly goals are stated in a positive manner.
2. Both academic and behavioral goals are included.
3. A small number of goals are targeted at a time.
4. The teacher provides quantitative feedback about student performance.
5. Feedback is provided by subject or class periods.
6. Communication is made on a regular basis (either daily or weekly).
7. Home-based contingencies are tied to school performance. Both short- and long-term consequences are employed.
8. Parental cooperation and involvement are solicited prior to implementation.
9. Student input into goals and contingencies is solicited, particularly with older children and adolescents.
10. Goals and procedures are modified as necessary.

*Source.* From DuPaul and Stoner (2003). Copyright 2003 by The Guilford Press. Reprinted with permission.

signatures can be provided to allow for frequent brief communications between parents and teachers. Figure 1 shows an example of what a school-home daily report card might look like.

### 3.5. Self-Mediated Interventions

Students can also be taught to monitor their own behavior or completion of academic-related steps. Although each student acts as the primary mediator for such strategies, it is important to note that a teacher must be present to signal recording and/or to monitor the accuracy of the student's self-recording. Checklists for work completion, or for procedures for completing tasks in specific academic areas, can be created. The presence of these checklists while the student is completing work can provide a prompt as to what steps are necessary to complete the task as well as provide for communication between the teacher and the student as to what steps have been completed.

Students can also be taught to observe and record occurrences of their own behavior—both positive and negative. A checklist or grid can be provided for each student to keep on his or her desk. The student could then record occurrences of a behavior or be taught to

respond to a signal (e.g., a beep from a tape recorder) and record whether or not he or she is currently performing a specific behavior (i.e., on task) at that time. In this way, not only is the student monitoring the occurrence of his or her own behavior, but the signal also acts as a prompt to demonstrate the behavior. To be most effective, it might be advantageous to have the teacher also record occurrences. These records can then be “matched” and reinforcement can be provided based on how well the results from the teacher and the student match up. Frequency of teacher-student match checks and reinforcement schedules can then be decreased as the child becomes more adept at the procedures.

### 3.6. Intervention Considerations for Secondary School-Aged Students

Typically, AD/HD symptoms are chronic and can be associated with myriad additional difficulties for secondary school-aged students. Adolescents with AD/HD often exhibit problems with study and organizational skills, test performance, and social and emotional adjustment. Unfortunately, minimal research has been conducted examining school-based interventions for secondary-level students with this disorder. Although

Child's Name: _____	Date: _____				
	<u>Special</u>	<u>Language Arts</u>	<u>Math</u>	<u>Reading</u>	<u>SS/Science</u>
Follows class rules with no more than 3 rule violations per period	Y N	Y N	Y N	Y N	Y N
Completes assignments within the designated time	Y N	Y N	Y N	Y N	Y N
Completes assignments at 80% accuracy	Y N	Y N	Y N	Y N	Y N
Complies with teacher requests (no more than 3 instances of noncompliance per period)	Y N	Y N	Y N	Y N	Y N
No more than 3 instances of teasing per period	Y N	Y N	Y N	Y N	Y N
<u>OTHER</u>					
Follows lunch rules	Y	N			
Follows recess rules	Y	N			
Total number of “yes” responses _____					
Total number of “no” responses _____					
Percentage _____					
Teacher's initials: _____					
Comments:					

FIGURE 1 Sample school-home daily report card. Copyright 2001 by William E. Pelham, Jr., Ph.D. Reprinted with permission.

many of the interventions discussed for children with AD/HD may be helpful for adolescents, modifications must be made to account for expected differences in developmental level and degree of independence. For example, token reinforcement and/or response cost systems must be modified in several ways. First, the adolescent should be involved in “negotiating” the specific responsibilities and privileges that are included in the contingency management system. Second, the system will be less concrete than what typically is used with younger children. In particular, a written contract specifying reinforcement and punishment is used rather than token reinforcers such as stickers and poker chips. Third, the time period between completion of a responsibility and receipt of reinforcement might be longer than with younger children because adolescents are presumably able to handle longer delays in reinforcement. As is the case for younger students, interventions for adolescents with AD/HD should include a variety of mediators (e.g., teachers, parents, computer, peers, self), especially because secondary school students have multiple teachers who might not have the time to implement complex treatment protocols. Self-mediated interventions have particular promise for those adolescents who have demonstrated progress with other-mediated strategies and who appear to have the requisite skills for self-monitoring and self-evaluation.

In addition to the various interventions enumerated previously, secondary school students with AD/HD may require other interventions to address study and organizational skills deficits as well as difficulty in interacting with peers and authority figures. First, students can be provided with direct instruction in study and organizational skills, including taking notes, preparing for and taking tests, and completing long-term assignments. A second promising avenue for intervention for adolescents with AD/HD is the implementation of “coaching” to support students in achieving self-selected academic, behavioral, and social goals. Peers, older students, siblings, or adults can serve as coaches for students, who are assisted in setting goals, developing plans to reach stated goals, identifying and overcoming obstacles to goal attainment, and evaluating progress. Third, adolescents with AD/HD and their families are likely to require some form of counseling support to address possible difficulties in adherence with household rules as well as in interactions with family members. Support may include promoting an accurate understanding of AD/HD and its influence on families’ interaction patterns; planning the education

of adolescents with AD/HD; and/or negotiating privileges and responsibilities.

#### 4. CONCLUSIONS AND FUTURE CONSIDERATIONS

Children and adolescents with AD/HD frequently experience difficulties in school settings, most notably in the areas of academic achievement and interpersonal relationships. Thus, a comprehensive treatment plan often includes a variety of school-based interventions in combination with psychostimulant medication and home-based behavioral strategies. Effective classroom interventions include both proactive strategies that are designed to modify antecedent events and reactive strategies that focus on changing consequent events. Although treatment strategies are usually implemented by teachers, plans should include peer-mediated, self-mediated, computer-mediated, and/or parent-mediated components. Interventions that directly address the function of disruptive behavior and that are applied as close to the point of performance as possible are more likely to be effective. Students with AD/HD at the secondary school level will require additional support to enhance study and organizational skills as well as coaching to improve social relationships. Empirical investigations conducted to date have emphasized the use of reactive, consequent-based interventions. Research efforts must be directed toward establishing efficacious, antecedent-based strategies as well as toward providing more support for interventions to address the unique needs of secondary school-aged students with this disorder.

#### See Also the Following Articles

Attention Deficit/Hyperactivity Disorders (ADHD) ■ Child Development and Culture

#### Further Reading

- Barkley, R. A. (1998). *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd ed.). New York: Guilford.
- Dawson, P., & Guare, R. (1998). *Coaching the ADHD student*. New York: Multi-Health Systems.
- DuPaul, G. J., & Eckert, T. L. (1997). School-based interventions for children with attention-deficit/hyperactivity disorder: A meta-analysis. *School Psychology Review*, 26, 5–27.

- DuPaul, G. J., & Power, T. J. (2000). Educational interventions for students with attention-deficit disorders. In T. E. Brown (Ed.), *Attention-deficit disorders and comorbidities in children, adolescents, and adults* (pp. 607–635). Washington, DC: American Psychiatric Press.
- DuPaul, G. J., & Stoner, G. (2003). *ADHD in the schools: Assessment and intervention strategies* (2nd ed.). New York: Guilford.
- Goldstein, S., & Goldstein, M. (1998). *Managing attention deficit hyperactivity disorder in children: A guide for practitioners* (2nd ed.). New York: John Wiley.
- Greenwood, C. R., Maheady, L., & Delquadri, J. (2002). Classwide peer tutoring programs. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventative and remedial approaches*. Bethesda, MD: National Association of School Psychologists.
- Pelham, W. E. (2001). *Attention deficit hyperactivity disorder: Diagnosis, nature, etiology, and treatment*. (Available: Center for Children and Families, 318 Diefendorf Hall, 3435 Main Street, Buffalo, NY 14214)
- Pelham, W. E., Wheeler, T., & Chronis, A. (1998). Empirically-supported psycho-social treatments for ADHD. *Journal of Clinical Child Psychology*, 27, 190–205.
- Rathvon, N. (1999). *Effective school interventions: Strategies for enhancing academic achievement and social competence*. New York: Guilford.
- Robin, A. L. (1998). *ADHD in adolescents: Diagnosis and treatment*. New York: Guilford.
- Shapiro, E. S., & Cole, C. L. (1994). *Behavior change in the classroom: Self-management interventions*. New York: Guilford.
- Shinn, M. R. (Ed.). (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.
- Watson, T. S., & Steege, M. W. (2003). *Conducting school-based functional behavioral assessments: A practitioner's guide*. New York: Guilford.



# Attention Deficit/Hyperactivity Disorders (ADHD)

Anne Teeter Ellison

University of Wisconsin, Milwaukee, Wisconsin, USA

1. Core Characteristics of Attention Deficit/Hyperactivity Disorder
  2. Associated Problems
  3. ADHD with Other Disorders
  4. Gender and Cultural Issues
  5. Etiology of ADHD
  6. Developmental Context for ADHD
  7. Evidence-Based Interventions for ADHD
  8. Medication Monitoring and Adherence
  9. Risk and Resiliency Factors that Affect ADHD
- Further Reading

*self-management training* Education in methods of regulating a person's behavior (e.g., self-instruction, anger control).

Attention deficit/hyperactivity disorder (ADHD) is a psychiatric disorder affecting 3 to 5% of children. Recent data suggest that the incidence of ADHD may be even higher (5–7%). ADHD is a chronic disorder, with approximately 75 to 80% of affected individuals showing evidence of significant impairment during adolescence and adulthood. Estimates suggest that 1.5 to 2.0% of adults and 2 to 6% of adolescents have ADHD. Research indicates that children with persistent ADHD have more severe symptoms, significant impairment in functioning, stressful family environments, and more adverse risk factors. There are a number of factors that interact with ADHD and further compromise adjustment over the life span.

## GLOSSARY

*behavioral disinhibition* Poor self-regulation or the inability to control one's activity level, attention, and emotions.

*inattention* Difficulties with distractibility, alertness, and arousal as well as with selective, sustained, and persistent attention to tasks.

*negative parental interactions* Parental interactions that are undesirable (e.g., fewer interactions, less affection, higher parental stress) or produce undesirable effects in their children.

*preventive parental measures* Behaviors of social control, including reinforcing prosocial behaviors, training preschool teachers, and using behavioral management principles.

*psychosocial stressors* Conditions in a person's social environment that affect levels of stress (e.g., poverty, family dysfunction).

## 1. CORE CHARACTERISTICS OF ATTENTION DEFICIT/HYPERACTIVITY DISORDER

Core symptoms of inattention, impulsivity, and hyperactivity comprise the major characteristics of attention deficit/hyperactivity disorder (ADHD), with new conceptualizations emphasizing poor self-control and

behavioral disinhibition. Behavioral disinhibition is synonymous with poor self-regulation or the inability to control one's activity level, attention, and emotions. In 1997, Barkley posited that disinhibition interferes with executive control functions, including working memory, internalization of speech to guide one's behavior, motor control for goal-directed behavior, the ability to analyze and synthesize responses, and self-regulation of emotions, motivation, and arousal.

The fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) outlined three subtypes of ADHD: attention deficit disorder/predominantly inattentive type (ADD/PI), attention deficit disorder/predominantly hyperactive-impulsive type (ADD/PHI), and attention deficit disorder/combined type (ADD/C). Inattention is multifaceted and refers to difficulties with distractibility, alertness, and arousal as well as with selective, sustained, and persistent attention to tasks. Individuals with ADD/PI appear to have cognitive disabilities (e.g., spacey, "daydreamy," sluggish, easily confused) that arise from slow information processing and poor focused/selective attention. There is some debate in the literature as to whether ADD/PI is a separate disorder distinct from ADD/PHI or ADD/C. Further research is warranted on this issue, particularly as it relates to gender differences. For example, some studies report that girls have higher rates of inattention than of impulsivity and hyperactivity.

Impulsivity encompasses the inability to inhibit behavior, delay responding, and delay gratification to reach long-term goals and perform tasks. Hyperactivity is one of the more obvious core features of the disorder and refers to excessive activity (both verbal and physical). Individuals with ADD/PHI display primary problems with hyperactivity and impulsivity, whereas those with ADD/C show deficits in all core symptoms, including impulsivity, inattention, and hyperactivity.

## 2. ASSOCIATED PROBLEMS

The core symptoms of ADHD lead to impairment in all aspects of life activities, including school attainment, family adjustment, social relations, occupational functioning, and self-sufficiency. Sleep disorders, health problems (e.g., allergies), and accidental injuries (e.g., bone fractures, poisoning) also appear to be higher in children with ADHD than in controls. More accidental injuries may occur in youth with high levels of aggression than in those with hyperactivity alone. Long-term complications associated with ADHD

include risk of substance abuse problems; problems in marriage, family cohesiveness, and chronic family conflict; employment difficulties such as frequent job changes, stress on the job, and underemployment; increased health risks such as early cigarette smoking, early sexual activity, increased driving accidents and eating disorders, and sleep disorders in children and adolescents; and risk of comorbid psychiatric disorders.

## 3. ADHD WITH OTHER DISORDERS

ADHD places individuals at risk for other psychiatric disorders. Oppositional defiant disorder (ODD, 54–67%) and conduct disorders (20–56%) are among the most frequent comorbid disorders in children due to impulsivity and an inability to follow rules. Conduct disorders are also common in teens (44–50%). The presence of ADHD with conduct problems increases the risk of later difficulties, including drug use and abuse, driving accidents, and additional psychiatric problems. Social problems are common in individuals with ADHD, where intrusive, inappropriate, awkward, and/or ineffective behaviors lead to rejection or strained relations with others. Mood disorders are also high (20–36%), with 27 to 30% having anxiety and 15 to 70% having major depression or dysthymia. Stimulant medication might not be as effective in individuals with ADHD plus anxiety, and side effects might be higher than in nonanxious ADHD groups. Learning disabilities are also frequent, with as many as 20 to 50% of children exhibiting significant learning problems due to difficulties with attention, work completion, and disruptive problems in the classroom.

In 1997, Jensen and colleagues cautioned that clinic-based longitudinal studies may increase the appearance of comorbidity because persons with more severe and comorbid conditions may be more likely to participate in ongoing studies. Rates of psychiatric comorbidity differ somewhat in community-based studies of children with ADHD. For example, the Multimodal Treatment Assessment (MTA) study of youth with ADHD reported comorbidity rates as follows: In the sample of children, 31.8% had ADHD alone, 33.5% had anxiety, 14.3% had conduct disorders, 39.9% had ODD, 3.8% had an affective disorder, and 10.9% had tic disorders.

Adults with ADHD have comorbidity rates of 16 to 31% for major depression, 24 to 35% for ODD, 17 to 25% for conduct disorders, 4 to 14% for obsessive-compulsive disorders, 35% for alcohol dependence or abuse, 24% for substance abuse (i.e., cannabis or other

drugs), 43% for generalized anxiety disorder, and 52% for overanxious disorder. Rates of conduct disorders or antisocial personality disorder are also high in adults with ADHD (22%).

The mechanisms of comorbidity are not well understood, but it appears that ADHD places an individual at risk for other psychiatric disorders and the presence of comorbid disorders interacts with and alters the developmental trajectory and treatment responsiveness of ADHD. In 2003, Pliska and colleagues provided an assessment and treatment approach for children with ADHD and comorbid disorders. It is recommended for addressing more complex cases of ADHD.

#### 4. GENDER AND CULTURAL ISSUES

National reports have highlighted the disparity of diagnosis and treatment for mental illness for girls and non-White children. Despite similarities in the incidence rates of hyperactivity in Black children, service delivery is lower in Black children than in Whites with similar problems. Referral and treatment rates are lower for girls (three times less likely to receive treatment compared with boys) and for Black children (three times less likely to be referred compared with White children).

There is some evidence that gender differences may be a function of the referral source, that is, clinic-referred children versus non-clinic-referred children. In a review of 18 studies on girls with ADHD in 1997, Gaub and Carlson found that nonreferred samples of girls with ADHD, as compared with boys with ADHD, had more intellectual impairments but were less aggressive, were less inattentive, and had lower levels of hyperactivity. However, when compared with clinic-referred samples, girls and boys with ADHD had more similarities than differences on core ADHD symptoms.

Other studies of clinic-referred girls indicate that girls with ADD/C are indistinguishable from boys with ADHD on measures of comorbid disorder, behavioral ratings of core symptoms, psychological functioning, and family history of psychopathology. However, when differences did occur, girls had lower reading scores and higher parent-rated measures of inattention. Furthermore, referred girls were a more extreme sample than were boys, with the former having higher rates of familial ADHD. Girls also showed a more positive response to stimulant medication than did boys. In 1999, Biederman and colleagues found similar results in girls referred to pediatricians and psychiatrists. Girls referred for ADHD were more likely to show conduct problems, mood and

anxiety disorders, a lower intelligence quotient (IQ), and more impairment on social, family, and school functioning than were nonreferred girls. Conduct problems were lower in girls than in boys with ADHD, and this may account for lower referral rates. However, high rates of mood and anxiety disorders suggest the need for comprehensive treatment. It was of particular note that girls in this study also had high rates of substance abuse disorders, including alcohol, drug, and cigarette use, and were at an increased risk for panic and obsessive-compulsive disorders.

In sum, clinic-referred girls with ADHD present with more symptoms than do girls without ADHD and are indistinguishable from clinic-referred boys with ADHD. It is important to note that girls who are referred for ADHD may represent a more severely impaired group than do community-based samples of girls. Further evidence suggests that girls who are in need of treatment for ADHD may be overlooked because they tend to be less disruptive than boys. Additional longitudinal research investigating gender differences would be helpful to resolve these critical issues.

#### 5. ETIOLOGY OF ADHD

Research indicates that problems in behavioral inhibition or self-control are a result of dysfunction in fronto-striatal networks, whereas other brain regions (e.g., basal ganglia that includes the caudate nucleus and cerebellum) are also implicated. The evidence of genetic transmission of ADHD, primarily involving the dopamine systems that innervate frontal-striatal regions, is strong. Studies estimate that 70 to 95% of deficits in behavioral inhibition and inattention are transmitted genetically. Research investigating the manner in which the environment interacts with subtle brain anomalies and genetic mutations is ongoing. Traumatic events, the presence of comorbid disorders, and other psychosocial stressors (i.e., poverty, family dysfunction) complicate ADHD but are not considered to be causal.

Although neurological and genetic substrates appear to be compromised, multiple interacting factors are likely involved in the expression of ADHD. It is likely that compromised neural systems influence adaptive functioning and that family, school/work, and community environments affect how ADHD is manifested and may contribute to the development of various coexisting disorders. There may be other factors, including exposure to environmental toxins (e.g., elevated lead exposure), prenatal smoking, and alcohol use, that

increase the risk for ADHD. However, these risk factors are not present in all children with ADHD. Some environmental explanations are inaccurate and non-scientifically based, including high sugar ingestion, allergies or sensitivities to foods, family discord, parental alcoholism, poor or ineffective parenting, and poor motivation. Furthermore, inaccurate beliefs about the nature of ADHD often lead to ineffective treatment approaches.

## **6. DEVELOPMENTAL CONTEXT FOR ADHD**

Although ADHD has been considered to be a disorder of childhood, there is compelling evidence that a majority of children do not outgrow ADHD. This section highlights the major challenges of ADHD throughout the life span.

### **6.1. ADHD During Early Childhood**

Symptoms of ADHD typically first appear during early childhood. Infants are often described as temperamental, difficult to care for due to excessive crying and irritability, difficulty in calming, overly sensitive to stimulation, and overly active. These complications often interfere with normal parent–child bonding and often lead to negative parental interactions (e.g., fewer interactions, less affection, higher parental stress). These patterns affect child compliance and lead to frustrating and challenging interactions. In preschool, hyperactivity levels are pronounced and lead to difficulty in adjusting to expectations to sit, listen, and get along with other children. Impulsivity interferes with play and often leads to rejection. Referral rates are high during this stage of development as children come into contact with other adults and face greater demands for self-control. Interventions at the stage frequently focus on increasing parenting skills, building positive parent–child relationships, increasing parent support, and implementing other preventive measures (e.g., reinforcing prosocial behaviors, training preschool teachers, using behavioral management principles).

### **6.2. ADHD During Middle Childhood**

Most research on ADHD has been conducted on children between 6 and 12 years of age. Deficits in self-control continue to be problematic and are highlighted by disruptive, noncompliant, and off-task behaviors at home and at school. Poor attention to schoolwork,

poor work completion, low motivation, low persistence to challenging tasks, and poor organizational skills negatively affect academic and school adjustment. Difficulties with social situations, negative peer and adult interactions, poor anger control, and low self-esteem create secondary problems that can be chronic. Parent–child relationships are often strained due to noncompliance, failure to complete household chores, and the need for constant monitoring of everyday activities (e.g., bathing, eating, getting dressed, going to bed). These difficulties can increase family stress and interfere with sibling relationships. Other comorbid disorders, including oppositional deviance, conduct problems, depression, and anxiety, may also emerge during this stage. Severe oppositional deviance is problematic and often presages antisocial behavior during later adolescence and adulthood. Treatments during this stage typically are multimodal, including parent training, behavior classroom management, academic interventions, self-management training (e.g., self-instruction, anger control), and medication.

### **6.3. ADHD During Adolescence**

Longitudinal studies reveal significant difficulties for approximately 70 to 80% of teens who had ADHD as young children. Although cognitive deficits and learning disabilities are common in children with ADHD, they are less well documented in adolescents and adults. Longitudinal studies show that youth with ADHD have significant academic difficulties, including high suspension rates (46%), high dropout rates (10%), and placements in special education for learning disabilities (32.5%), emotional disturbance (35.8%), and speech language disorders (16.3%). Negative academic outcomes were present even after intensive treatments, including medication, individual therapy, family therapy, and special education placement. In general, children with hyperactivity are less well educated, have higher rates of grade retention, and have lower grades compared with controls at 5- and 10-year follow-ups. Adolescents with ADHD also show higher rates of automobile accidents and speeding tickets than do teens without ADHD. Both cigarette use and marijuana use are higher according to parental reports, whereas teens with ADHD report higher rates of cigarette use but not of alcohol use, or of marijuana, cocaine, heroin, and other illegal substances use, compared with non-ADHD teens. Antisocial behaviors, including theft, breaking and entering, disorderly conduct, carrying a weapon, assault with a weapon, assault with fists, setting fires,

and running away from home, were reported in the Milwaukee Longitudinal Study.

In general, treatment options for teens are less well researched than are those for children. Treatment for adolescents with ADHD typically focuses on increasing problem solving and communication between parents and teens, psychopharmacotherapy, and classroom accommodations for academic difficulties. More systematic study is needed to investigate the strength of these various interventions. It has been suggested that children with ADHD might not receive needed treatment.

#### **6.4. ADHD During Adulthood**

In 1996, Barkley and colleagues found that adults with ADHD had similar levels of educational achievement and occupational adjustment but differed from controls on symptoms of ADHD and oppositional problems in college and at work. They also had shorter duration of employment, more psychological distress and maladjustment, and more antisocial acts and arrests for disorderly conduct and thefts compared with controls. Even though conduct problems and risk for comorbid antisocial personality disorder appear in approximately 25% of individuals with ADHD, the majority of adults with hyperactivity do not engage in criminal behaviors.

According to driving instructors and self- and parent-reports of driving skills, young adults with ADHD were more distractible and impulsive while driving. High rates of driving-related difficulties, including license suspensions or revocations, serious accidents (i.e., involving a wrecked car), and hit-and-run accidents, were also reported. In 1998, Barkley indicated that young adults with ADHD had sexual intercourse at an earlier age, more sexual partners, and higher rates of pregnancy. Contraceptive use was lower, sexually transmitted diseases were higher, and testing for HIV/AIDS was higher in the ADHD group than in controls.

Treatments for adults with ADHD are not well documented but often include multiple approaches, including family and couples therapy, occupational and career counseling, occupational accommodations, medication, and treatment for comorbid disorders (e.g., alcohol or drug treatment, depression, bipolar disorders). Others have emphasized the need for counseling to change the negative mind-set that results from years of failure and coaching for everyday responsibilities. Although empirical studies are needed to determine the efficacy of various treatment options, studies do show the efficacy of stimulant medication in the treatment of ADHD in adults.

### **7. EVIDENCE-BASED INTERVENTIONS FOR ADHD**

The American Academy of Child and Adolescent Psychiatry (AACAP) has developed practice guidelines for the diagnosis and treatment of ADHD. The American Academy of Pediatrics (AAP) recommends that stimulant medication should not be used as the only treatment for ADHD and should be administered only after a careful evaluation. Practice guidelines recommend a comprehensive multimethod approach for the diagnosis of ADHD in children and youth.

#### **7.1. Multimodal Treatment Regimens**

The MTA study, funded by the National Institute for Mental Health, reported that children with ADHD received suboptimal care in the community. Even though two-thirds of the sample received stimulant medication, care in the community was less effective than were carefully managed medication, behavioral treatment, and combined treatments. Only 25% of children receiving care in the community were normalized after a 14-month trial, whereas 68% of the combined group, 56% of the carefully managed medication group, and 34% of the behavioral treatment group showed normalization. When community care is provided, it is not carefully monitored, nor is it as effective as multimodal intensive treatment. The MTA study showed that children with ADHD had the best response to multimodal treatment that included 35 sessions of parent training, a full-time summer treatment program for children to learn social and sports skills and to practice academic skills, weekly teacher consultation, a paraprofessional aide in the classroom, contingency management in the classroom, and medication. This extensive treatment was highly effective in reducing the major symptoms of ADHD and was superior to treatment generally found in the community. It is difficult to discern whether the quality and level of treatment described in the MTA study can be easily implemented in the community.

#### **7.2. Other Evidence-Based Interventions**

Other empirically supported treatments for children with ADHD include behavioral therapy and contingency management techniques; a summer treatment program with a systematic reward/response cost program, sports skills

training, a 1-hour daily academic special education class, training in effective social skills, daily report cards, and parent training; parent training combined with contingency management and didactic counseling to increase parent knowledge of ADHD; a community-based family therapy program; the good behavior game, response cost, using the “attention trainer”; modification of classroom assignments and task demands; and the Irvine Paraprofessional Program. Self-management, direct contingency management, and intensive behavioral and social skills training have also been shown to be effective.

## **8. MEDICATION MONITORING AND ADHERENCE**

Research on the short-term efficacy of stimulant medication is well documented for 75 to 80% of children with ADHD; however, medication monitoring and adherence is problematic. For example, in 2001, Vitiello stated, “For optimal pharmacological treatment of children with ADHD, medication adjustments are needed for long-term treatment even when the initial dose is chosen in a careful, comprehensive, and unbiased manner.” In the MTA study, more than 70% of children assigned to the medical management group were on different doses after a 13-month trial. Although the majority of children receiving community care in the MTA study were treated with stimulant medication, as a whole, children in the medical management group received more careful medication monitoring from physicians and showed greater improvement of symptoms than did children in the care in the community group.

In 2001, Thiruchelvam and colleagues investigated medication adherence in children 6 to 12 years of age. The most salient factors affecting compliance were the absence of ODD, the severity of ADHD symptoms, and the age of the children. Children with ODD were 11 times more likely to refuse medication. Youth with more symptoms were more responsive to the medication and were more compliant when taking medication as well. Positive stimulant response may encourage parents and children to stick to the medication regimen at higher rates than in cases where medication is not very helpful. Older children were also more likely to refuse medication. Because there is a decrease in hyperactivity symptoms with age, older children may perceive less benefit from medication and choose not to adhere. Social stigma may also play a role in adherence at this stage. Physicians are advised to develop adherence plans for youth on medication.

In sum, research indicates that multiple therapies are needed to adequately address the problems associated with ADHD. Stimulant medication with behavioral and psychosocial interventions, including classroom behavior management and parent management training, improve ADHD symptoms and associated problems in children with ADHD. Initial research indicates that stimulant medication is effective for adults and that other cognitive-behavioral interventions (e.g., self-management) show promise. Currently, treatment within a single modality (medication vs behavioral) appears to have very little long-term impact. Effective treatments are less well documented for adolescents and adults with ADHD. Although studies have shown that stimulant medication is effective, less is known about the effects of multimodal treatment in older groups.

## **9. RISK AND RESILIENCY FACTORS THAT AFFECT ADHD**

Although ADHD presents challenges throughout the life span, some factors complicate the disorder, whereas others appear to be protective. Risk factors that alter the course of ADHD include child characteristics (e.g., severity of ADHD symptoms, intelligence levels, comorbidity), family discord or environmental distress, and early treatment for ADHD and the presence of coexisting disorders. In an effort to optimize outcome, comorbid disorders should be targeted for treatment along with the ADHD symptoms. Oppositional defiant behavior problems are among the most debilitating difficulties over time because they often lead to conduct disorders and antisocial personality disorders during adolescence and adulthood. There is evidence that effective parenting skills can interrupt this progression in many children. Other family factors that increase the complexity of ADHD include parental psychopathology such as maternal depression and paternal antisocial personality disorder. These parental difficulties often interfere with effective parenting and produce added stress to vulnerable families whose members are already challenged by disruptive noncompliant child behaviors. There is strong evidence that raising a child with ADHD is stressful, so additional parental problems make this challenge even more overwhelming. Furthermore, as noted by Goldstein in 2002, “Living in a household, above the poverty level, with parents who are free of serious psychiatric problems, consistent in their parenting style, and available to their children appear to be among the most

powerful variables at predicting good outcome.” Parents are also advised to seek individual and family therapy for their own problems in an effort to strengthen interpersonal effectiveness and family cohesiveness.

The extent to which treatment alters the developmental course of ADHD is not well understood. Some studies of children who received extensive treatment showed that these youth still had poor outcomes during adolescence. However, the MTA study showed short-term improvement (over a 14-month period) in children receiving comprehensive multimodal treatment in highly controlled and monitored programs. Furthermore, growing evidence suggests that medication may buffer some of the negative effects of ADHD. Although there are reasons to be cautious about the use of stimulant medication in young children, recent studies suggest that early treatment may alter the neurodevelopmental pathways of ADHD in positive ways. Stimulant treatment also appears to improve outcomes for adults. In 1999, Wilens and colleagues also found that 70% of adults receiving a year of cognitive therapy and stimulant medication showed a reduction of ADHD symptoms and were less anxious and depressed.

Current studies of ADHD are focusing on the long-term effects of multimodal treatment for children and are investigating what works best for adolescents and adults with ADHD. Current research is promising, but researchers are still exploring how the outcome and course of ADHD can be altered with effective treatments. The impact of environmental events on the development of attention and self-regulation is of particular interest.

### See Also the Following Articles

Child Development and Culture ■ Educational and Child Assessment ■ Family and Culture

### Further Reading

- Barkley, R. A. (1997). *ADHD and the nature of self-control*. New York: Guilford.
- Barkley, R. A. (1998). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. New York: Guilford.
- Barkley, R. A., & Gordon, M. (2002). Research on comorbidity, adaptive functioning, and cognitive impairments in adults with ADHD: Implications for a clinical practice. In S. Goldstein, & A. Teeter Ellison (Eds.), *A clinician's guide to adult ADHD: From evaluation to treatment* (pp. 43–69). San Diego: Academic Press.
- Barkley, R. A., Murphy, K., DuPaul, G., & Bush, T. (2002). Driving in young adults with attention deficit hyperactivity disorder: Knowledge, performance, adverse outcomes, and the role of executive functioning. *Journal of the International Neuropsychological Society*, 8, 655–672.
- Castellanos, F. X., Lee, P., Sharp, W., Jeffries, N., Greenstein, D., Clasen, L., Blumenthal, J., James, R., Ebens, C., Walter, J., Zijdenbos, A., Evans, A., Giedd, J. N., & Rapoport, J. L. (2002). Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Medical Association*, 288, 1740–1748.
- DuPaul, G., & Stoner, G. (2003). *ADHD in the schools: Assessment and intervention strategies* (2nd ed.). New York: Guilford.
- Gaub, M., & Carlson, C. L. (1997). Gender differences in ADHD: A meta-analysis and critical review. *Journal of the Academy of Child and Adolescent Psychiatry*, 36, 1036–1045.
- Goldstein, S. (2002). Continuity of ADHD in adulthood: Hypothesis and theory meet reality. In S. Goldstein, & A. Teeter Ellison (Eds.), *Clinician's guide to adult ADHD: Assessment and intervention* (pp. 25–45). London: Academic Press.
- Leibson, C. L., Katusic, S. K., Barbaresi, W. J., Ransom, J., & O'Brien, P. C. (2001). Use and cost of medical care for children and adolescents with and without attention-deficit/hyperactivity disorder. *Journal of the American Medical Association*, 285, 60–66.
- MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 56, 1073–1086.
- Pliszka, S. R., Carlson, C. L., & Swanson, J. (2003). *ADHD with comorbid disorders: Clinical assessment and management*. New York: Guilford.
- Tannock, R. (1998). Attention deficit hyperactivity disorder: Advances in cognitive, neurobiological, and genetic studies. *Journal of Child Psychology and Psychiatry*, 39, 65–100.
- Teeter, P. A. (1998). *Interventions for ADHD: Treatment in developmental context*. New York: Guilford.
- Teeter Ellison, P. A. (2002). An overview of childhood and adolescent ADHD: Understanding the complexities of development into the adult years. In S. Goldstein, & A. Teeter Ellison (Eds.), *Clinician's guide to adult ADHD: Assessment and intervention* (pp. 2–19). London: Academic Press.
- Thiruchelvam, D., Charach, A., & Schachar, R. J. (2001). Moderators and mediators of long-term adherence to stimulant treatment in children with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 922–928.
- Vitiello, B. (2001). Psychopharmacology for young children: Clinical needs and research opportunities. *Pediatrics*, 108, 983–989.
- Wilens, T. E., McDermott, S. P., Biederman, J., & Abrantes, A. (1999). Cognitive therapy in the treatment of adults with ADHD. *Journal of Cognitive Psychotherapy*, 13, 215–226.



# Attitude Measurement

David Trafimow

New Mexico State University, Las Cruces, New Mexico, USA

1. Introduction
  2. Two Meanings of Reliability
  3. Two Meanings of Validity
  4. The Historical Problem: Attitudes Did Not Predict Behaviors
  5. Factor Analysis
  6. Attitudes Toward Objects or Behaviors and Behavioral Categories
  7. Attitude Specificity, Behavioral Intentions, and Behaviors
  8. Open Attitude Measures
  9. Implicit Attitude Measures
  10. Direct and Indirect Attitude Measures
  11. Distinctions That Are Important for Attitude Measurement
  12. Consumer Applications
  13. Conclusion
- Further Reading

## GLOSSARY

**accessibility** The ease with which a memory can be activated; it is often measured by the time it takes a person to react to a stimulus.

**action** What a person is actually going to do or not do; it is usually the verb in the attitude measure.

**associative hypothesis** The hypothesis that a person forms associations between behavioral beliefs and other behavioral beliefs, between normative beliefs and other normative beliefs, but not between behavioral beliefs and normative beliefs.

**construct validity** The extent to which a measure of a variable actually measures what it is supposed to measure.

**context** The situation in which a person will perform or not perform the behavior of concern.

**factor analysis** A statistical technique that reduces a large number of test items or scales to a smaller number of dimensions.

**internal consistency** The extent to which different scales that comprise a measure correlate with each other.

**open measurement** A way in which to measure repeated behaviors that may increase the validity of the measures of corresponding attitudes.

**predictive validity** The extent to which a measure of a variable predicts the measure of another variable.

**principle of correspondence** The extent to which an attitude and behavior measure match each other with regard to action, target, time, and context.

**repeated behavior** A behavior that is performed several times or more; such a behavior presents a challenge for the valid measurement of its corresponding attitude.

**skewed distribution** A distribution that deviates from normality as a result of more extreme scores at one tail than at the other tail; response latencies are notorious for having extreme scores at the tail that represents slow latencies (i.e., long times).

**target** The object that is or is not acted on by a person whose attitude is being measured.

**test-retest reliability** The correlation between a test and itself when it has been taken at two different times.

**time** When the behavior will or will not be performed by a person whose attitude is being measured.

Because attitudes have been assumed to be predispositions for behaviors, consumer researchers

have assumed that attitudes should predict behaviors. However, the success of predicting behaviors from attitudes largely depends on how reliably and validly attitudes are measured. The validity of attitude measures, in turn, depends on a variety of issues, the most important of which is correspondence of measurement. Other concerns are whether one is measuring attitudes toward objects or toward behaviors, whether one wishes to obtain implicit or explicit attitude measures, whether the attitude pertains to a single behavior or to repeated behaviors, and whether attitudes can be distinguished from other variables.

## 1. INTRODUCTION

The vast majority of applied researchers who measure attitudes do so because they wish to predict or affect people's behaviors. Health psychologists wish to increase healthy behaviors such as exercise, healthy eating, and use of condoms; consumer researchers wish to induce people to buy products; and politically motivated researchers wish to predict and affect people's votes and financial contributions. Researchers usually assume that attitudes are an important determinant of behaviors, with the accompanying assumption that knowledge of people's attitudes implies an ability to predict and control behaviors. To obtain such knowledge of people's attitudes, however, it is absolutely crucial to be able to measure them reliably and validly—the main topic of this article.

## 2. TWO MEANINGS OF RELIABILITY

An attitude measure can be reliable in one of two general ways. To understand the simplest type of reliability, suppose that a set of participants completes an attitude measure at a particular time and then completes it again at a later time. The correlation between the two test-taking occasions is the attitude measure's test–retest reliability. However, there is a second meaning of reliability in that the term can be used to denote internal consistency. To understand internal consistency, suppose that an attitude measure is composed of several scales. If these scales truly measure the participants' attitudes, all of the scales should correlate with each other. Another way of saying that the scales correlate with each other is to say that they have a high degree of internal consistency. Although there are different formulas for calculating different kinds of internal consistency, all of these formulas

require an assumption that all of the scales that compose the measure are equally good. Based on this assumption, all of the formulas imply that increasing the number of scales increases the internal consistency of the whole attitude measure. Because both test–retest reliability and internal consistency are usually considered to be prerequisites for validity, it is common practice to use several scales to measure attitudes.

## 3. TWO MEANINGS OF VALIDITY

How does one know whether a particular attitude measure is valid? To answer this question, one must first ask “valid for what?” An attitude measure can be valid if it predicts another variable that is considered desirable to predict such as behaviors and behavioral intentions (i.e., motivations to perform behaviors). For example, a measure of condom use attitudes that actually predicts whether people will intend to use condoms, or whether they actually use condoms, can be said to have predictive validity.

Another answer to the “valid for what?” question is that the measure must actually measure attitude and not something else. It is possible that a particular attitude measure might predict behavior, not because it measures attitude but rather because it measures something else that is correlated with behavior. Researchers who are concerned with this are forced to deal with the issue of how to define an attitude. Not surprisingly, various researchers have used different definitions, depending on their theories. From this perspective, the valid measurement of attitudes depends on the theories in which attitudes play an important role. Empirically obtained evidence in support of a particular theory also supports the validity of the attitude measurement as it is specified by that theory, whereas empirical evidence against a particular theory tends to disconfirm the validity of the attitude measurement as it is specified by that theory. When the theory in which the attitude construct is embedded is empirically supported, the attitude measure that is part of that theory is said to have construct validity.

## 4. THE HISTORICAL PROBLEM: ATTITUDES DID NOT PREDICT BEHAVIORS

Since the 1920s, researchers have assumed that attitudes are predispositions for behaviors. If this is true, it

follows that attitudes should predict behaviors. However, the vast majority of researchers from the 1920s through the 1960s obtained correlations that were either low or not statistically significant. The consistent inability of researchers to predict behaviors from attitudes suggested one of two possibilities. First, perhaps attitudes really do not matter very much for predicting behaviors. Second, perhaps attitudes do matter, despite the lack of empirical support, but the attitude measures used in previous studies were not valid. Although many researchers argued for the first possibility, other researchers eventually achieved a good deal of success by assuming the second possibility. These latter researchers redefined the idea of attitude in two ways, both of which resulted in attitude measures with greater construct and predictive validity.

#### 4.1. Attitude Accessibility

One view derives from the social cognition tradition, but it does not receive much attention here because it has not been used much by applied researchers. Fazio used the idea of attitude accessibility to explain low correlations between attitudes and behaviors. He assumed that attitudes affect behaviors, but only when they are easy to retrieve from memory. When attitudes are not accessible, there is no reason to believe that they would affect behavior. Therefore, researchers who subscribed to this view explained previous findings of low attitude–behavior correlations by assuming that they were due to a failure on the part of researchers to measure accessible attitudes. In support of the accessibility view, much research during the 1980s and 1990s indicated that accessible attitudes are more predictive of behaviors than are less accessible attitudes.

Nevertheless, applied researchers have, for the most part, ignored the accessibility view. The reason for this is not clear. One possible reason is that many applied researchers are not particularly knowledgeable about this view. Another possible reason is that researchers who have favored the accessibility view have not been clear about how this view can be applied to predict and control behaviors.

#### 4.2. Principle of Correspondence

Fishbein proposed the principle of correspondence, which provided the other solution to the problem of low attitude–behavior correlations. His main assumption is that behaviors have four components: action, target, time, and context. It is easiest to understand this

idea with an example. Suppose that a researcher wishes to predict whether people will give blood at the campus blood drive on Tuesday. The action is “give,” the target is “blood,” the time is “on Tuesday,” and the context is “at the campus blood drive.” If the attitude measure does not correspond with the behavior measure in regard to action, target, time, and context, it is invalid and a respectable attitude–behavior correlation is not likely to be obtained. For example, if a researcher measures people’s attitudes toward “giving blood,” there is no reason to expect that this attitude measure will successfully predict whether people will “give blood at the campus blood drive on Tuesday.” Rather, to predict this behavior, it is necessary to measure people’s attitudes toward “giving blood at the campus blood drive on Tuesday.” This is often done by having people make a check mark on a scale with various options such as the following:

I extremely like/quite like/slightly like  
/neutral/slightly dislike/quite dislike  
/extremely dislike giving blood at the  
campus blood drive on Tuesday.

(Whether people actually performed the behavior could be assessed in a variety of ways such as checking the lists of people who donated at the campus blood drive on Tuesday and simply asking people whether they had given blood at the campus blood drive on Tuesday.)

The correspondence view explained previously obtained low attitude–behavior correlations in a straightforward way. The reason for the low correlations is that the attitude measures used in those studies were not valid because they did not obey the principle of correspondence. Results from two types of research paradigms supported this view. First, researchers measured attitudes and behaviors (or behavioral intentions, which are often assumed to be direct precursors to behaviors) according to the principle of correspondence and obtained much higher correlations than those that researchers had been able to obtain previously. Second, some researchers manipulated the degree of correspondence of attitude and behavior (or behavioral intention) measures experimentally and obtained much higher correlations when there was a high degree of correspondence than when there was not. In combination, the findings provided a very convincing case for the importance of measuring attitudes in accordance with the

principle of correspondence. Therefore, it is now common practice for applied researchers to measure attitudes in this way, and the result has been a dramatically improved ability to predict behavior in a variety of domains such as dieting, exercising, drinking, smoking, voting, using seat belts, using condoms, and getting screened for cervical cancer.

## 5. FACTOR ANALYSIS

As was discussed earlier, increasing the number of scales can increase the internal consistency of the attitude measure. Consequently, many researchers do not stop at having one attitude scale and instead have several scales. For example, in addition to the like-dislike example given previously, participants could respond to scales that include pairs such as wise-foolish, beneficial-harmful, enjoyable-not enjoyable, good-bad, and pleasant-unpleasant. When multiple scales are used to measure an attitude, the mean of a participant's responses to the scales is often taken as the representation of his or her attitude.

Depending on the behavior of interest, it may happen that not all of the scales are equally good for measuring attitude. It may even turn out that, for some behaviors, some of the scales will measure something other than attitude. To test this possibility, many researchers habitually submit the various scales to a factor analysis. Factor analysis is a statistical technique that reduces a large number of items down to a smaller number of underlying dimensions. For example, on intelligence quotient (IQ) tests, a large number of items may be reduced down to dimensions such as verbal ability and mathematical ability. As another example, on personality tests, hundreds of items have been factor analyzed and reduced to five basic factors of personality. Ideally, when various attitude scales are submitted to a factor analysis, one factor that represents attitude should result. Although one factor is often obtained, it sometimes happens that more than one factor is obtained. In this case, the researcher must determine which factor is the "true" attitude factor—a determination that depends on a variety of issues that are too complicated and numerous to discuss here. In addition, obtaining more than one factor can be interpreted to mean that there is more than one component to the attitude. In this case, the answer to the question of which factor represents attitude is that all of them may do so.

## 6. ATTITUDES TOWARD OBJECTS OR BEHAVIORS AND BEHAVIORAL CATEGORIES

The principle of correspondence implies a distinction between attitudes toward objects and attitudes toward behaviors. Suppose that a researcher is interested in using advertising to reduce discrimination and wishes to measure relevant attitudes. Unfortunately, it is not clear what the relevant attitudes are. In general, researchers in this area have measured attitudes toward the groups of interest such as women, African Americans, Jews, handicapped people, and people with AIDS. However, the principle of correspondence suggests that if one wishes to predict discriminatory behavior, this is not going to work. This is because although a measure of attitudes toward a particular group includes the target component of the behavior to be predicted, it does not include the action, time, and context components. If one wishes to predict a particular discriminatory behavior, it is necessary to use an attitude measure that includes all of the components of the criterion behavior. At the very least, this implies that the attitude measure will have to include the action component, which makes it a measure of an attitude toward a behavior rather than a measure of an attitude toward an object. For example, an attitude toward "hiring women" is an attitude toward a behavior, whereas an attitude toward "women" is an attitude toward an object. For a more mundane example, one could measure an attitude toward "Skippy peanut butter" or toward "buying Skippy peanut butter"; the former is an attitude toward an object, whereas the latter is an attitude toward a behavior.

An argument has been made against using the principle of correspondence in some situations. In brief, the argument is that researchers are sometimes interested in a large number of behaviors. For example, a researcher might be interested in a wide range of behaviors that discriminate against women and not just hiring behaviors. Obviously, an attitude measure toward a single behavior is likely to be inadequate for predicting a variety of discriminatory behaviors. One solution to this problem is for the researcher to measure attitudes toward all of the discriminatory behaviors that are of interest and to predict each discriminatory behavior from its corresponding attitude measure. Another solution is for the researcher to list a set of discriminatory behaviors of interest and then lump all of them under the general category of "discriminatory behaviors." Participants can then give their attitudes toward the

whole behavioral category, and these can then be used to predict the behaviors in that category. Although there is some evidence that either of these solutions may have some validity, there is insufficient evidence for a strong conclusion. However, there is strong evidence that attitudes toward behaviors are better predictors of behaviors than are attitudes toward objects.

### **7. ATTITUDE SPECIFICITY, BEHAVIORAL INTENTIONS, AND BEHAVIORS**

Attitudes have been shown to be good predictors of behaviors, but they have also been shown to be good predictors of people's intentions to perform behaviors. In addition, such behavioral intentions have often been shown to be good predictors of behaviors, and most researchers believe that intentions are proximate causes of behaviors. This combination of empirical findings and theorizing in the area, along with the practical point that behavioral intentions tend to be much easier to measure than are real behaviors, has resulted in the widespread use of behavioral intentions as a substitute for behaviors. This substitution has resulted in two controversies that have not yet been settled. First, there is a controversy about whether behavioral intentions are close enough to real behaviors to justify using the former as a substitute for the latter. A large part of this controversy is based on the issue of how well behavioral intentions predict behaviors. To the extent that intentions do a good job of predicting behaviors, the substitution would be supported, whereas a lack of prediction would support the reverse conclusion. In fact, results vary widely. In general (but there are exceptions), behavioral intentions do a better job of predicting behaviors when the intention and behavior measures conform to the principle of correspondence than when they do not. Thus, the substitution of behavioral intentions for behaviors is most acceptable when one uses correspondent measures.

The use of behavioral intentions as a criterion measure brings up a second controversy. To understand the underlying reason for the controversy, consider that all behaviors are performed with a target, an action, a time, and a context, meaning that the measured behavior will automatically have these four elements. Therefore, the trick to obeying the principle of correspondence is to make sure that the attitude measure specifies all of these elements in a way that matches the behavior measure. But matters change when attitudes are used to predict

behavioral intentions rather than actual behaviors. A behavioral intention does not necessarily have a target, an action, a time, and a context, as is illustrated by the following example. Imagine the behavior of "buying a television set." For a person to actually perform this behavior, he or she must buy a particular brand of television set, at a particular time, at a particular store or from a particular Web site. But for this person to intend to perform the behavior, many of these elements need not be specified. In the case of this example, the intention specifies a rather vague target (television set rather than a specific brand of television set) and an action (buy) but not a time or a context. Therefore, to have an attitude measure that is correspondent with the intention measure, it too should specify the target (in a similarly vague way) and the action but not a time or a context. Consequently, when one attempts to predict behavioral intentions from attitudes, it is possible to fully specify target, action, time, and context or to not fully specify these four elements. In either case, the principle of correspondence is obeyed so long as the degree of specification of the four elements is the same for both the attitude and behavioral intention measures. Does the degree of specification affect the size of the obtained correlation between the attitude and behavioral intention measures? Currently, there is insufficient empirical evidence to answer this question. Moreover, even for the prediction of actual behaviors, it is possible for the researcher to be uninterested in a particular element (e.g., the store at which the television set was bought). In this case, the behavior measure can ignore the uninteresting element or not ignore it, and the attitude measure can ignore it or not ignore it, and so long as the two measures match, the principle of correspondence is obeyed. Will the prediction of the behavior from the attitude measure be greater if the uninteresting element is specified? Again, there is insufficient empirical evidence to know. In sum, whether one predicts behavioral intentions or behaviors, whether one predicts clear-cut behaviors or behavioral categories, and whether one is interested in more of the elements (e.g., target, action, time, context) or fewer of them, there is insufficient research to determine whether it is better to have measures that are as specific as possible or not.

### **8. OPEN ATTITUDE MEASURES**

Although there is widespread agreement that it is crucial to measure attitudes in accordance with the principle of

correspondence, it has not always been clear how to do this. In particular, it has not always been clear how to measure attitudes toward repeated behaviors in accordance with this principle. To see the problem, suppose that a researcher is interested in predicting exercise behaviors from attitudes toward exercising. If the researcher were interested in only one performance of an exercise behavior, it would be easy to obey the principle of correspondence as follows:

*Attitude measure:* I like/dislike to engage in vigorous physical activity at least one time during the month of October. (Participant makes a check mark on a scale.)

*Behavior measure:* I engaged in physical activity at least one time during the month of October. (Participant confirms or disconfirms that he or she performed the behavior.)

Unfortunately, this method of measuring attitudes and behaviors does not work well for repeated behaviors. Suppose that a researcher is interested in regular exercise. There is no clear dividing line between “regular” and “not regular” exercise; therefore, it is not satisfactory to arbitrarily specify a number of exercise occasions that qualifies as regular. Courneya proposed the idea of open attitude and behavior measures to address this problem. His idea was to have the participant, rather than the researcher, specify the number of behaviors. Although the use of open measures has not been tested exhaustively (and the tests have focused more on predicting behaviors from intentions than from attitudes), preliminary evidence suggests that it results in an increase in the prediction of behaviors compared with other methods. The following is an example of open measures:

*Attitude measure:* I like to engage in vigorous physical activity \_\_\_\_\_ times during the month of October.

*Behavior measure:* I engaged in vigorous physical activity \_\_\_\_\_ times during the month of October.

To see why open measures often improve the prediction of behaviors from attitudes, consider an example. Suppose that a person believes that exercising 16 days during the month of October is “regular exercise,” whereas exercising 15 days or less is not. In addition, suppose that the person has a positive attitude toward regular exercise but that he or she exercised only 15 days during October. With traditional measures, the person’s behavior would seem to be inconsistent with his or her attitude; exercising only 15 days is inconsistent with a positive attitude toward regular exercise

(16 days). In contrast, using open measures, it can easily be seen that exercising for 15 days is quite consistent with the person’s positive attitude toward exercising for 16 days.

## 9. IMPLICIT ATTITUDE MEASURES

Attitude researchers have often found it useful to distinguish between implicit and explicit attitude measures. Explicit attitude measures are usually characterized as being conscious, deliberative, and controllable, whereas implicit attitude measures are usually characterized as being unconscious, unintentionally activated, and not controllable. Whether the distinction between explicit and implicit measures should be considered to be a dichotomy or a continuum is not a settled issue, although many researchers who favor implicit measures seem to favor a dichotomous interpretation.

There are several examples of implicit attitude measures. One kind of implicit measure makes use of response latency, where participants respond to a computer-presented stimulus by pressing a key, and faster reaction times are interpreted to indicate a stronger attitude toward the stimulus. Other kinds of implicit measures include memory tasks, physiological measures (e.g., galvanic skin response, heart rate), and indirect questionnaires (where participants are not asked directly about the attitude object or behavior).

Implicit measures, when compared with explicit ones, have both advantages and disadvantages. Some advantages are as follows. First, if the issue of interest is a socially sensitive one, it may be difficult to get honest responses through explicit measures, and so implicit measures provide a way in which to avoid this difficulty. A finding that supports this argument is that implicit and explicit attitude measures have been found to be more highly correlated for socially insensitive attitude objects than for socially sensitive ones. Second, if the researcher is interested in unconscious attitudes, a case can be made that implicit measures are more valid than explicit ones. Supporting evidence indicates that implicit attitude measures tend to be more predictive of implicit measures of other variables than do explicit attitude measures. Third, it is possible that implicit measures are more direct than explicit ones. This is because explicit attitude measures must be funneled through the conscious processing system, whereas implicit ones do not.

On the other hand, implicit measures also have disadvantages. First, if one is interested in predicting

behavior, it is reasonably clear how to use the principle of correspondence to obtain valid explicit attitude measures, whereas it is less clear how the principle of correspondence can be used for implicit attitude measures. Second, the reliability of implicit attitude measures has not been thoroughly investigated, and the preliminary evidence suggests that they have less test-retest reliability than do explicit measures. Because reliability is a precursor to validity, this preliminary evidence also suggests that implicit measures may also be less valid than explicit ones. Third, some of the same studies that suggest that implicit attitude measures are better than explicit ones for predicting implicit measures of other variables also suggest that implicit attitude measures work less well than explicit ones for predicting explicit measures of other variables (e.g., behavioral intentions). Finally, the presumed advantages for implicit attitude measures depend on assumptions about the nature of consciousness, cognitive resources, controllability, and others that, although supported in the literature, have not been proven beyond a reasonable doubt.

Once a researcher has decided to use an implicit measure such as response latency, there is a further issue of how to characterize the central tendency of the person's responses to repeated exposures to the attitude object. Although many researchers have used mean scores, there are problems with this. For one, response latency distributions tend to be highly skewed in the direction of longer response latencies. This is because a large number of factors, such as blinking, distraction, and daydreaming, can increase response latencies. Consequently, mean scores are likely to present a distorted view of participants' actual central tendencies.

Several solutions to this problem have been proposed. The simplest solution is to use median, rather than mean, response latencies as the attitude measure. Another solution is to perform a reciprocal transformation of the data to reduce their skewness [ $1/x$ ,  $1/(x + 1)$  if any of the response latencies are less than 1 second, where  $x$  is the raw response latency]. A third solution is to perform a logarithmic transformation, which has the consequence of bringing the tail involving slower latencies closer to the center of the distribution. An additional recent solution involves a transformation of all of the latency data to  $z$  scores, and then mean or median  $z$  scores can be used as the attitude measure. In sum, there are several ways in which to deal with the problem of skewed latency distributions, and the choice of which method to use depends on considerations that are too numerous and complex to describe here.

## 10. DIRECT AND INDIRECT ATTITUDE MEASURES

Several expectancy-value attitude theories were proposed during the 1950s and 1960s. According to these theories, attitudes are a function of people's assumptions about the probability of various consequences arising from the performance of a behavior and evaluations of how good or bad those consequences are. Although the theories differ in the precise ways in which people are postulated to combine subjective probabilities and evaluations of consequences, they nevertheless have a common implication. Because attitudes are caused by a combination of beliefs about consequences and evaluations, this combination can be used as an indirect attitude measure. Consistent with expectancy-value theories, a large number of findings indicate that indirect and direct attitude measures are highly correlated, at least when the measures are created in accordance with the principle of correspondence. Consequently, many researchers have used indirect attitude measures, rather than direct ones, to predict other variables of interest, notably behavioral intentions and behaviors.

There has been some controversy about whether direct or indirect attitude measures are better predictors of behavioral intentions or behaviors. Although indirect measures have the advantage of specifying the subjective probabilities and evaluations of consequences that determine attitudes, there are both theoretical and empirical reasons to prefer direct attitude measures if one wishes to predict behavioral intentions or behaviors. The theoretical reason stems directly from the assumption that indirect measures assess variables (subjective probabilities and evaluations) that determine attitudes that, in turn, are a determinant of behavioral intentions or behaviors. According to this reasoning, attitudes are a more proximal cause of behavioral intentions or behaviors than are subjective probabilities and evaluations. Under the assumption that more proximal predictor variables work better than do less proximal ones, it follows that because direct attitude measures are assessing a variable (attitude) that is more proximal to behavioral intentions and behaviors than are indirect measures (which assess subjective probabilities and evaluations), direct measures should be a better predictor of behavioral intentions and behaviors than are indirect measures.

Numerous studies have provided tests of this theoretical reasoning. In the vast majority of cases, the reasoning has been supported; direct measures are generally

superior to indirect ones for predicting behavioral intentions and behaviors. An exception is when the attitude measure is not in accord with the principle of correspondence.

## **11. DISTINCTIONS THAT ARE IMPORTANT FOR ATTITUDE MEASUREMENT**

With the problem of low attitude–behavior correlations having been solved by the principle of correspondence, recent researchers have changed their focus to three other problems that are now addressed. First, there is an important issue of whether attitudes are really an amalgamation of a cognitive (thinking) and affective (feeling) component. If there are separate cognitive and affective components, an implication is that each of these components should be measured separately, thereby improving the prediction of behaviors in a variety of applied domains. Second, there has been a great deal of controversy over whether attitudes should be measured separately from subjective norms. It was not until recently that this controversy was finally resolved. Finally, there are some attitudes that are not easily measured according to the principle of correspondence, and it might be necessary to make an adjustment. These issues are addressed in the following subsections.

### **11.1. The Separate Measurement of Cognition and Affect**

Most researchers believe that there are both cognitive (thinking) and affective (feeling) components to attitudes. Three types of evidence support this belief. First, many researchers have used factor analysis to reduce a large number of attitude scales down to a smaller number of factors. In most cases, two factors result, with cognitive items loading on one factor and affective items loading on the other.

A second type of evidence comes from hierarchical regression analyses. Put simply, it is possible to consider the unique contribution of cognition to predicting general attitudes (or intentions) or to consider the unique contribution of affect. The results of these studies tend to indicate that, for any particular behavior, either cognition or affect will make a statistically significant unique contribution—a finding that should not occur regularly if attitude does not have these two components. An additional nicety of hierarchical regression analyses is

that the results are often consistent with researchers' intuitions. For example, cognition has been shown to be the more important contributor for what seems to be the cognitively controlled behavior of "studying over winter break," whereas affect has been shown to be the more important contributor for what seems to be the affectively controlled behavior of "smoking cigarettes."

Finally, Trafimow and Sheeran have made use of the associative hypothesis. They assumed that people have different types of beliefs about the consequences of behaviors and that some beliefs are more cognitive, whereas other beliefs are more affective. Given this, people are assumed to compare cognitive beliefs with other cognitive beliefs in the interest of forming the cognitive component of an attitude, and people are assumed to compare affective beliefs with other affective beliefs in the interest of forming the affective component of an attitude. But when people compare cognitive beliefs with other cognitive beliefs, or compare affective beliefs with other affective beliefs, they form associations; people form associations between cognitive beliefs and other cognitive beliefs, or between affective beliefs and other affective beliefs, but not between cognitive beliefs and affective beliefs. Consequently, when people are later asked to write down their beliefs about a behavior, writing a cognitive belief should cue the retrieval of another cognitive belief, whereas writing an affective belief should cue the retrieval of another affective one, thereby causing people's belief lists to be clustered by belief type. In contrast, if people do not distinguish between cognitive and affective beliefs (or between cognitive and affective components of attitudes), such clustering should not occur. In fact, such clustering is obtained, further supporting the distinction between cognitive and affective components of attitudes.

Empirical support for the distinction between cognitive and affective attitude components has led researchers to measure them separately to maximize the prediction of other variables such as behavioral intentions. Recent findings indicate that the prediction of behavioral intentions is significantly enhanced when the cognitive and affective attitude components are measured separately, and this finding has been replicated across a wide range of behavioral domains.

### **11.2. The Separate Measurement of Attitudes and Subjective Norms**

Since the 1960s, researchers have assumed that attitudes and subjective norms are different causes of behaviors

(or, more often, behavioral intentions). Whereas attitudes have been assumed to be caused by beliefs about the personal consequences of performing a behavior, subjective norms have been assumed to be caused by beliefs about what important others think one should do. Because attitudes and subjective norms have been assumed to be different causes of behavioral intentions, it made sense to measure them separately to maximize the prediction of behavioral intentions. However, during the 1970s and 1980s, many researchers questioned the distinction between attitudes and subjective norms and argued that these were really different names for the same underlying idea. Obviously, if attitudes and subjective norms are different names for the same underlying construct, there is no reason to have distinct measures of each of them. Thus, there is an important measurement issue at stake in this controversy: Should attitudes be measured separately from subjective norms or not?

The argument against the distinction between attitudes and subjective norms was based on three issues. First, attitudes and subjective norms were often found to be highly correlated with each other, consistent with the notion that they are merely different names for the same underlying construct. Second, path analyses have sometimes indicated “crossover” effects, whereby attitudes and subjective norms affect each other. Third, philosophical arguments have been made that beliefs about consequences (i.e., behavioral beliefs) that are presumed to cause attitudes are not different from beliefs about the opinions of important others (i.e., normative beliefs) that are presumed to cause subjective norms. An example should make this issue clear. Suppose that someone has the behavioral belief that “my father will disagree if I eat chocolate” or the normative belief that “my father thinks I should not eat chocolate.” Many researchers have argued that these two statements are just different ways of saying the same thing; therefore, if the cause of attitudes and subjective norms is the same, attitudes and subjective norms must also be the same.

Five kinds of findings have resolved most of the disagreement in favor of the distinction between attitudes and subjective norms. First, in general, behavioral beliefs have been found to be more highly correlated with attitudes than with subjective norms, and normative beliefs have been found to be more highly correlated with subjective norms than with attitudes—precisely what one would expect if behavioral beliefs cause attitudes and normative beliefs cause subjective norms. Second, although attitudes and subjective norms sometimes have been found to be highly correlated, the correlation often has been

found to be low or moderate. Third, attitudes and subjective norms have been manipulated experimentally, with different effects on behavioral intentions pertaining to different types of behaviors; manipulating attitudes has a larger effect on some behaviors, whereas manipulating subjective norms has a larger effect on others. Fourth, behavioral beliefs have been shown to be more strongly associated with other behavioral beliefs than with normative ones, and normative beliefs have been shown to be more strongly associated with other normative beliefs than with behavioral ones. Clearly, regardless of philosophical validity of the distinction between the two types of beliefs, people do make the distinction, contradicting the argument that attitudes and subjective norms have the same cause. Finally, there are individual differences between people in that some people are more under attitudinal control across a large number of behaviors, whereas others are more under normative control. If attitudes and subjective norms were the same thing, reliable individual differences in attitudinal or normative control should not be obtained. In sum, there is a great deal of evidence that attitudes and subjective norms are different from each other, and they should be measured separately.

## 12. CONSUMER APPLICATIONS

Attitude measurement is of more than just theoretical interest. Consumer and marketing researchers have applied the forgoing principles of attitude measurement in a variety of domains such as substance abuse among adolescents and tobacco use among college athletes as well as trying a new diet suppressant, getting mammograms, using condoms, drinking and driving, eliciting donations, purchasing environmentally friendly products, paying more for energy from renewable sources, purchasing food, voting, purchasing software, and even purchasing attitude research (by marketing directors). In these domains, as well as in many other domains, improved attitude measurement has resulted in an improvement in the prediction of behavioral intentions and behaviors from attitudes.

Improving the prediction of behavioral intentions and behaviors is not the only function of good attitude measures. It sometimes happens, although not very often, that an attitude does not do a good job of predicting a particular behavior. Before psychologists and consumer researchers had valid attitude measures, a low correlation between an attitude and a behavior was susceptible to at least two explanations. First, the

low correlation could show that the behavior is caused by something other than an attitude. Second, the low correlation could be due to an invalid attitude measure. An advantage of valid attitude measures is that they decrease the plausibility of the latter explanation and thereby increase the plausibility of the former one. Thus, when valid attitude measures nevertheless result in low correlations between attitudes and behaviors, researchers can be more confident in exploring other variables. A well-researched example of such an area is condom use. Although attitudes are capable of predicting condom use to some degree, other variables have also been shown to be good (perhaps better) predictors. Two of these are subjective norms and confidence that one knows whether others (e.g., one's sexual partner) think a condom should be used.

To illustrate the importance of valid attitude measures for consumer research regardless of whether the measures show that attitudes are a strong or weak predictor of behaviors, consider an example of a consumer researcher who wishes to increase sales of a particular product. Before investing money in an ad, the researcher needs to know what variables to address in the ad. If the behavior of buying the product is under attitudinal control for the population of interest, it makes sense for the ad to focus on variables that are likely to affect people's attitudes toward buying the product. Some of these variables might be beliefs about the product and affect. But what if the behavior is not under attitudinal control? In that case, there is little point in focusing an ad on variables designed to affect attitudes; for a behavior that is not under attitudinal control, there is no reason to believe that causing attitude change will increase sales. Consequently, the consumer researcher would be better served by creating an ad that focused on a different variable such as subjective norms. In general, having valid attitude measures increases the confidence that researchers can have in their data and provides a more solid basis for creating ads. Data showing that attitudes are good predictors of the behavior of concern provide a strong reason for creating an ad that is designed to affect attitudes; otherwise, the data provide a strong reason for creating an ad that focuses on other variables.

A further issue in attitude measurement, as applied to evaluating the effects of advertising, concerns the types of attitudes that should be measured. Consumer researchers have measured the effects of an ad on attitudes toward the ad, the product, or the brand. It has been rarer for researchers to measure the effects of an ad on attitudes toward the behavior the ad was

designed to influence (e.g., buying the advertised product, voting for the advertised candidate). According to the principle of correspondence, if the reason for creating an ad is to influence behavior, it is precisely this last type of attitude that should be measured. Although there have been some demonstrations that an ad affects attitudes toward the behavior rather than merely attitudes toward the ad, the product, or the brand, more research is needed to establish the size of these effects. This is particularly so because the effects are likely to depend on a large number of variables such as the product domain, the type of ad, the type of use to which the product is put, and the frequency with which people are exposed to the ad.

One final example illustrates the importance of whether one measures attitudes toward the ad, the brand, the product, or the behavior of buying the product. Suppose that a consumer researcher for a company wishes to increase sales of a particular product. In addition, suppose that this researcher is evaluating an ad that has been shown to cause a change in people's attitudes toward the ad, the brand, or the product. It should be clear from the principle of correspondence that these attitudes, despite their seeming importance, are likely to not be particularly relevant to whether people will buy the product.

There may be cases where the consumer researcher's goal is something other than increasing sales of a particular product. Perhaps the goal is to increase sales of all of the products made by the company. In that case, it might be worthwhile to run an ad that focuses on increasing people's attitudes toward the brand. Even though such an ad might be unlikely to cause much of an increase in the sales of a particular product, a small increase in the sales of several products may justify the cost of the ad.

### 13. CONCLUSION

There has been a great deal of progress in how attitudes are measured. Current attitude measures are more reliable, more valid, and more correspondent to behaviors than were attitude measures in the past. In addition, more is known about some of the relevant issues that underlie attitude measurement. For example, recent research indicates that cognition and affect are both components of attitudes and should be measured separately. Recent research also indicates that attitudes are different from subjective norms and that the two variables should be measured separately. The consequence of such progress is that the prediction of behavioral intentions and behaviors from attitudes is

much greater than it has ever been before, and this has been demonstrated in a wide variety of domains.

### See Also the Following Articles

Advertising and Culture ■ Advertising Psychology  
 ■ Intentional Behavior ■ Measurement and Counseling

### Further Reading

- Ajzen, I. (1988). *Attitudes, personality and behavior*. Chicago: Dorsey.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Courneya, K. S. (1994). Predicting repeated behavior from intention: The issue of scale correspondence. *Journal of Applied Social Psychology*, 24, 580–594.
- Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown, & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175–197). Malden, MA: Blackwell.
- Fazio, R. H. (1990a). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego: Academic Press.
- Fazio, R. H. (1990b). A practical guide to the use of response latency in social psychological research. In C. Hendrick, & M. S. Clark (Eds.), *Review of personality and social psychology*, Vol. 2. *Research methods in personality and social psychology* (pp. 74–97). Newbury Park, CA: Sage.
- Fishbein, M. (1980). Theory of reasoned action: Some applications and implications. In H. Howe, & M. Page (Eds.), *Nebraska Symposium on Motivation, 1979* (pp. 65–116). Lincoln: University of Nebraska Press.
- Trafimow, D. (1998). Attitudinal and normative processes in health behavior. *Psychology and Health*, 13, 307–317.
- Trafimow, D., & Sheeran, P. (1998). Some tests of the distinction between cognitive and affective beliefs. *Journal of Experimental Social Psychology*, 34, 378–397.



# Attitudes

George Y. Bizer

Eastern Illinois University, Charleston, Illinois, USA

1. Introduction
  2. The Tripartite Model of Attitudes
  3. Attitude Measurement
  4. Attitude Strength
  5. Attitude Change
  6. Conclusion
- Further Reading

## GLOSSARY

**attitude** A relatively enduring evaluation that a person holds about a target.

**attitude change** A process by which a person's evaluation toward a target is made more positive or negative.

**attitude strength** The extent to which an attitude is persistent over time, resists attempts to persuade, influences behavior, and influences cognition.

**central route to persuasion** The process of attitude change that relies on thoughtful processing of [the arguments within] a persuasive message.

**Likert scale** A tool for measuring attitudes in which respondents indicate to what extent they believe a series of items are characteristic of them regarding an attitude object.

**peripheral route to persuasion** The process of attitude change that does not rely on thoughtful processing of [the arguments within] a persuasive message.

**semantic differential** A tool for measuring attitudes in which respondents indicate how well a series of adjectives describes an attitude object.

**tripartite model of attitudes** Model suggesting that attitudes are derived from and/or influence affective, behavioral, and cognitive [components].

Attitudes are the relatively enduring evaluations that people hold toward all sorts of things in the world. Attitudes can be held toward people, objects, issues, and concepts. Researchers have devoted a great deal of effort to understanding attitudes because attitudes have been shown to be useful in predicting human behavior. Because of this interest, researchers have developed theories to explain where attitudes come from and tools with which to measure attitudes. Some research has focused on the distinctions between strong and weak attitudes, whereas other research has examined how attitudes can be changed. Because they are such an important part of everyday life, attitudes are one of psychology's most important constructs.

## 1. INTRODUCTION

A college student may like her psychology course but dislike her literature course. A consumer may prefer liquid laundry detergent over the powdered variety. And a city's populace may prefer the incumbent mayor over the challenger. Liking and disliking the objects in our world is a fundamental part of human nature. Psychologists use the term "attitude" to refer to these likes and dislikes that people hold—the relatively enduring evaluations that people hold toward all sorts of things. We can hold attitudes toward people (e.g., friends, celebrities), objects (e.g., automobiles, candy bars), political or social issues (e.g., capital punishment, immigration policy), and even abstract concepts (e.g., conservatism, democracy).

For years, researchers have spent a great deal of effort in attempting to better understand attitudes. One of the reasons why the concept has attracted so much attention is that attitudes serve an important role in our daily lives. The attitudes of college students influence which courses they decide to attend, the attitudes of consumers guide their decisions about which products to purchase, and the attitudes of a city's population affect voters' decisions about which candidate to elect. It is obvious that human behavior has a variety of causes, but understanding a person's attitudes is often a reliable way in which to predict the person's behavior. But attitudes also serve other roles. Sometimes, attitudes help us to gain rewards and avoid punishments. Other times, attitudes help us to define who we are as people; people may feel so strongly about an attitude that they define themselves in terms of that attitude (e.g., people may consider themselves to be "staunch Democrats" or "hardcore pro-lifers"). In addition, some research has even shown that holding negative attitudes toward other people may make us feel better about ourselves. Thus, because the attitudes we hold play important roles in many arenas, they are indeed a fundamental part of our everyday lives.

## 2. THE TRIPARTITE MODEL OF ATTITUDES

Where do attitudes come from? Research on the tripartite model of attitudes suggests that attitudes can be based on

three different sources (for a visual representation of the tripartite model, see Fig. 1). One part of the tripartite model is the affective component. This affective basis of attitudes refers to the feelings, moods, and emotions that are associated with the attitude object. Consider a person who likes to exercise—a person who could be said to have a positive attitude toward exercise. This person's feelings of being "invigorated" or "distressed" whenever he or she exercises could serve as the affective basis of the attitude. Next, the behavioral basis refers to how the person has acted or would act in the future in regard to the target. If the person visits the gym to work out four times per week, this activity would serve as the behavioral basis. Finally, the cognitive basis refers to the overt thoughts or beliefs that the person has in regard to the target. If the person thinks that exercise is "healthy" and "worthwhile," such thoughts would serve as the cognitive component of the attitude.

In this example, the affective, behavioral, and cognitive components of the attitude all are positive, and this would suggest that the person will likely hold an overall positive attitude toward exercise, that is, the attitude object. However, the three components will not necessarily be consistent. A person may think that chocolate cake tastes wonderful (affective) and may eat chocolate cake often (behavioral) but may also think that chocolate cake is loaded with calories and fat (cognitive). In such cases, the overall attitude toward the object may be based more heavily on some components than on others.

It is important to note, however, that the components of attitudes can also be influenced by the attitude itself.

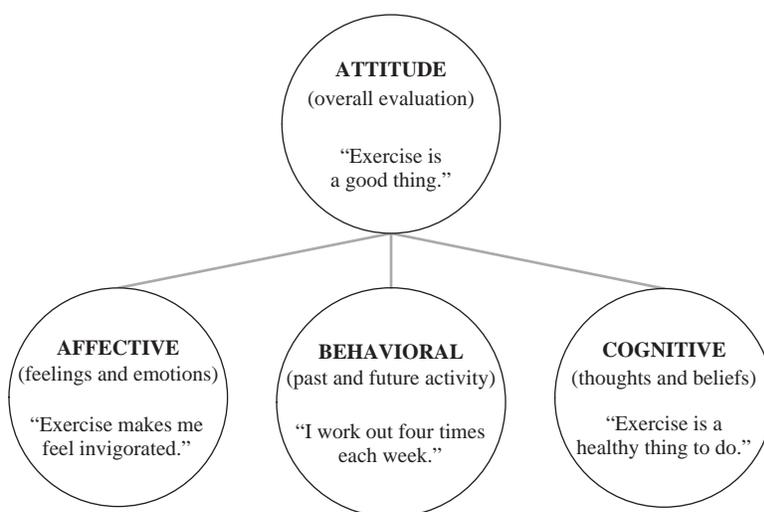


FIGURE 1 Visual representation of the tripartite model.

For example, in 1993, Eagly and Chaiken discussed how the attitude can also influence the three components. In other words, a person's affect, behaviors, and cognitions not only may guide and create the attitude but also may be driven or influenced by the attitude. As such, the three components of the tripartite model can be conceptualized in terms of both antecedents and consequences of their associated attitudes.

### 3. ATTITUDE MEASUREMENT

Because attitudes are so important, researchers have developed a variety of tools with which to measure them. Two of the most commonly used techniques are the Likert scale and the semantic differential. As shown in Fig. 2, participants reporting attitudes on a Likert scale are asked to what extent a variety of statements are characteristic of them. In a similar vein, as shown in Fig. 3, participants reporting attitudes on a semantic differential measure are asked to indicate how well a series of adjectives describe the attitude object. In either case, the researcher will typically provide a variety of items for the participant to complete. The scores from the items are combined to create an overall measure of the person's attitude toward the target.

There are several guidelines that researchers can follow in attempting to tap a person's attitude with as little error as possible. One guideline is to reverse code some of the items by framing them negatively as done in the third and fifth items in Figs. 2 and 3. Doing so

Oranges:						
Healthy	<u>  x  </u>	—	—	—	—	Unhealthy
Good-tasting	<u>  x  </u>	—	—	—	—	Bad-tasting
Expensive	—	—	—	<u>  x  </u>	—	Affordable
Pleasant	<u>  x  </u>	—	—	—	—	Unpleasant
Messy	—	—	<u>  x  </u>	—	—	Clean

FIGURE 3 Example of a semantic differential scale.

serves two purposes. First, it reduces the likelihood that a person will simply check the leftmost or rightmost box down the entire instrument in attempts to save time or cognitive resources. A person is less likely to do so if he or she realizes that various items are indeed reverse coded. Similarly, if a person does simply check the leftmost or rightmost box down the entire instrument, the researcher may recognize this and treat the data accordingly. Also, some research suggests that providing respondents with five to seven options from which to select provides the most valid data. Scales with fewer than five options lack the precision necessary to measure attitudes accurately, whereas scales with more than seven options tend to suffer from additional noise without benefiting from any additional precision.

Likert scales and semantic differential scales are ubiquitous. Students typically use Likert scales when evaluating their professors at the end of a semester, and departments and universities use these data in part to make personnel decisions. Companies also use such techniques to measure customers' attitudes toward the products and/or services the companies offer. By learning which aspects of the companies customers like and dislike, managers can direct resources to improve deficient areas of their companies. In short, proper attitude measurement can gauge the likes and dislikes of a person or a group of people. Behavior of a person or a group of people can then be predicted based on the attitude measure.

### 4. ATTITUDE STRENGTH

Although attitudes typically predict behavior, this is not always the case. Some attitudes have tremendous

Please indicate the extent to which the following statements are or are not characteristic of you:	
1 = very uncharacteristic of me	
2 = somewhat uncharacteristic of me	
3 = neither characteristic nor uncharacteristic of me	
4 = somewhat characteristic of me	
5 = very characteristic of me	
1. Oranges are a healthy fruit.	<u>    5    </u>
2. Oranges taste good.	<u>    5    </u>
3. Oranges are expensive.	<u>    2    </u>
4. Oranges are pleasant.	<u>    5    </u>
5. Oranges are messy.	<u>    3    </u>

FIGURE 2 Example of a Likert scale.

influence over our lives, whereas others are largely inconsequential. For example, consider two people's attitudes toward environmental protection. The two people both may think that protecting the environment is a good thing—they both may hold positive attitudes toward environmentalism—but the strength of those attitudes may be quite different. One person's attitude may lead her to vote only for pro-environment candidates, drive a fuel-efficient car, and donate money to pro-environment organizations. On the other hand, the other person's attitude may have very little impact. He will occasionally recycle and has on occasion voted for pro-environment candidates. Why would two people who hold equivalent attitudes act so differently?

The difference lies in attitude strength. In 1995, Petty and Krosnick defined attitude strength as the extent to which an attitude persists over time, resists change when exposed to persuasive information, influences cognitive processes, and influences behavior (as in the preceding example). Although two people may hold what appear to be equivalent attitudes—perhaps as observed by identical scores from Likert scales—the attitudes might not be functionally equivalent. Researchers have identified a wide variety of ways in which attitudes can become stronger, including thinking about attitudes, learning about the attitude object, or expressing attitudes repeatedly.

Attitude strength can be measured by observing the four defining features directly: how long attitudes persist over time, how well attitudes resist persuasive attempts, the extent to which attitudes predict cognitive processes, and the extent to which attitudes predict behavioral processes. But there is a wide variety of other features that are associated with these defining features. Some researchers have investigated the certainty with which attitudes are held: when a person feels certain of an attitude, that attitude is likely to show the four defining features of attitude strength. In a similar vein, researchers have shown that accessible attitudes, attitudes based on much knowledge, and attitudes on issues that people find to be personally important also tend to manifest the four defining features of attitude strength.

Therefore, many researchers measure not only attitudes but also the strength of those attitudes. For example, consider an incumbent governor's campaign manager determining in which counties to invest scarce advertising funds. Although two counties may show roughly equivalent positive attitudes toward the incumbent, such a report tells only part of the story. Based on this information alone, the manager might decide to split the advertising dollars across both counties equally. However, it is

possible that the citizens of County A may be, on average, highly certain of their positive attitudes, whereas the citizens of County B may be, on average, not certain at all. In such a case, one would expect that the positive attitudes would predict vote choice better among the residents of County A than among those of County B. Therefore, the manager would benefit more from investing advertising in the county with the weaker attitudes (as measured by certainty) than in the county with the stronger attitudes. Thus, measuring attitudes alone is only part of the story. One must also measure the strength of those attitudes to predict the behavioral and cognitive consequences of attitudes most effectively.

## 5. ATTITUDE CHANGE

Attitudes are thought to be relatively enduring evaluations. However, there is a wealth of research on how attitudes can and do change in the face of a persuasive communication. One of the most widely studied theories of persuasion is Petty and Cacioppo's elaboration likelihood model. In their 1986 article, these researchers posited that there are two routes through which persuasion can take place. When people have the motivation and the ability to carefully process and think about the arguments within a counterattitudinal message, they will use the central route to persuasion. When people process messages centrally, they evaluate the arguments within the message—the logical factual information that is directly relevant to the attitude object. People will then change their attitudes to the extent that their cognitive responses—the thoughts they have about the arguments while reading the message—are positive. However, when people lack either the motivation or the ability to carefully process and think about the arguments within a counterattitudinal message, they will use the peripheral route to persuasion. When people process a message peripherally, instead of elaborating on the arguments within the message, they respond to cues that exist within the message—images or heuristics not directly relevant to the message that can persuade without requiring much thought.

For example, consider an advertisement for a new car that appears in a general-interest magazine. At the top of the advertisement is the name of the car, and at the bottom of the page are three paragraphs detailing the car's resale value, horsepower, gas mileage, and safety record. In the background is an image of the car being driven across a sun-drenched beach with beautiful palm trees swaying in the background. A reader who is in the market for a new automobile may have the motivation to

use the central route to persuasion when processing the advertisement. This person will carefully read the arguments at the bottom of the page and, assuming that the reader's cognitive responses are positive, will end up with a more positive attitude toward the car. However, consider another reader who does not plan to purchase a car for years and therefore has no motivation to use the central route. This person may instead simply observe the beautiful palm trees and the beach and therefore simply create an association between the positive imagery and the car. Thus, the second person may also end up with a more positive attitude toward the car but will do so through the peripheral route instead of the central route.

If both the central and peripheral routes can lead to the same effect (e.g., attitude change), why does it matter which route to persuasion people take? Perhaps the most important reason is that attitudes resulting from central route persuasion tend to be stronger than attitudes resulting from peripheral route persuasion. That is, attitudes formed from central route processing tend to be more persistent over time, more resistant to future persuasion, and more impactful on behavior and cognitive processes than do attitudes formed from peripheral route processing. So, from this standpoint, it would seem better to engage an audience in central route processing due to the enhanced strength that would result. However, peripheral route processing may be the best option when (a) the audience does not have the ability or motivation to use the central route or (b) there are very few compelling arguments that can be presented in a message in the first place. In short, persuasion practitioners must balance the costs and benefits when determining whether to use primary central route or peripheral route persuasive appeals.

## 6. CONCLUSION

Attitudes, as the representations of what we like and dislike, play an important role in our daily lives. From

important decisions such as choosing one presidential candidate over another to more mundane decisions such as which flavor of ice cream to purchase, attitudes play a fundamental role in all sorts of behavioral and cognitive processes. Because they are such an important part of our lives, they have all sorts of important applications. Political scientists study attitudes in attempting to predict voters' choices, consumer psychologists study attitudes in attempting to understand consumers' purchasing decisions, and health promotion researchers study attitudes in attempting to understand the healthy and unhealthy behaviors in which people engage. In short, because they play a fundamental role in explaining behavioral and mental processes, attitudes are one of psychology's most important constructs.

### See Also the Following Articles

Attitude Measurement ■ Intentional Behavior ■ Measurement and Counseling

### Further Reading

- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 269–322). New York: McGraw–Hill.
- Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly*, 24, 163–204.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205.
- Petty, R. E., & Krosnick, J. A. (Eds.). (1995). *Attitude strength: Antecedents and consequences*. Hillsdale, NJ: Lawrence Erlbaum.
- Zimbardo, P. G., & Lippé, M. R. (1991). *The psychology of attitude change and social influence*. New York: McGraw–Hill.



# Authoritarianism

Klaus Boehnke

International University Bremen, Bremen, Germany

Andreas Hadjar

Chemnitz University of Technology, Chemnitz, Germany

1. Authoritarianism: The Concept
  2. Correlates of Authoritarianism
  3. Treatment Approaches
- Further Reading

## GLOSSARY

**attitude** Persistent mental state of readiness to react to a certain object or class of objects in a certain way; learned disposition to behave in a consistent way toward a given class of objects.

**disposition** Relatively long-lasting emotionally rooted attitude.

**ethnocentrism** Tendency to divide the social world into groups with which one identifies and to which one submits (in-groups) and groups of outsiders (out-groups) to which one is hostile; characterized by glorification of the in-group and defamation of the out-group.

**personality structure** Unit that underlies individual ways of behaving and gives consistency to otherwise contradictory-seeming mannerisms and behavioral or attitudinal specialties.

**personality syndrome** Acquired pattern of personality characteristics bearing resemblance to the personality structures of others who shared similar experiences and problems and made similar adaptations.

**xenophobia** Abnormal fear of strangers.

Authoritarianism is a psychological concept that assumes the existence of a more or less unconscious

general orientation of individuals toward being antidemocratic, prejudiced, and fascist. The so-called “authoritarian personality” is characterized by two ambivalent tendencies: subordination to authorities and the desire to become an authority himself or herself. Authoritarianism as a personality syndrome is predominantly measured by using the so-called F-scale and later revisions of that instrument.

## 1. AUTHORITARIANISM: THE CONCEPT

There is no concept in the social sciences that is more closely connected to the events in Germany and throughout the world in the 1930s and 1940s than the authoritarianism concept. For decades, this approach served as the main explanation of fascism and anti-democratic thoughts or action in psychology.

### 1.1. Historical Background

Precursors of the authoritarianism concept can be traced back to the early and mid-1930s. At that time Marxist psychoanalyst Wilhelm Reich studied voting behavior of Germans between 1929 and 1933 to find reasons for the increase in the number of voters for the fascist NSDAP party from less than 1 million to approximately 17 million voters during that period. Within his framework of thinking, he concluded that ideologies of subordination to authority were

internalized by subordinate individuals and eventually became a stable personality structure. Some years later, researchers of the so-called Frankfurt School, such as Erich Fromm, Theodor Adorno, Max Horkheimer, and Herbert Marcuse, carried out a study on “Authority and Family” and systematized the idea of an authoritarian personality structure that is strongly determined by the societal context. In the 1940s, members of the Frankfurt School, who had to leave Germany due to their Jewish and Marxist background, engaged in a much larger research project—partly supported by the American Jewish Committee—in their American exile. Influenced by their experience of racism and inhumanity in fascist Germany, the group around Adorno set out to uncover psychological forces that promote fascist and antidemocratic attitudes and to find possible ways to fight fascist tendencies in society. They used quantitative and qualitative research methods (e.g., questionnaire-based surveys and qualitative semistructured interviews). Individuals surveyed in the U.S.-based studies were mainly students and members of the American middle class.

## 1.2. Concept of the Authoritarian Personality

The main thesis of Adorno was that the political, economic, and social attitudes of an individual form a coherent structure of thought, which is an expression of an underlying, hidden personality structure. In accordance with Freudian psychodynamic theory, this personality structure is assumed to have developed during early childhood. Authoritarian relations within the family, an authoritarian parental style, and a general lack of love and warmth determine it. Antidemocratic individuals often experience feelings of hatred against authorities—namely their father as the family’s bread winner—during their childhood. Because they never could express this hatred in any way against their father or their parents, they direct such feelings of hatred toward convenient scapegoats—inferiors and social minorities (e.g., immigrants/foreigners, homosexuals, and the handicapped)—later in life. Main characteristics of authoritarian personalities include a preoccupation with superiority and one-upmanship, an intolerance of behavior and objects that are different, a tendency to classify all things into black and white/right and wrong, a strict adherence to received views and prevailing social trends, and superficial respect for authority figures. Adorno and

collaborators developed the so-called F-scale (fascism scale) to measure such an antidemocratic personality structure (Fig. 1).

The scale is constructed to measure a syndrome that comprises nine dimensions or subscales not always present simultaneously:

1. Conventionalism: rigid adherence to the conventional values of the middle class
2. Authoritarian submission: an uncritical and submissive attitude toward idealized in-group authorities
3. Authoritarian aggression: a tendency to look out for and to condemn, reject, and punish people who violate conventional values
4. Anti-intracception: rejection of all inwardness, of the subjective, the imaginative, the tender-minded, and of self-criticism
5. Superstition and stereotypicality: the belief in mystical determinants of the individual’s fate, a disposition to think in rigid categories
6. Power and toughness: a preoccupation with the dimension of dominance–submission, strong–weak, leader–follower; identification with power figures; overemphasis on the conventionalized attributes of the ego; exaggerated assertion of strength and toughness
7. Destructiveness and cynicism: a generalized hostility and vilification of the human
8. Projectivity: to project unconscious emotional impulses onto the outside world and to believe that wild, evil, and dangerous things go on in the world
9. Sexuality: an exaggerated concern with sexual matters

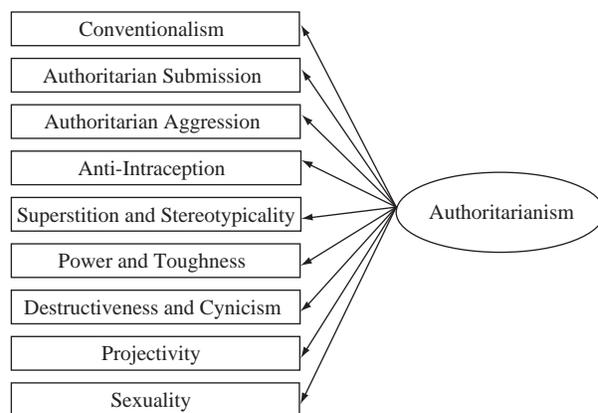


FIGURE 1 The authoritarianism construct.

### 1.3. Criticism

The accomplishment of Adorno and collaborators was to create an insight into the phenomenon of an authoritarian personality. However, there are methodological and empirical limitations. One of the foremost points of criticism is that the theory attempts to explain societal processes (Germany's turn to fascism) on the basis of knowledge about individual personality development. Another major criticism is that the authoritarianism concept and the F-scale questions appear highly ideological. This criticism is repudiated by a number of psychologists, including Gerda Lederer, who points out that F-scale items are related to moral and individual values, human relations, the self, family, and sexuality and are not at all ideological. At least the popular reception of the concept, however, is highly politicized. Particularly in the 1950s, it was used to characterize not only fascist but also communist movements. Yet other points of criticism are concerned with methodological problems of Adorno's empirical work, indicating, for instance, a sampling problem. Adorno generalized from a nonprobability sample that consisted of white middle-class Americans. Also, several studies showed problems with the construct validity of the F-scale, i.e., they doubt that the F-scale measures exactly what it intends to measure. Furthermore, semi-structured interviews conducted with individuals in order to gain information on their parents' educational style were criticized as being more of a memory exercise than a pool of information about parental authoritarianism. Some studies replicating Adorno's work found no evidence to support the central findings, particularly on the impact of authoritarian socialization. One study even reports evidence for a positive relationship between authoritarianism and xenophilia. However, taking into account the multitude of existing studies, one can see the existence of the phenomenon authoritarian personality as confirmed.

### 1.4. Contemporary Developments of the Authoritarianism Concept

The initiator of the new research stream on authoritarianism is Robert Altemeyer. His approach leaves the theoretical basis of the Freudian psychodynamic theory behind and focuses on learning processes highlighting agents of socialization, such as family and peer group, and the context of society. In his work the so-called right-wing authoritarianism is influenced by situation or environment rather than just personality

development. Altemeyer reduces the dimensions of authoritarianism to three: authoritarian submission, authoritarian aggression, and conventionalism. In the 1980s, many social scientists attempted to develop new instruments to measure the authoritarian personality. Others tried to explore by qualitative research methods the conditions that lead to such a personality structure.

### 1.5. Similar Contemporary Concepts

There are two contemporary competitors to the authoritarianism concept. During the past decade, Sidanius and Pratto studied what they call the social dominance orientation (SDO). SDO refers to fundamental values to accept social hierarchies. These values form a personality structure in the Freudian sense and determine so-called legitimizing myths, which are moral and intellectual justifications of individual or institutional discrimination. Whereas SDO still recognizes psychodynamic assumptions, the concept of hierarchic self-interest (HSI) is based on a paradigm of lifelong socialization in a market-oriented economy. HSI is an expression of the individual effort to perform "better than others" in a hierarchically structured society, an internalized elbow mentality, and it appears as a syndrome of values that comprises three core dimensions (success orientation, competitiveness, and individualism), as Hagan and collaborators showed in their youth studies. Since HSI is a syndrome and second-order construct (see Fig. 2 for an example from empirical research), it is not ultimately determined but can

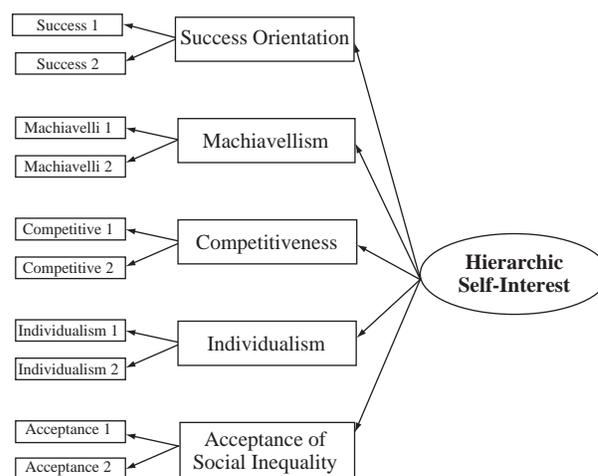


FIGURE 2 Hierarchic self-interest (structural model).

be modified by integrating other first-order factors (e.g., machiavellism and acceptance of social inequality). HSI turned out to be a stable predictor of xenophobia. This finding is in a way paradoxical because market-oriented HSI values apparently lead—under the condition of increasingly limited resources—to a tendency of ethnocentric market restriction. Roots of HSI are seen in modes of workforce participation of men and women and parental styles so that high HSI values may be interpreted as expressions of familial or socioeconomic disintegration.

## 2. CORRELATES OF AUTHORITARIANISM

### 2.1. Antecedents

#### 2.1.1. Class and Socioeconomic Status

Empirical evidence suggests that authoritarianism is found among the lower classes to a higher degree than among the middle classes. Lipset assumes in his working class authoritarianism thesis that lower classes tend to have an authoritarian view on politics; therefore, they are more likely to support extremist movements that promise fast and simple solutions to social problems and that blame inferior scapegoats for the problems. This authoritarian view results from low education, the degree of isolation of the class, economic and psychological uncertainty, and the particular family life in lower classes. Depending on the exact kind of authoritarianism scale that was used and how class was measured, results of empirical studies support this thesis, particularly the influence of education on authoritarianism. A higher level of education seems to be the decisive characteristic of nonauthoritarian people.

#### 2.1.2. Family Socialization

Adorno *et al.* emphasize the role of the family for the development of the authoritarian personality. Other scientists studied the mechanisms that lie behind the relationship of socialization and authoritarianism in more detail. They suggested that the transmission of authoritarian attitudes from the parents onto the children depends heavily on the strength of social and emotional ties between parents and children. Central to an authoritarian socialization are parental styles that include violence and do not recognize the

needs of the children. Authoritarian families usually are ruled by the father and do not allow children to participate in family decisions.

#### 2.1.3. Cross-Cultural Differences

In dealing with the concept of authoritarianism, it must be discussed whether this concept applies only to the Western world or whether it is universal. Based on data from more than 100 countries, Meehan shows that there is a strong relationship between culture, attitudes, and politics. State authoritarianism is highly related to authoritarian attitudes among the citizens. These attitudes seem to result from a culture that is based on hierarchies and a traditional family structure. Comparisons of authoritarianism among adolescents in East and West Germany after the fall of the Berlin Wall show very clearly that different socialization cultures lead to different degrees of authoritarianism: East German adolescents agreed with authoritarian statements more than West Germans did. However, scales to measure authoritarianism cannot be used throughout the world in the same way and the same composition. Words and contexts are interpreted differently in different cultures. Studies in Japan, for example, showed that authoritarian structures and authoritarian attitudes do exist but cannot be measured in terms of American authoritarianism instruments.

### 2.2. Consequences

The best studied and most powerful relation is the one obtained for authoritarianism and ethnocentrism. Ethnocentrism is an attitude characterized by the glorification of one's own group (in-group) and the defamation and discrimination of other groups (out-group). Xenophobia, racism, and nationalism are other orientations similar to ethnocentrism. An authoritarian personality tends to be more ethnocentric. In certain circumstances (situational factors), the just mentioned attitudes can lead to discrimination, violence, and hate crimes.

Authoritarianism is also connected to other attitudes, namely sexism, anti-Semitism, cognitive rigidity (dogmatism), and political and economic conservatism. Authoritarian personalities tend to devalue women, homosexuals, and Jews. They strongly support the market economic system and back conservative political movements. Authoritarianism also may explain oppression, chauvinism, and negative put-downs on

the individual level and the level of society. Many of the empirical findings on consequences of authoritarianism may, however, rightly be accused of bearing an element of tautology because almost all consequences addressed could also be seen as elements of the authoritarian syndrome per se.

### 3. TREATMENT APPROACHES

From knowledge about explanatory factors of authoritarianism, particularly socialization, a list of treatment measures can be derived. The founders of the Frankfurt School suggested enlightenment and an education fostering responsibility. That means making the mechanisms of society transparent to a broad audience. The psychologist Kurt Lewin, who conducted several experiments on group conflicts in small groups, pointed out that democratically acting group leaders are necessary to prevent authoritarianism. Group leaders as well as all other group members must be enabled to play their role in the democratic process. Authoritarian leadership styles must be replaced by consultative and participative leadership styles. To support this process of democratization, the democratic leader has to leave space for self-determination with clearly defined and reasoned borders to the group members rather than predetermining a more or less restricted space with rigid borders. Democratic skills cannot be passed on in a simple way; they must be learned and interactively developed in every generation, in all institutions of society, including the family, the school, and the workplace. Whereas a Freudian approach to authoritarianism would suggest that there is no treatment for authoritarian personalities in adulthood, from the perspective of lifelong socialization a cure seems possible. To push back authoritarian tendencies in daily life settings, people's integration into groups (working groups, peer groups, and families) must be supported. People must be given participation opportunities in group decisions; rules, actions, and decisions shall be explained to them in a detailed and reasoned way; they shall be encouraged to look at themselves and others from different perspectives; and every kind of rigid discipline and physical or

mental violence must be absent in dealing with people who tend to be authoritarian.

### See Also the Following Articles

Family and Culture ■ Industrial/Organizational Psychology across Cultures ■ Interpersonal Behavior and Culture ■ Power, Authority, and Leadership ■ Stereotypes

### Further Reading

- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.
- Brown, R. (1995). *Prejudice. Its social psychology*. Oxford: Blackwell.
- Christie, R., & Jahoda, M. (1954). *Studies in the scope and method of the authoritarian personality*. Glencoe, IL: Free Press.
- Hadjar, A. (2004). *Ellenbogenmentalität und Fremdenfeindlichkeit bei Jugendlichen. Die Rolle des Hierarchischen Selbstinteresses [Elbow mentality and xenophobia among adolescents. The role of hierarchic self-interest]*. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Hagan, J., Hefler, G., Classen, G., Boehnke, K., & Merkens, H. (1998). Subterranean sources of subcultural delinquency beyond the American Dream. *Criminology*, 36.
- Lewin, K. (1948). *Resolving social conflicts*. New York: Harper & Brothers.
- Martin, J. L. (2000). The authoritarian personality, 50 years later: What lessons are there for political psychology. *Political Psychology*, 22.
- Meleon, J. D., Hagendoorn, L., Raaijmakers, Q., & Visser, L. (1988). Authoritarianism and the revival of political racism: Reassessments in The Netherlands of the reliability and validity of the concept of authoritarianism by Adorno et al. *Political Psychology*, 9.
- Rippl, S., Seipel, C., & Kindervater, A. (2000). *Autoritarismus. Kontroversen und Ansätze der aktuellen Autoritarismusforschung [Authoritarianism: Controversies and approaches of contemporary research on authoritarianism]*. Opladen: Leske & Budrich.
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.



# Autism and Asperger's Syndrome

Fred R. Volkmar and Ami Klin

Yale University, New Haven, Connecticut, USA

1. Definition and Diagnosis
  2. Epidemiology
  3. Biological Mechanisms
  4. Psychological Processes
  5. Treatment
  6. Course and Prognosis
- Further Reading

## GLOSSARY

**circumscribed interests** Areas of profound interest pursued with great interest and intensity; the usual focus is on the acquisition of facts. These interests come to dominate the child, and the family's, life and interfere with acquisition of skills in other areas.

**echolalia** Repetition of words/phrases/sentences; may be immediate (i.e., occurring in the immediate context) or delayed (repetition of words used in the past).

**nonverbal learning disability** A profile of strengths and weaknesses on neuropsychological testing with detrimental effects on social skills.

**savant skills** Areas of unusual ability typically seen in people who otherwise exhibit mental retardation/intellectual disability.

**stereotypies** Repetitive motor mannerisms typically involving the whole body (body rocking) or extremities (hand flapping and toe walking).

Autism and Asperger's syndrome are disorders characterized by marked disturbances in the basic

aspects of social development and are associated with unusual behaviors. In autism, marked disturbances in language and communication are present and the disturbances in behavior can take the form of unusual purposeless mannerisms (stereotypies), behavioral rigidity and compulsive adherence to rituals and routines, unusual preoccupations, or preoccupation with nonfunctional aspects of materials. In Asperger's syndrome, language skills are, in general, preserved, although social use of language may be impaired; typically unusual, all encompassing, circumscribed interests are present. The relationship of these two conditions is the topic of some debate. There is general agreement on the important role of genetic and other neurobiological factors in causing these conditions. Treatment, particularly if delivered early, can significantly improve outcome.

## 1. DEFINITION AND DIAGNOSIS

Autism was first described by Leo Kanner in 1943 when he described a group of children who he believed were born without the usual predisposition to be social. Kanner's description has, in general, proven remarkably accurate. He emphasized two essential features for autism: (i) The autism involved what appeared to be self-absorption, and (ii) the insistence on sameness involved the unusual preoccupation with objects and difficulties in dealing with change. Kanner described many of the other clinical features often observed in autism

(e.g., unusual “splinter” skills on psychological testing, profound problems in communication, and difficulties with echolalia and language use for those children who did speak). In 1944, Hans Asperger, unfamiliar with Kanner’s paper, published his description of a small group of boys with what he termed autistic personality disorder or, as it was more commonly translated, autistic psychopathy. Like Kanner, he emphasized the marked social difficulties in these cases; unlike Kanner, he mentioned that in his cases words seemed to serve as the “lifeline” for the child. Asperger also noted the presence of strong circumscribed interests that dominated the child’s life (and often the family life as well) to the point that these interfered with acquiring other skills. His paper also mentioned some other point of divergence from Kanner’s description; for example, Asperger believed that the disorder tended to run in families (particularly in fathers) and was associated with motor clumsiness.

In the years following Kanner’s description (which was much more widely known in the English-speaking world) there was much controversy about the possible continuity of autism with schizophrenia and other forms of psychosis; the etiology of the condition was also controversial, with some clinicians implicating deviant experience. However, as data accumulated it became clear that autism was not an early form of schizophrenia and various lines of evidence strongly implicated neurobiological factors in pathogenesis. Asperger’s work was largely ignored in the English-speaking world until a paper by Lorna Wing (1981) introduced his concept and stimulated much research.

Autism was first officially recognized in 1980 as diagnosis apart from childhood schizophrenia. Asperger’s disorder was only recognized in 1994, and its definition remains the topic of debate.

## 2. EPIDEMIOLOGY

More than 30 studies have been conducted on the epidemiology of autism, with most focusing on more stringently diagnosed autism. A prevalence rate of approximately 1 case per 1000 children appears reasonable, with lower rates observed for Asperger’s disorder (approximately 2.5/10,000 children). Recent studies have generally reported higher rates than older ones, but changes in diagnosis, an increased awareness of the disorder, and other factors complicate our understanding of this increase.

Although Kanner’s initial report suggested high rates of professional and educational achievement in parents, it appears that this was due to a bias in cases referred to him; children with autism and Asperger’s come from parents of all levels of educational and occupational achievement and from all races and cultures. Studies consistently report higher rates in boys than in girls for both conditions (3.5 or 4 boys for every girl); girls with autism are more likely to have significant mental retardation, raising the possibility that males have a lower threshold for expressing the disorder.

Autism is associated with a range of medical conditions. Most strikingly, perhaps 20% of children with autism will develop seizure disorder (epilepsy). Although autism has been associated with a range of other medical conditions, the two strongest associations have been for Fragile X syndrome and tuberous sclerosis—two strongly genetic disorders.

## 3. BIOLOGICAL MECHANISMS

Various lines of evidence suggest the importance of neurobiological factors in causing autism and Asperger’s disorder. For example, children with autism exhibited an increased rate of persistent primitive reflexes, delayed development of hand dominance, EEG abnormalities, and seizure disorder. During the past few decades, there has also been an increased appreciation of the role of genetic factors in both conditions. This evidence includes the observation of much higher rates of concordance for autism in identical as opposed to (same-sex) fraternal twins, an approximately 50-fold increase in the disorder among siblings, and an increased risk to siblings of a range of other developmental problems. Although the data for Asperger’s disorder are more limited, there is evidence for significant rates of this condition in family members as well.

Studies of neurochemistry have not found specific biological markers, although it is clear that as a group, individuals with autism exhibited elevated levels of the neurotransmitter serotonin. Neuroimaging and neuropathological findings have included a general increase in brain size and abnormalities in the cerebellum and in structures such as the amygdala.

## 4. PSYCHOLOGICAL PROCESSES

In his initial report of infantile autism, Kanner mentioned that children with autism likely had good

cognitive potential given their ability to perform well on some parts of intelligence tests. Whereas subsequent research has confirmed the observation that there are some areas of strength in autism, it also has made it clear that approximately 70% of children with autism do exhibit mental retardation. Areas of strength on psychological testing typically include nonverbal skills, whereas verbal skills are generally areas of weakness. In Asperger's disorder, a reverse profile (often of the type termed nonverbal learning disability) is often identified.

People with autism do, at times, exhibit unusual skills ("islets of special abilities") that stand in marked contrast to overall intellectual level. For example, there might be unusual facility in decoding letters and numbers (hyperlexia) or feats of memory, calendar calculation, drawing, or musical abilities. Such individuals are sometimes termed autistic savants. In Asperger's disorder, special interests often take the form of considerable knowledge about some circumscribed area of interest (e.g., clocks, time, dinosaurs, the operas of Wagner, and snakes).

Various psychological theories have been proposed to account for the unusual constellation of abilities in autism. One theory focuses on what is termed a lack of central coherence with a resulting difficulty in integrating information. Another theory posits deficits in executive function (i.e., in organizational and forward planning skills). Yet another theory suggests that individuals with autism lack a theory of mind (i.e., have difficulties in putting themselves in another person's place). This theory has been highly productive for research but fails to account for the significant social disability of high functioning individuals who usually can perform theory of mind tasks with ease; a lack of developmental orientation is also problematic (i.e., children with autism exhibit social deficits in areas that typically would precede emergence of theory of mind).

Recent work has highlighted the importance of studying specific social processes in autism, for example, in the ways individuals with autism view social scenes or look at faces.

## 5. TREATMENT

Educational and behavioral interventions are the mainstay of treatment for autism and Asperger's disorder. Goals of treatment include minimizing disruptive behaviors, promoting learning and "learning to learn" skills, and increasing social, communication, and self-help skills. Generally, a fairly intensive and highly structured

program is needed, particularly for younger children with autism. Changes in syndrome expression with increased age and development require periodic updates of treatment goals. Given the tendency toward rigidity and perseveration, generalization of skills is an important aspect of intervention. Augmentative communication systems may be helpful. For children with Asperger's disorder, better verbal skills sometimes mask the extent of problems in other areas. Consideration of the child's pattern of strengths and weaknesses should be included in educational planning. Most educational programs for children with autism utilize behavioral management techniques to some extent. Psychodynamic, play therapies have a limited role in treatment; for verbal individuals, structured and focused psychotherapy/counseling may be helpful.

Although no drug treatment has proven curative, important target symptoms, such as self-injury, aggression, and overactivity, can be significantly improved. For individuals with autism, the major tranquilizers (particularly the atypical neuroleptics) have been used. For higher functioning individuals with autism and Asperger's, antidepressants and selective serotonin reuptake inhibitors may be of some benefit.

## 6. COURSE AND PROGNOSIS

Earlier detection and intervention has significantly improved the outcome in autism. Prior to provision of educational services, relatively few individuals with autism were able to attain personal self-sufficiency and independence as adults. However, longitudinal data suggest that with provision of early diagnosis and intervention perhaps 15% of individuals with autism are able to be relatively independent and self-sufficient as adults, with another 20% of cases able to function in the community with some support. Positive prognostic factors include overall intelligence and communication skills (particularly the presence of communicative speech by age 5). Although data are relatively more limited, individuals with Asperger's disorder appear to have more positive outcomes (as a group, they also have higher IQs and good verbal skills). A major medical complication is the development of seizures in perhaps 20% of individuals with autism; the risk for developing seizures is increased through childhood and adolescence. Depressive and anxiety symptoms may appear in higher functioning adolescents, who may become more aware of their areas of disability, become painfully aware of their inability to form friendships despite

a desire to do so, and begin to suffer the cumulative effect of years of failed contact with others and teasing.

### See Also the Following Articles

Child Development and Culture ■ Educational and Child Assessment ■ Psychological Assessment, Standards and Guidelines for

### Further Reading

Asperger, H. (1944). Die "Autistischen Psychopathen" im Kindersalter. *Archiv für Psychiatrie und Nervenkrankheiten*, 117, 76–136.

Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update.

*Journal of Autism and Developmental Disorders*, 33(4), 365–382.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 358(1430), 345–360.

National Research. (2001). *Educating young children with autism*. Washington, DC: National Academy Press.

Volkmar, F. R., Klin, A., Siegel, B., Szatmari, P., Lord, C., Campbell, M., et al. (1994). Field trial for autistic disorder in DSM-IV. *American Journal of Psychiatry*, 151(9), 1361–1367.

Wing, L. (1981). Asperger's syndrome: A clinical account. *Psychology and Medicine*, 11(1), 115–129.



# Aviation

Evan A. Byrne

National Transportation Safety Board, Washington, D.C., USA

1. History
  2. Scope and Extent of the Domain
  3. Sources of Basic Principles and Theoretical Approaches
  4. Applications and Ongoing Issues
  5. Education and Training
  6. Occupations and Employment
- Further Reading

## GLOSSARY

- aviation psychology** Scientific discipline focused on understanding and improving human performance in aviation.
- crew resource management** A collection of procedures, techniques, and communications strategies that fosters the use of all available resources (including people and available technology) to provide optimal levels of safety and mitigate the consequences of error.
- cultural factors** Procedural, communications, equipment, and design-related issues that are unique to an organization or a society.
- electroencephalogram (EEG)** Measurement of aggregate brain electrical activity recorded at the scalp.
- FAA** Abbreviation for Federal Aviation Administration.
- NASA** Abbreviation for National Aeronautics and Space Administration.
- National Airspace System** The complex system of people, equipment, and procedures that guides how aircraft are operated in the United States.
- NTSB** Abbreviation for National Transportation Safety Board.
- situation awareness** How people keep track of critical features in their environment, such as their position in space and time, and the current and future status of systems.

Aviation psychology is a simple name for a deceptively broad multidisciplinary applied domain. Its goal is plain: understanding and improving human performance in aviation. As a formal discipline, it draws on other established areas of psychology and engineering. For example, within the discipline of psychology, aspects of applied, experimental, social, clinical, industrial, organizational, and engineering specializations have been incorporated into aviation psychology.

## 1. HISTORY

The origins of aviation psychology are nearly as old as powered flight. Whether it played a role in the selection of the Wright *Flyer's* first pilot on the sands of Kitty Hawk, North Carolina, a century ago is debatable. Lindbergh's self-account of his 33-hour transatlantic crossing in the *Spirit of St. Louis* in 1927 is among the earliest accounts of the effects of sustained wakefulness on pilot performance. However, aviation psychology is generally described as having its formal beginning coincident with the increased military use of aviation in the two world wars. During those times, there was a fundamental need to rapidly select and adequately train the pilots required to fly the fighters, bombers, and transport airplanes for the war efforts. As aviation became increasingly important to worldwide commerce, and with the advent of the space program, aviation psychology as a discipline expanded during the postwar years. This expansion paralleled technological

advances that increased the complexity of some facets of aviation and the continuing need to understand human performance capabilities and limitations in these domains. Several detailed recent reviews of the domain exist that expand on the history and basic applications in this area.

## 2. SCOPE AND EXTENT OF THE DOMAIN

Enhancing pilot performance is a major focus of aviation psychology; however, its reach in aviation extends far beyond pilots. For example, early efforts of aviation psychologists included enhancing the training of aircraft gunners and selection of aircraft mechanics. Aviation psychology focuses on understanding the operator's basic limitations and needs. Basic needs for optimal performance typically include adequate selection and training, sufficient real-time information, appropriate human-machine interfaces and equipment, and compensated environmental conditions. The scope of aviation psychology has generally kept pace with the rapid expansion of aviation itself since the advent of powered flight.

The National Airspace System is described on the Federal Aviation Administration's (FAA) Web site as "a complex collection of systems, procedures, facilities, aircraft, and . . . people [that] represents the overall environment for the safe operation of aircraft." With its focus on the people and equipment involved in this complex system, aviation psychology is broad in scope but singular in purpose: to understand and promote optimal human performance. Aviation psychologists have contributed to our knowledge of human performance in both civilian and military applications. They have examined and sought ways in which to improve the performance of pilots, flight attendants, mechanics, air traffic controllers, and ground support personnel. It is clear that aviation psychologists now also have an important contribution to make in the study of human performance factors in civil aviation security, for example, in selecting and training security screening personnel and in identifying methods to augment and support human search capabilities during repetitive inspection tasks.

A peer-reviewed journal, the *International Journal of Aviation Psychology*, serves as a central focus for the dissemination of research in this area, and a survey of its contents can quickly provide a comprehensive

overview of this area. Other professional journals, such as *Human Factors* (published by the Human Factors and Ergonomics Society) and *Aviation, Space, and Environmental Medicine* (published by the Aerospace Medical Association), also contain work by aviation psychologists. In the latter case, there are some who may be more appropriately titled "aerospace psychologists" given that their focus involves understanding human performance in space-related applications such as low-g environments and crew coordination issues during long-duration space flights.

## 3. SOURCES OF BASIC PRINCIPLES AND THEORETICAL APPROACHES

Aviation psychology draws on basic principles, findings, and tenets from other specialized domains of psychology, sociology, physiology, and human factors. For example, knowledge of sensory processes, memory limitations, psychomotor skills, and cognitive functioning translates directly to the aviation environment and can be used to optimize the design of instruments, procedures, and warning devices. Other areas of experimental and cognitive psychology that provide insight into attention and decision making also contribute to aviation psychology in areas such as evaluating instrument scan behavior, understanding visual detection capabilities for pilots (or inspectors detecting defects in manufacturing/maintenance inspection processes), and aeronautical decision-making skills. Finally, knowledge of basic human performance capabilities under exposure to noise, vibration, motion, and altitude facilitates aviation psychologists' understanding of human capabilities and limitations.

Methods derived from industrial, organizational, and educational psychology can aid in the selection and training of aviation personnel. Components from cognitive psychology, social psychology, and sociology can contribute to the study of crew coordination and cultural factors in aviation. For example, cultural factors in aviation are not just associated with design stereotypes for the movement of controls and switches. They also include crew coordination issues that can vary across cultures, include variations in the acceptance of automated technology, and can even affect the coordination of operating practices when airlines merge.

Physiological psychology has contributed to the measurement of workload and assessment of fatigue in aviation environments. Although few operational

examples of the former exist, the use of physiological measures to infer the state of operators has taken place in basic and applied aviation research environments. These include traditional measures such as heart rate and electroencephalogram (EEG). In addition, eye movements have been used to study operator scan patterns for complex displays or while performing aviation-related tasks.

There are few, if any, theoretical approaches unique to aviation psychology. Instead, basic theories generated elsewhere are applicable to basic and applied questions in aviation. Some theoretical or worldview approaches are especially salient as applied to the aviation environment. For example, information processing theories that outline and predict how operators process information from the environment and manage conditions of increasing workload are useful in evaluating aviation work environments to ensure that workload and capabilities are not exceeded. Also, during recent years, significant effort has been focused on understanding and establishing theoretical and methodological approaches to the study of situation awareness, that is, how operators learn about and keep track of critical features in their environment, such as their position in space and time, and the current and future status of systems. Theoretical models for the acquisition and maintenance of situation awareness by personnel involved in aviation tasks have been developed, although acceptance of their utility is not universal.

Aviation psychology presents a unique application and research environment. However, with few exceptions (e.g., operations in extreme environments, operations under conditions of *g* loading, specific examination of flight- or domain-specific tasks such as the visual processes involved in landing), knowledge, methods, and concepts derived elsewhere in psychology and engineering can readily translate to applications and research questions in this area. For additional information on the basic tenets, principles, and theoretical approaches used by aviation psychologists, there exist recent compendiums of work outlining these issues from both historical and contemporary perspectives.

#### **4. APPLICATIONS AND ONGOING ISSUES**

Aviation psychology is an incremental and evolving discipline that is generally responsive to the dynamic challenges it faces as aviation itself continues to evolve.

Following are some of the areas of inquiry in which aviation psychologists are engaged.

##### **4.1. Error Management**

Studies and training methods for crew coordination have evolved from attending static personality and “charm school” sessions to delivering strategies for the prompt identification and timely recovery from human errors that are inevitable. Work continues on determining optimal methods for enhancing crew coordination and error management, on developing instructional methods for distributing these error management concepts through technical training, and on reinforcing during routine operations through robust procedures.

##### **4.2. Decision Making and High-Stress Situations**

Research continues on understanding how personnel in the aviation environment arrive at decisions under conditions of stress, often having limited information. Some of these efforts are targeted to developing methods to enhance pilot decision making and judgment. Other efforts examine factors that affect human performance in stressful situations such as passengers evacuating an airplane under an emergency situation.

##### **4.3. Fatigue Countermeasures**

Substantial work in the area of fatigue has been performed by aviation psychologists at the National Aeronautical and Space Administration (NASA) Ames Research Center, and work continues to better understand the effects of fatigue on aviation operations. Research conducted by the FAA has looked at shiftwork and fatigue in air traffic control operations. Aviation psychologists working for the military have examined scheduling practices and pharmacological countermeasures to mitigate the effects of fatigue in military aviation.

##### **4.4. Automation and Human–Machine Interfaces**

The study of human interaction with automation remains a dominant focus for many aviation psychologists. The use of automated aids is increasing in both air carrier and general aviation, and this trend is expected to continue with the refinement of more complex aircraft

system status displays, horizontal and vertical navigation displays, airport situation awareness displays, highway-in-the-sky flight path aids, and the like. Automated aids are also increasingly being applied to other aviation operations such as air traffic control and maintenance.

#### **4.5. New Training Technologies**

Simulators have found widespread use in the training of air carrier flight crews. Methods for enhancing the training environment using these tools continue to be evaluated. The use of computer-based training and part task simulators is also spreading to other areas of aviation, including general aviation and maintenance technicians. Ensuring that the transfer of training effects is positive and predictable will continue to be a focus for aviation psychologists.

#### **4.6. Accident Reduction**

Methods to study human error in accidents continue to be examined to identify broad causal factors common to multiple accidents for which targeted intervention strategies could be applied. Strategies for this work include the use of surveys, the use of simulator studies, and review of accident data.

#### **4.7. Maintenance and Other Aviation Applications**

Some areas that have received relatively less attention in the past, as compared with flight deck and pilot issues, are receiving increasing attention by aviation psychologists. These include maintenance and ground support personnel, and recent accidents have shown that human error in these areas can also yield catastrophic outcomes. Strategies for increasing performance on the flight deck, such as crew resource management, are being translated to optimize performance by operators in these areas. Similarly, additional focus is being placed on the information needs of maintenance personnel to complement work done previously to examine environmental and ergonomics needs.

### **5. EDUCATION AND TRAINING**

Because of the multidisciplinary characteristics of aviation psychology, there is no single training or education path to become an aviation psychologist (although

for decades some institutions, such as the University of Illinois and Ohio State University, have produced professionals who made significant contributions in this area). Like the applied domain of human factors, successful performance in aviation psychology requires a professional who has developed a strong grounding in basic psychological principles, experimental methodology, and basic familiarity with engineering concepts. Creativity and the ability to adapt are essential traits that facilitate interaction with experts from the many facets of aviation, and formulation of strategies to cope with the short-cycle time periods (from question to required answer) characteristic of many issues aviation psychologists may be called on to study.

Equally important to becoming an effective aviation psychologist is a strong interest in, some observational experience in, or (optimally) direct experience in an operational facet of aviation (e.g., holding a pilot certificate). This experience will also help to facilitate communication with subject matter experts and provide the aviation psychologist with practical knowledge that will augment his or her professional training in a chosen area of psychology or in an allied discipline. Aviation has many regulatory and procedural boundaries that have been established by regulators, manufacturers, companies, and labor groups. These areas, whether they involve hours of service regulations for pilots or the maintenance and inspection routine, should be mastered to the extent possible for an aviation psychologist to contribute optimally.

## **6. OCCUPATIONS AND EMPLOYMENT**

Aviation psychologists can be found employed by organizations in and on the periphery of the aerospace industry. For example, they are employed by agencies such as NASA, the FAA, and the National Transportation Safety Board (NTSB) in the federal government and are involved in basic research, certification, design, oversight, and safety investigations. In addition, most branches of the U.S. Department of Defense (DoD) have similar areas in which aviation psychologists are employed to facilitate military aviation. In the private sector, aviation psychologists are employed by manufacturers, airlines, and private research institutions. Academia is another area where aviation psychologists can be found working on basic and applied research questions, often funded by DoD, NASA, and/or FAA grants and projects.

### See Also the Following Articles

Accidents in Transportation ■ Engineering Psychology  
 ■ Transportation Systems, Overview

### Further Reading

- Billings, C. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum.
- Caldwell, J. A., & Gilreath, S. R. (2002). A survey of aircrew fatigue in a sample of U.S. Army aviation personnel. *Aviation, Space, and Environmental Medicine*, 73, 472–480.
- Dismukes, R. K. (1994). Aviation human factors research in U.S. universities: Potential contributions to national needs. *International Journal of Aviation Psychology*, 4, 315–331.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32–64.
- Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors*, 27, 149–157.
- Garland, D. J., Wise, J. A., & Hopkins, V. D. (Eds.). (1999). *Handbook of aviation human factors*. Mahwah, NJ: Lawrence Erlbaum.
- Guilford, J. P. (1948). Some lessons from aviation psychology. *American Psychologist*, 3, 3–11.
- Helmreich, R. L., & Merritt, A. C. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences*. Aldershot, UK: Ashgate.
- Koonce, J. M. (1984). A brief history of aviation psychology. *Human Factors*, 26, 499–508.
- O'Hare, D., & Chalmers, D. (1999). The incidence of incidents: A nationwide study of flight experience and exposure to accidents and incidents. *International Journal of Aviation Psychology*, 9(1), 1–18.
- O'Hare, D., & Lawrence, B. (2000). The shape of aviation psychology: A review of articles published in the first 5 years of the *International Journal of Aviation Psychology*. *International Journal of Aviation Psychology*, 10(1), 1–11.
- Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors*, 43, 641–674.
- Tsang, P. S., & Vidulich, M. A. (2002). *Principles and practices of aviation psychology*. Mahwah, NJ: Lawrence Erlbaum.
- Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.
- Wickens, C. D., & Andre, A. D. (1999). Psychology applied to aviation. In A. M. Stec, & D. A. Bernstein (Eds.), *Psychology: Fields of application* (pp. 184–198). Boston: Houghton Mifflin.
- Wilson, G. F. (Ed.). (2002). Flight psychophysiology [special issue]. *International Journal of Aviation Psychology*, 12(1).
- Zsombok, C. E., & Klein, G. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ: Lawrence Erlbaum.



# Aviation

Evan A. Byrne

National Transportation Safety Board, Washington, D.C., USA

1. History
  2. Scope and Extent of the Domain
  3. Sources of Basic Principles and Theoretical Approaches
  4. Applications and Ongoing Issues
  5. Education and Training
  6. Occupations and Employment
- Further Reading

## GLOSSARY

- aviation psychology** Scientific discipline focused on understanding and improving human performance in aviation.
- crew resource management** A collection of procedures, techniques, and communications strategies that fosters the use of all available resources (including people and available technology) to provide optimal levels of safety and mitigate the consequences of error.
- cultural factors** Procedural, communications, equipment, and design-related issues that are unique to an organization or a society.
- electroencephalogram (EEG)** Measurement of aggregate brain electrical activity recorded at the scalp.
- FAA** Abbreviation for Federal Aviation Administration.
- NASA** Abbreviation for National Aeronautics and Space Administration.
- National Airspace System** The complex system of people, equipment, and procedures that guides how aircraft are operated in the United States.
- NTSB** Abbreviation for National Transportation Safety Board.
- situation awareness** How people keep track of critical features in their environment, such as their position in space and time, and the current and future status of systems.

Aviation psychology is a simple name for a deceptively broad multidisciplinary applied domain. Its goal is plain: understanding and improving human performance in aviation. As a formal discipline, it draws on other established areas of psychology and engineering. For example, within the discipline of psychology, aspects of applied, experimental, social, clinical, industrial, organizational, and engineering specializations have been incorporated into aviation psychology.

## 1. HISTORY

The origins of aviation psychology are nearly as old as powered flight. Whether it played a role in the selection of the Wright *Flyer's* first pilot on the sands of Kitty Hawk, North Carolina, a century ago is debatable. Lindbergh's self-account of his 33-hour transatlantic crossing in the *Spirit of St. Louis* in 1927 is among the earliest accounts of the effects of sustained wakefulness on pilot performance. However, aviation psychology is generally described as having its formal beginning coincident with the increased military use of aviation in the two world wars. During those times, there was a fundamental need to rapidly select and adequately train the pilots required to fly the fighters, bombers, and transport airplanes for the war efforts. As aviation became increasingly important to worldwide commerce, and with the advent of the space program, aviation psychology as a discipline expanded during the postwar years. This expansion paralleled technological

advances that increased the complexity of some facets of aviation and the continuing need to understand human performance capabilities and limitations in these domains. Several detailed recent reviews of the domain exist that expand on the history and basic applications in this area.

## 2. SCOPE AND EXTENT OF THE DOMAIN

Enhancing pilot performance is a major focus of aviation psychology; however, its reach in aviation extends far beyond pilots. For example, early efforts of aviation psychologists included enhancing the training of aircraft gunners and selection of aircraft mechanics. Aviation psychology focuses on understanding the operator's basic limitations and needs. Basic needs for optimal performance typically include adequate selection and training, sufficient real-time information, appropriate human-machine interfaces and equipment, and compensated environmental conditions. The scope of aviation psychology has generally kept pace with the rapid expansion of aviation itself since the advent of powered flight.

The National Airspace System is described on the Federal Aviation Administration's (FAA) Web site as "a complex collection of systems, procedures, facilities, aircraft, and . . . people [that] represents the overall environment for the safe operation of aircraft." With its focus on the people and equipment involved in this complex system, aviation psychology is broad in scope but singular in purpose: to understand and promote optimal human performance. Aviation psychologists have contributed to our knowledge of human performance in both civilian and military applications. They have examined and sought ways in which to improve the performance of pilots, flight attendants, mechanics, air traffic controllers, and ground support personnel. It is clear that aviation psychologists now also have an important contribution to make in the study of human performance factors in civil aviation security, for example, in selecting and training security screening personnel and in identifying methods to augment and support human search capabilities during repetitive inspection tasks.

A peer-reviewed journal, the *International Journal of Aviation Psychology*, serves as a central focus for the dissemination of research in this area, and a survey of its contents can quickly provide a comprehensive

overview of this area. Other professional journals, such as *Human Factors* (published by the Human Factors and Ergonomics Society) and *Aviation, Space, and Environmental Medicine* (published by the Aerospace Medical Association), also contain work by aviation psychologists. In the latter case, there are some who may be more appropriately titled "aerospace psychologists" given that their focus involves understanding human performance in space-related applications such as low-g environments and crew coordination issues during long-duration space flights.

## 3. SOURCES OF BASIC PRINCIPLES AND THEORETICAL APPROACHES

Aviation psychology draws on basic principles, findings, and tenets from other specialized domains of psychology, sociology, physiology, and human factors. For example, knowledge of sensory processes, memory limitations, psychomotor skills, and cognitive functioning translates directly to the aviation environment and can be used to optimize the design of instruments, procedures, and warning devices. Other areas of experimental and cognitive psychology that provide insight into attention and decision making also contribute to aviation psychology in areas such as evaluating instrument scan behavior, understanding visual detection capabilities for pilots (or inspectors detecting defects in manufacturing/maintenance inspection processes), and aeronautical decision-making skills. Finally, knowledge of basic human performance capabilities under exposure to noise, vibration, motion, and altitude facilitates aviation psychologists' understanding of human capabilities and limitations.

Methods derived from industrial, organizational, and educational psychology can aid in the selection and training of aviation personnel. Components from cognitive psychology, social psychology, and sociology can contribute to the study of crew coordination and cultural factors in aviation. For example, cultural factors in aviation are not just associated with design stereotypes for the movement of controls and switches. They also include crew coordination issues that can vary across cultures, include variations in the acceptance of automated technology, and can even affect the coordination of operating practices when airlines merge.

Physiological psychology has contributed to the measurement of workload and assessment of fatigue in aviation environments. Although few operational

examples of the former exist, the use of physiological measures to infer the state of operators has taken place in basic and applied aviation research environments. These include traditional measures such as heart rate and electroencephalogram (EEG). In addition, eye movements have been used to study operator scan patterns for complex displays or while performing aviation-related tasks.

There are few, if any, theoretical approaches unique to aviation psychology. Instead, basic theories generated elsewhere are applicable to basic and applied questions in aviation. Some theoretical or worldview approaches are especially salient as applied to the aviation environment. For example, information processing theories that outline and predict how operators process information from the environment and manage conditions of increasing workload are useful in evaluating aviation work environments to ensure that workload and capabilities are not exceeded. Also, during recent years, significant effort has been focused on understanding and establishing theoretical and methodological approaches to the study of situation awareness, that is, how operators learn about and keep track of critical features in their environment, such as their position in space and time, and the current and future status of systems. Theoretical models for the acquisition and maintenance of situation awareness by personnel involved in aviation tasks have been developed, although acceptance of their utility is not universal.

Aviation psychology presents a unique application and research environment. However, with few exceptions (e.g., operations in extreme environments, operations under conditions of *g* loading, specific examination of flight- or domain-specific tasks such as the visual processes involved in landing), knowledge, methods, and concepts derived elsewhere in psychology and engineering can readily translate to applications and research questions in this area. For additional information on the basic tenets, principles, and theoretical approaches used by aviation psychologists, there exist recent compendiums of work outlining these issues from both historical and contemporary perspectives.

#### **4. APPLICATIONS AND ONGOING ISSUES**

Aviation psychology is an incremental and evolving discipline that is generally responsive to the dynamic challenges it faces as aviation itself continues to evolve.

Following are some of the areas of inquiry in which aviation psychologists are engaged.

##### **4.1. Error Management**

Studies and training methods for crew coordination have evolved from attending static personality and “charm school” sessions to delivering strategies for the prompt identification and timely recovery from human errors that are inevitable. Work continues on determining optimal methods for enhancing crew coordination and error management, on developing instructional methods for distributing these error management concepts through technical training, and on reinforcing during routine operations through robust procedures.

##### **4.2. Decision Making and High-Stress Situations**

Research continues on understanding how personnel in the aviation environment arrive at decisions under conditions of stress, often having limited information. Some of these efforts are targeted to developing methods to enhance pilot decision making and judgment. Other efforts examine factors that affect human performance in stressful situations such as passengers evacuating an airplane under an emergency situation.

##### **4.3. Fatigue Countermeasures**

Substantial work in the area of fatigue has been performed by aviation psychologists at the National Aeronautical and Space Administration (NASA) Ames Research Center, and work continues to better understand the effects of fatigue on aviation operations. Research conducted by the FAA has looked at shiftwork and fatigue in air traffic control operations. Aviation psychologists working for the military have examined scheduling practices and pharmacological countermeasures to mitigate the effects of fatigue in military aviation.

##### **4.4. Automation and Human–Machine Interfaces**

The study of human interaction with automation remains a dominant focus for many aviation psychologists. The use of automated aids is increasing in both air carrier and general aviation, and this trend is expected to continue with the refinement of more complex aircraft

system status displays, horizontal and vertical navigation displays, airport situation awareness displays, highway-in-the-sky flight path aids, and the like. Automated aids are also increasingly being applied to other aviation operations such as air traffic control and maintenance.

#### **4.5. New Training Technologies**

Simulators have found widespread use in the training of air carrier flight crews. Methods for enhancing the training environment using these tools continue to be evaluated. The use of computer-based training and part task simulators is also spreading to other areas of aviation, including general aviation and maintenance technicians. Ensuring that the transfer of training effects is positive and predictable will continue to be a focus for aviation psychologists.

#### **4.6. Accident Reduction**

Methods to study human error in accidents continue to be examined to identify broad causal factors common to multiple accidents for which targeted intervention strategies could be applied. Strategies for this work include the use of surveys, the use of simulator studies, and review of accident data.

#### **4.7. Maintenance and Other Aviation Applications**

Some areas that have received relatively less attention in the past, as compared with flight deck and pilot issues, are receiving increasing attention by aviation psychologists. These include maintenance and ground support personnel, and recent accidents have shown that human error in these areas can also yield catastrophic outcomes. Strategies for increasing performance on the flight deck, such as crew resource management, are being translated to optimize performance by operators in these areas. Similarly, additional focus is being placed on the information needs of maintenance personnel to complement work done previously to examine environmental and ergonomics needs.

### **5. EDUCATION AND TRAINING**

Because of the multidisciplinary characteristics of aviation psychology, there is no single training or education path to become an aviation psychologist (although

for decades some institutions, such as the University of Illinois and Ohio State University, have produced professionals who made significant contributions in this area). Like the applied domain of human factors, successful performance in aviation psychology requires a professional who has developed a strong grounding in basic psychological principles, experimental methodology, and basic familiarity with engineering concepts. Creativity and the ability to adapt are essential traits that facilitate interaction with experts from the many facets of aviation, and formulation of strategies to cope with the short-cycle time periods (from question to required answer) characteristic of many issues aviation psychologists may be called on to study.

Equally important to becoming an effective aviation psychologist is a strong interest in, some observational experience in, or (optimally) direct experience in an operational facet of aviation (e.g., holding a pilot certificate). This experience will also help to facilitate communication with subject matter experts and provide the aviation psychologist with practical knowledge that will augment his or her professional training in a chosen area of psychology or in an allied discipline. Aviation has many regulatory and procedural boundaries that have been established by regulators, manufacturers, companies, and labor groups. These areas, whether they involve hours of service regulations for pilots or the maintenance and inspection routine, should be mastered to the extent possible for an aviation psychologist to contribute optimally.

## **6. OCCUPATIONS AND EMPLOYMENT**

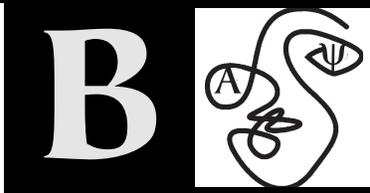
Aviation psychologists can be found employed by organizations in and on the periphery of the aerospace industry. For example, they are employed by agencies such as NASA, the FAA, and the National Transportation Safety Board (NTSB) in the federal government and are involved in basic research, certification, design, oversight, and safety investigations. In addition, most branches of the U.S. Department of Defense (DoD) have similar areas in which aviation psychologists are employed to facilitate military aviation. In the private sector, aviation psychologists are employed by manufacturers, airlines, and private research institutions. Academia is another area where aviation psychologists can be found working on basic and applied research questions, often funded by DoD, NASA, and/or FAA grants and projects.

### See Also the Following Articles

- Accidents in Transportation ■ Engineering Psychology
- Transportation Systems, Overview

### Further Reading

- Billings, C. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum.
- Caldwell, J. A., & Gilreath, S. R. (2002). A survey of aircrew fatigue in a sample of U.S. Army aviation personnel. *Aviation, Space, and Environmental Medicine, 73*, 472–480.
- Dismukes, R. K. (1994). Aviation human factors research in U.S. universities: Potential contributions to national needs. *International Journal of Aviation Psychology, 4*, 315–331.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.
- Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors, 27*, 149–157.
- Garland, D. J., Wise, J. A., & Hopkins, V. D. (Eds.). (1999). *Handbook of aviation human factors*. Mahwah, NJ: Lawrence Erlbaum.
- Guilford, J. P. (1948). Some lessons from aviation psychology. *American Psychologist, 3*, 3–11.
- Helmreich, R. L., & Merritt, A. C. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences*. Aldershot, UK: Ashgate.
- Koonce, J. M. (1984). A brief history of aviation psychology. *Human Factors, 26*, 499–508.
- O'Hare, D., & Chalmers, D. (1999). The incidence of incidents: A nationwide study of flight experience and exposure to accidents and incidents. *International Journal of Aviation Psychology, 9*(1), 1–18.
- O'Hare, D., & Lawrence, B. (2000). The shape of aviation psychology: A review of articles published in the first 5 years of the *International Journal of Aviation Psychology*. *International Journal of Aviation Psychology, 10*(1), 1–11.
- Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors, 43*, 641–674.
- Tsang, P. S., & Vidulich, M. A. (2002). *Principles and practices of aviation psychology*. Mahwah, NJ: Lawrence Erlbaum.
- Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.
- Wickens, C. D., & Andre, A. D. (1999). Psychology applied to aviation. In A. M. Stec, & D. A. Bernstein (Eds.), *Psychology: Fields of application* (pp. 184–198). Boston: Houghton Mifflin.
- Wilson, G. F. (Ed.). (2002). Flight psychophysiology [special issue]. *International Journal of Aviation Psychology, 12*(1).
- Zsombok, C. E., & Klein, G. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ: Lawrence Erlbaum.



# Behavioral Assessment in Schools

Steven W. Evans, Amy L. Williams,  
and Brandon Schultz

James Madison University, Harrisonburg, Virginia, USA

Mark D. Weist

University of Maryland at Baltimore, Baltimore,  
Maryland, USA

1. Rationale for School-Based Assessment
  2. Evaluating Behavioral and Emotional Problems
  3. Evaluating Outcomes
  4. Legal and Ethical Issues
- Further Reading

## GLOSSARY

**analogue assessment** An observational measure of the child's behavior(s) of interest elicited through the use of simulated conditions to predict behavior in real-life settings.

**direct observation** Observation and recording of problem behaviors in the context in which they occur.

**effect size** A calculation of preintervention and postintervention scores as well as the standard deviation or standard error of the instrument to produce an index of the amount of change assessed for the individual or group.

**error variance** The contribution of various sources of error (e.g., biases, drift) to a set of data.

**functional behavioral assessment (FBA)** Assessment of environmental conditions that are correlated with the occurrence and nonoccurrence of behaviors for the purpose of identifying those things that serve to initiate and maintain behavior.

**generalization** The degree to which learned skills or changes in behavior patterns are displayed in untrained situations.

**informed consent** The legal and ethical requirement to inform children and their legal guardians regarding the pur-

pose of assessment procedures and to receive appropriate agreement to these procedures prior to their initiation.

**interviews** Direct questioning of the child himself or herself, or of the child's parents, teachers, and/or other caregivers, for the purpose of assessing strengths, weaknesses, and patterns of behavior; interviews can be highly structured (following a strict line of questioning), unstructured (based entirely on the interviewer's clinical direction), or semistructured (some combination thereof).

**mediators/moderators** Characteristics of the individuals receiving interventions and characteristics of the intervention process that affect treatment outcomes.

**normative sample** A group of people identified as being representative of a population who are assessed, with their data being used to compare with results obtained from subsequent administrations of the assessment instrument.

**rating scales** Assessment instruments that provide a standardized format to gather information about an individual's behavioral characteristics; these are usually paper-and-pencil forms that may be completed outside of an assessment setting.

**response bias** Trends in the way in which informants respond to assessments that result in error variance.

**traditional assessment** Behavioral assessment strategies that focus on assessing the presence or absence of diagnostic categories.

The assessment of social and emotional functioning in the schools is needed for purposes of classification, intervention development, and measurement of

outcomes. Procedures for completing these evaluations, as well as the relative strengths and weaknesses of the techniques, are described in this article. The descriptions are organized by purpose, beginning with techniques for problem identification and followed by the application of these techniques to measuring outcomes. Legal, ethical, and cultural issues as they relate to these procedures are described at the end.

## **1. RATIONALE FOR SCHOOL-BASED ASSESSMENT**

There are two primary contexts for behavioral assessment in schools: special education and expanded school mental health (ESMH). When considering behavioral assessment in the context of special education, it is important to keep in mind that these services are guided in part by legislative acts such as the Individuals with Disabilities Education Act (IDEA) and the amendments to the IDEA in 1997 (IDEA '97). These legislative acts provide funds for public schools to fulfill the mandate of a "free and appropriate public education" for all children and adolescents with disabilities, including those with emotional and behavioral disabilities. In this context, behavioral assessment of students is critical for two reasons: (a) to assist in understanding how behavioral and emotional issues interfere with learning and (b) to develop appropriate intervention strategies to enhance learning success. Within IDEA '97, there is specific reference to the role of "functional behavioral assessment" (FBA) for the analysis and monitoring of student behavior. Although the law stops short of defining this requirement, FBA is widely assumed to include interviews, structured observations, and assessment of the function or purpose of student behavior. Based on the language of IDEA '97, FBA is intended to go beyond the mere question of eligibility for special education services or simple "gatekeeping." Rather, its intended purpose is to address those issues surrounding the creation and monitoring of appropriate interventions. To this end, IDEA '97 requires school professionals to assess behaviors that impede a student's ability to learn or that impede the ability for his or her peers to learn, although the language of the law does not specifically define those behaviors. Legal precedents based on case law suggest that physical aggression, noncompliance, verbal abuse, and disruptive classroom behaviors can rise to this level. For those students with disabilities who exhibit these types of behaviors, a decision must

be made as to the severity of the behaviors and whether or not they warrant specific interventions. If so, school professionals are legally bound to design a behavioral intervention plan (BIP) to address these needs. IDEA is being reauthorized in the U.S. Congress, and many hope that with this reauthorization there will be increased funding and an emphasis on prevention and support for youth before formal referral into special education. These changes would increase the need for assessment earlier in the process than is currently required.

Whereas IDEA '97 is applicable only to students in public schools who meet diagnostic criteria for specific areas of disability, Section 504 of the Rehabilitation Act of 1973 protects individuals from discrimination in any institution receiving federal funding, based on a much wider range of disabling conditions. Section 504 is not program legislation but rather civil rights legislation. In addition to its application to schools as a mechanism to receive special services, this act also requires that school districts ensure that all of their procedures and techniques are nondiscriminatory. When conducting behavioral assessment, professionals must be cognizant of potential cultural biases that may occur within the process. For example, behavioral assessment instruments that are standardized and norm referenced with samples that do not reflect the cultural differences of minority students may classify their behavior inaccurately. In such instances, it is imperative for school professionals to interpret results within the appropriate framework of the students' cultures and backgrounds. In other words, results from instruments that are based on norm samples that do not accurately reflect students' cultural backgrounds must be qualified. This is particularly true if these instruments have the potential for overidentifying minority students for disabilities based solely on cultural differences. Conversely, it is also important to not deny minority students special education services as a result of attempting to be culturally competent or politically correct.

The second primary context for behavioral assessment in schools is ESMH programs. ESMH programs involve collaboration between schools and community agencies to move toward a full continuum of mental health promotion, including early intervention and treatment for youth in special and regular education. These programs are developing progressively in the United States, as well as in other countries, as a result of their documented ability to reach underserved youth and preliminary evidence that they lead to outcomes valued by schools and communities (e.g., improved student attendance and behavior, decreased inappropriate referrals to

special education). Behavioral assessment is of foundational importance in the ESMH context and for special education and Section 504 plans to evaluate the nature of the presenting behavioral and emotional problems, provide guidance in the development of effective interventions, and measure the outcomes of those interventions.

## 2. EVALUATING BEHAVIORAL AND EMOTIONAL PROBLEMS

### 2.1. Purpose

The evaluation of behavioral and emotional problems in the classroom is warranted by the difficulties experienced by children who contend with these problems on a day-to-day basis. A number of behavioral and emotional problems that can prevent a child from functioning at a level commensurate with his or her same-age or same-grade peers are commonly evidenced in the school setting. Some of the more common behavioral and emotional problems that cause such impairments include anxiety, attention deficit/hyperactivity disorder (ADHD), anger/aggression, attachment disorders, conduct problems, depression, delinquency, posttraumatic stress disorder (PTSD), and bullying. The school environment is unique in that it places demands on children that are not otherwise required of them. For example, in the school setting, children are expected to sit quietly and attend to classroom activities. Furthermore, schools require students to function within demanding social contexts that youth may avoid or not be exposed to outside of school. The difficulties these expectations pose for children with behavioral and emotional problems are substantial, and assessment of these problems in the setting where they often occur has important implications for the development of effective interventions. It logically follows that assessment should occur, at least in part, in the school setting.

Evaluation of behavioral and emotional problems in schools is frequently done to determine whether a child or an adolescent is eligible for, or would benefit from, special services. Under these circumstances, evaluation is done to answer questions pertaining to classification and to guide the development of effective interventions. Although determining eligibility for special services is important, it is the development of interventions that is vital for changing problem behaviors. Effective interventions are those that target the presenting behaviors and not the label. Therefore, school professionals should place an emphasis on the development and

implementation of interventions that target the specific needs of the students who are referred for evaluation rather than their diagnoses or labels.

### 2.2. Traditional Assessment

There are a number of assessment methods used to evaluate behavioral and emotional problems in the schools. Traditional assessment involves the examination of scores from norm-referenced measures to determine a particular individual's difference from or similarity to a normative sample. In other words, traditional assessment asks the question, "How does this student's behavior compare with that of similar children?" Standard tests of cognitive abilities and academic achievement, as well as many rating scales, are structured in this way. Using traditional assessment techniques, the inferred construct being measured (e.g., attention, depression) is thought of as a cause of a particular behavior in a specific setting. Although identifying the underlying syndrome or "label" related to a child's behavior is certainly important for diagnosis and classification, this information is not always the most useful for designing interventions. In contrast, measurement approaches such as behavioral observation, analogue assessment, interviews, and rating scales provide highly useful information about the functional relationship between the environment and particular behaviors. These techniques attempt to uncover the causes and consequences of behavior that often lead directly to effective intervention strategies.

### 2.3. Direct Observation

Direct observation is likely the most straightforward of all assessment techniques because it involves just what it implies: the direct observation and recording of problem behaviors in the context in which they occur. Three methods of direct observation are narrative recording (i.e., writing down what is happening in a narrative fashion, often following the antecedent-behavior-consequence [ABC] pattern or what occurs before, during, and after specific behaviors), event recording (i.e., tallying the instance of a target behavior), and time sampling (i.e., breaking down observations into intervals and recording the number of times a behavior occurs relative to the number of intervals observed to form a ratio). Direct observation differs from other methods of assessment in that little inference is required, making it one of the less biased methods for evaluating potential target behaviors.

There are some measurement issues related to observational methods that can negatively influence the data that are collected, including poorly trained observers, observer drift, poor operational definitions of target behaviors, poorly constructed data-recording systems, and reactivity (i.e., when the act of observation affects the student's behavior). Establishing interobserver agreement is one way in which to ensure that the data collected from direct observation are reliable and useful. A common procedure used to measure and enhance the quality of direct observation data is having another trained observer collect data simultaneously with the primary observer and then comparing data across observers to yield a measure of interobserver agreement. Direct observation is a labor intensive procedure that requires significant training time and effort to maintain reliability.

An additional problem with direct observation is the limited availability of the technique to assess infrequently occurring behavior. For example, children who are perceived as aggressive by teachers might exhibit only two or three episodes of these behaviors per week (or even fewer). The likelihood of observing these defiant or aggressive behaviors during discrete classroom observations is small. The technique is probably most practical for high-frequency behaviors such as on-task behavior and "talking out."

Finally, most observational data are difficult to put into context because there are no norms. In the school environment there is considerable variability in behavioral expectations across classroom activities and between teachers, resulting in various and shifting definitions of what is "normal." There are observational techniques that allow one to begin to account for this shifting context. For example, some time-sampling procedures require the observer to shift back and forth between observing the target child and same-sex classmates or a randomly selected peer. Pooling the data from the target and peer intervals provides the evaluator with a definition of what was normal during the context observed. However, no strategy will always work to overcome the problem of shifting contexts, and school professionals must interpret the results of their observations carefully.

#### **2.4. Analogue Assessment**

Analogue assessment provides an alternative to direct observation and is useful in situations where direct observation is not feasible. Analogue assessment is an observational measure of the child's behavior of

interest elicited through the use of simulated conditions; observations are made in the simulated setting to predict behavior in real-life settings. Assessments of this type can be administered by paper and pencil, by audiotape or videotape, or through the use of role-plays or enactments. In a controlled environment, the evaluator creates a situation parallel to one that the child experiences in school and observes and records the behavior of interest. As with direct observation methods, this technique can be labor intensive; however, it does allow the evaluator to assess specific conditions that may occur only rarely. Some of these situations may be easy to recreate, for example, asking a child who is poor in math to complete some math and reading work. In this example, differences in the child's reactions to the various academic tasks may be observed, and these data may contribute to an understanding of behavior problems experienced during math class.

Among the problems associated with this method of assessment is the lack of standardization, issues related to reliability and validity, and problems with generalization. The most basic of these is the fact that analogue assessments are contrived situations, and regardless of how similar the situation is to the targeted environment, the child usually knows that it is contrived, and this knowledge may influence his or her behavior. Whenever analogue assessment techniques are used, all of these potential issues should be considered when interpreting the results.

#### **2.5. Interviews**

Interviews to assess behavioral and emotional problems in children can be conducted with the child himself or herself or with the child's parents, teachers, and/or other caregivers. A variety of different interview formats exist. An omnibus interview is designed for gathering a wide range of information, whereas a behavior-specific interview is narrower in scope and focuses on the assessment of specific problem areas. An interview that is problem solving in nature focuses on presenting concerns with the goal of developing an intervention plan. Interviews can be structured, unstructured, or semistructured.

Although it is often the case that interviews focus on reaching a diagnostic conclusion, interviews that focus on both problem assessment and the development of intervention plans are useful in school settings. Behavioral interviewing addresses the assessment of the current behaviors and environmental factors that contribute to the initiation and maintenance of problem

behaviors. These interviews are designed to detect specific problems that can be targeted for intervention and include phases of problem identification, problem analysis, treatment implementation, and treatment evaluation.

Interviews may be a valuable and efficient method for collecting information about a wide range of behavioral and emotional problems. The evaluator can efficiently collect data about behavior over a long period of time and in a variety of situations. In addition, the adults being interviewed may provide information about their reactions to this behavior or interventions related to the targeted behavior. Ideally, interviews are conducted by staff members who will continue to be involved in developing and implementing interventions with students in need. In these cases, the interview can be used to establish the initial relationship and communicate empathy and support to the child while collecting valuable information.

## 2.6. Rating Scales

Rating scale measures provide a standardized format to gather information about an individual's behavioral characteristics and can be administered to children to obtain self-report data or to parents, teachers, and/or other caregivers to obtain informant-report data. A few of the advantages of rating scales are their ability to provide behavioral information in a short amount of time, their moderate cost, and their technical precision and practical utility. Rating scales can be used in the comprehensive assessment of potential behavior problems and deficits in functioning for screening, diagnosis/classification, treatment selection/intervention planning, progress monitoring, and outcome evaluation.

As with most other forms of assessment, there are some problems associated with the use of rating scales. These problems typically fall under one of two categories: response bias and error variance. Response bias refers to trends in the way in which informants respond to rating scales. Some examples of response bias are halo effects (i.e., rating the student in a positive or negative manner based on characteristics unrelated to what is being measured), leniency or severity effects (i.e., the tendency to have an overly generous or overly critical response set), and central tendency effects (i.e., the tendency to rate down the middle and avoid end point ratings). Error variance refers to the differences among several ratings of the same individual. The different types of error variance include source variance (i.e., differences in ratings provided by various persons), setting variance (i.e., differences in ratings across

situations), temporal variance (i.e., inconsistency of behavior ratings over time), and instrument variance (i.e., differences in ratings using different instruments).

Although rating scales are valuable tools, it is frequently tempting to overrely on them due to their minimal costs. Although they are efficient, there are many questions they do not answer (e.g., the role of environmental contingencies). Although environmental concerns are critical to the development of interventions, this information is generally not provided by rating scales. Furthermore, rating scales are only as useful as the sources. For example, children are fairly poor raters of their own disruptive, off-task, and inappropriate behaviors; consequently, adults have been identified as a better source for this information. On the other hand, children appear to be better at reporting their internal affective states and, as a result, are generally better raters of their own depressive and anxious characteristics. However, although these are general trends that should be considered when interpreting rating scales, there is considerable individual variability.

## 2.7. Functional Behavioral Assessment

A primary goal of FBA is to identify environmental conditions that are correlated with the occurrence and nonoccurrence of behaviors. This type of assessment provides a systematic analytic method for identifying those things that serve to initiate and maintain behavior. FBA incorporates many methods of assessment to gain a complete understanding of an individual's behavior in a particular environmental context. Approaches to FBA can be indirect (i.e., gathering behavioral information from rating scales, interviews, etc.), direct (i.e., directly observing the behavior in the context of the natural environment in which it occurs), or experimental (i.e., generating hypotheses about the maintaining contingencies for the problem behavior and testing them with the individual serving as his or her own control). Completing these assessments is an iterative process of data collection, hypotheses generation, and collection of additional data. The hypotheses provide the specific questions that guide the subsequent data collection that either confirms or redirects the hypotheses.

The identification of the cause-effect relations in the interaction of an individual and his or her environment is integral to the development of interventions aimed at changing behavior. Because FBA focuses on the relationship between particular environmental circumstances

and behavior, it has the potential to contribute greatly to the development of effective interventions for problem behaviors. A successful FBA provides information about these environmental circumstances affecting the problem behavior and also provides a clear direction for treatment.

### **2.8. Current Issues**

Because all of the assessment techniques discussed present their own unique advantages and limitations, the use of multiple assessment techniques with multiple informants is often considered the best method for evaluating behavioral and emotional problems. Although the use of multiple informants (e.g., parents, teachers, students) has the potential to provide a wealth of information, inconsistencies in data from one source and method to another is an issue that is often encountered in behavioral and emotional assessment. Such inconsistencies must be taken into consideration when assessing behavioral and emotional problems and especially when identifying behaviors to target for intervention. Frequently, careful evaluation relying on multiple informants using multiple methods will yield data that are contradictory and inconsistent, making interpretation difficult. Nevertheless, this comprehensive strategy is recommended to reduce the influence of specific sources of error and to produce the opportunity to recognize and interpret convergent themes used to generate explanations and interventions.

## **3. EVALUATING OUTCOMES**

### **3.1. Purpose**

As noted in the previous section, children and adolescents manifest many of their problems related to behavioral and emotional problems at school. This not only warrants school-based evaluation of these problems, it frequently results in school-based provision of services. These services take many forms, ranging from universal interventions, to disorder-specific treatment, to school-based day treatment programs. Professionals providing these services are expected to assess the progress of children and adolescents participating in them so as to modify interventions and know when interventions are no longer needed. As a result, the questions that guide outcome evaluations pertain to understanding an individual's progress in treatment with implications for modifying or discontinuing services. These assessment

procedures frequently include observations, rating scales, and performance measures.

### **3.2. Observations**

Observations have inherent advantages and disadvantages that are described in the previous section. The use of observation measures to measure outcomes is warranted with high-frequency behavior or with behavior that has a high degree of situational specificity. Observed rates of on-task behavior is a common outcome measure of a high-frequency behavior. Defiant and disruptive behavior is usually difficult to evaluate with observational techniques because its frequency is low even for students who teachers perceive as defiant or disruptive. This is especially true in secondary schools, where episodes of defiant behavior can be quite intimidating and need not occur very often for a teacher to report a problem. Low-frequency behavior with a high degree of situational specificity is conducive to observational measures because the observer can efficiently target the specific situation that elicits the problematic behavior. For example, a child referred for problematic behavior in the cafeteria may warrant observational assessment because the observation period is limited and the likelihood of observing the problematic behavior within this situation may be quite high. Observational measures have been found to be sensitive to intervention effects when administered frequently and reliably.

### **3.3. Rating Scales**

The use of rating scales can be a very efficient method for collecting outcome data. Compromises to validity can occur when raters are invested in the intervention process, resulting in a bias to perceive improvement. Collecting data from some raters who are blind to treatment or not invested in its success can be a valuable assessment strategy; however, sometimes these people are not available or are not able to provide useful assessment data. Collecting rating scale data from multiple sources and then interpreting convergent findings has been recommended so as to minimize individual biases. This technique can be helpful, but it also presents evaluators with unique challenges. It is quite common to receive rating scale data with tremendous variability and little convergence around any conclusion. Although this may indicate poor reliability, it may also indicate that the child's behavior changes as a function of classroom, activity, or teacher. In these situations, it is sometimes helpful to evaluate improvement in relation to

specific situations and settings. In other words, a child may be making considerable improvement in Mr. Smith's math class but demonstrating no improvement in Mrs. Martin's English class.

One of the problems with the use of rating scales as measures of outcomes pertains to the instruments possessing the necessary sensitivity to detect meaningful change. Changes resulting from effective interventions might not appear in changes in scores on rating scales because the targeted behavior change may have been more specific than can be indicated on most rating scales. For example, interventions may effectively reduce a set of behavior that annoys peers in math class; however, this may produce very little change in the child's overall social functioning or acceptance by other children and adults. As a result, the use of most rating scales as an outcome measure in this situation is probably inappropriate.

To measure changes in behavior, it is frequently useful to have repeated measurement of target behavior so as to track change. One of the practical problems with administering rating scales is the timely completion and return of assessments by parents and teachers. It is usually necessary to closely monitor their return, check for completion (e.g., whether the rater completed the back of the form), and follow-up with raters whose forms are incomplete or late.

### 3.4. Performance Measures

Performance measures assess the ability of a child or an adolescent to exhibit a skill. They are frequently used in academic achievement testing, but they also have applications related to behavioral and emotional problems. Pre- and post-assessments can be administered to assess whether a child had a skill (e.g., social skill, problem-solving skill) prior to the intervention and possesses the skill after the intervention. These assessments tend to be straightforward and based on an appraisal of the child's ability to demonstrate through role-play, written responses, and/or verbal description a set of skills that have been taught directly. These techniques are useful for determining whether a child or an adolescent is ready to begin applying the skills in actual situations because generalization is not likely to occur if the child cannot produce the behavior on cue. A common misapplication of these measures in evaluating outcomes is an overreliance on them as the end point of the need for interventions. Many socially impaired youth have been successfully taught numerous social skills that have resulted in no change in

behavior in other settings. Generalization, or the degree to which learned skills are displayed in novel or untrained situations, is usually the most challenging part of any behavioral intervention, and measures assessing behavior change in the target settings is a necessary outcome measure.

### 3.5. Current Issues

Successfully evaluating outcomes requires giving attention to many issues, including generalization, mediators and moderators, and size of effects. Attention to generalization requires evaluators to focus key outcome measures on the settings in which the problems that led to the initial evaluation occurred. Although measurement of the process of interventions can help to inform treatment decisions, it is possible to understand the ultimate success of interventions only when assessments target the problems in the context in which they have occurred. Moderators and mediators refer to characteristics of the individuals receiving interventions and characteristics of the intervention process that affect treatment outcomes. For example, interventions provided by different teachers in different classrooms may result in mixed results, not as a result of the intervention but rather as a result of whether or not each teacher implemented the techniques correctly. In this example, it could be said that the implementation of the technique "mediated" the outcomes.

Although data indicating improvements are encouraging, questions will remain as to how much improvement is necessary before it is concluded that meaningful changes have taken place. This issue is at the heart of indexes of effect size and clinical significance. These calculations rely on data from the same assessment techniques as described previously but include analyses to assist in the interpretations of the size of the effects. There are various approaches to this topic, including measures of whether change moved a student from outside of the normal range to within the normal range as well as indexes of the size of change in relation to the standard error of measurement and standard deviation. Unlike many other analyses, these techniques can be applied to data for an individual child or a group of children.

## 4. LEGAL AND ETHICAL ISSUES

For states to receive federal funding, there must be strict adherence to the principles of the applicable

program legislation and civil rights legislation passed at the federal level. State laws are often more specific than federal legislation and act to flesh out federal mandates, such as IDEA '97, by defining the means that school professionals use to meet the requirements outlined in these laws. With states operating independently in this fashion, there are often differences among states in terms of how the federal laws are interpreted. One state's procedures for conducting an FBA or a BIP may differ from those of another state. Furthermore, requirements set forth in IDEA that are not addressed by the state may be interpreted and defined at the school district level, sometimes resulting in dramatic differences in assessment and classification procedures between adjacent neighborhoods. Therefore, the onus is placed on schools to train their staff to be able to conduct behavioral assessment in accordance with the pertinent laws and local procedures.

In terms of behavioral assessment, perhaps the two most relevant areas of legal and ethical concern are informed consent and procedures for ensuring cultural competence in assessment. As with any form of non-academic assessment in the school setting, legal and ethical issues arise around consent for assessment. IDEA and the American Psychological Association's Ethical Principles of Psychologists specifically require that informed consent be obtained from the parents or guardians of a minor before conducting an initial evaluation. For consent to be "informed," the signor must (a) fully understand the purposes and means of the assessment, (b) provide consent voluntarily, and (c) be legally competent. It is generally assumed that children are incompetent to provide independent consent, although there are situations where exceptions are made. For example, students who are emancipated from their parents may be eligible to give consent on their own behalf, and many states allow independent consent for mental health services under 18 years of age. In general, however, parents or guardians are typically required to consent to any assessment services provided to children.

When obtaining consent, professionals must keep in mind that the school environment creates a unique situation for behavioral assessment because, in contrast to clinical settings, it is not automatically assumed that a child's participation will include such services. The clinical practice of "passive consent," or the assumption of consent based merely on participation, is not applicable in the school setting. In fact, some argue that the school setting is inappropriate for mental health services in general and

that the practice of assessing children in a school in domains outside of academia is inappropriate. Therefore, professionals in the school setting have the legal and ethical duty to ensure that parents or guardians understand the purpose and means of any assessment that is conducted. It is strongly recommended that school professionals obtain "active consent," which includes formal documentation of parental consent and meets all of the requirements of consent described previously. Ultimately, this might mean that students will not receive the services they need due to parental disapproval or apathy. Although this is unfortunate, the need for active consent presents such an ethical, and at times legal, concern that it is suggested that professionals in the schools always seek active consent despite the possible consequences.

Another area of legal and ethical concern is cultural competence. In simplest terms, there are two components of cultural competency. First, professionals must understand and respect the values, beliefs, customs, and traditions of the families they serve. This component of cultural competence involves interacting appropriately with people from diverse cultures, implementing a communication process whereby practitioners interact with people as equals (and not as experts), and demonstrating a genuine commitment to understanding students' strengths and weaknesses in the context of the culture, resources, and environment in which the behaviors occur. As mentioned previously, when a professional uses assessment instruments that are normed on non-representational samples, there is a need to interpret the results with great care. Cultural competence in this respect may involve the ability to interpret the results within the correct framework. For example, a child of Hispanic background who appears to be significantly "introverted" or to lack "leadership" skills on some rating scales may in fact exhibit behaviors well within ranges that would be expected of his or her family-oriented cooperative culture. It is the culturally competent professional who can interpret these results in the correct context.

The second component to cultural competence is an awareness of how professionals' own cultural backgrounds affect their services to the students. In some cases, whole schools can adopt a white middle-class culture as a result of the dominant cultural makeup of the faculty and staff. In these instances, there may be a tendency within the school to place an emphasis on timeliness, parental involvement (to the exclusion of extended family members), and a lack of emotional display. Although this might not be self-evident

among the faculty and staff members of the school, these cultural norms can be at odds with families of diverse cultural backgrounds. For example, school personnel can mistakenly interpret students who arrive late for an assessment, or parents who arrive late to meetings, as uncaring or uninvolved. However, in a family's culture, there might not be the same emphasis on timeliness. Similarly, students who openly weep at unfortunate news or seem to be unmoved by positive feedback may be mistakenly considered to be unstable or unconcerned. Again, these reactions might not fit with the dominant culture of the school or professionals themselves, but they still fall well within the norm for the student's culture. As a result, professionals must continually assess their own cultural influences and be willing to recognize instances where their own cultural values do not match those of the students and their families.

### See Also the Following Articles

Attention Deficit/Hyperactivity Disorders (ADHD) ■ Autism and Asperger's Syndrome ■ Cross-Cultural Psychology, Overview ■ Cultural Complexity ■ Cultural Psychology ■ Educational and Child Assessment ■ Emotional and Behavioral Problems, Students with ■ Ethics and Social Responsibility ■ Forensic Mental Health Assessment ■ Learning Disabilities ■ Psychological Assessment, Standards and Guidelines for ■ Psychology and the Law, Overview ■ School Discipline and Behavior Management ■ School Violence Prevention ■ Suicide Intervention in Schools

### Further Reading

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/Adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232.
- Dunlap, G., Kern-Dunlap, L., Clarke, S., & Robbins, F. R. (1991). Functional assessment, curricular revision, and severe behavior problems. *Journal of Applied Behavior Analysis*, *24*, 387–397.
- Ervin, R. A., Radford, P. M., Bertsch, K., Piper, A. L., Ehrhardt, K. E., & Poling, K. (2001). A descriptive analysis and critique of the empirical literature on school-based functional assessment. *School Psychology Review*, *30*, 193–210.
- Ghuman, H. S., Weist, M. D., & Sarles, R. M. (Eds.). (2002). *Providing mental health services to youth where they are*. New York: Brunner–Routledge.
- Gresham, F. M., Watson, T. S., & Skinner, C. H. (2001). Functional behavioral assessment: Principles, procedures, and future directions. *School Psychology Review*, *30*, 156–172.
- Mash, E. J., & Terdal, L. G. (Eds.). (1997). *Assessment of childhood disorders* (3rd ed.). New York: Guilford.
- Pellegrini, A. D. (2001). Practitioner review: The role of direct observation in the assessment of young children. *Journal of Child Psychology and Psychiatry*, *42*, 861–869.
- Shapiro, E. S., & Kratochwill, T. R. (Eds.). (2000). *Behavioral assessment in schools: Theory, research, and clinical foundation* (2nd ed.). New York: Guilford.
- Sterling-Turner, H. E., Robinson, S. L., & Wilczynski, S. M. (2001). Functional assessment of distracting and disruptive behaviors in the school setting. *School Psychology Review*, *30*, 211–226.
- Weist, M., Evans, S. W., & Lever, N. (Eds.). (2003). *Handbook of school mental health: Advancing practice and research*. New York: Kluwer/Plenum.



# Behavioral Medicine Issues in Late Life—Pain, Sleep, etc.

Bruce Rybarczyk

Rush Medical College, Chicago, Illinois, United States

1. Applying Behavioral Medicine to Older Adults
  2. Depression and Health in Older Adults
  3. Behavioral Sleep Medicine and Aging
  4. Chronic Pain Interventions for Older Adults
  5. Behavioral Medicine Issues in the Frail Old
  6. Summary
- Further Reading

**pseudodementia** Reversible symptoms of dementia (i.e., memory, attention, and other cognitive deficits) that are caused by an underlying depression.

**sleep hygiene** An educational treatment for insomnia that provides instruction on how the sleep environment, caffeine, alcohol, nicotine, food, and exercise affect sleep.

**sleep restriction** A treatment for insomnia that involves restricting the time allowed for sleep, beginning with a very short period and gradually increasing. This method promotes a state of sleep deprivation in order to facilitate learning to fall asleep faster and stay asleep.

**stimulus control** A treatment for insomnia that requires the individual to leave the bed when he or she is not able to sleep and to not use the bed or bedroom for any activities during the day.

## GLOSSARY

**depressed psychophysiologic functioning** A complication from hospitalization (increased incontinence, confusion, agitation, and loss of appetite) often observed in the frail old.

**excess disability** A substantially lower than expected level of physical functioning for a given medical condition, attributable to behavioral factors (i.e., noncompliance, lack of activity, and depression).

**frail elderly** The old-old (age 85+ years) segment of the population with significant, multiple medical illnesses that limit functioning. These individuals typically do not live independently.

**masked depression** A presentation of depression in older adults characterized by the presence of more physical symptoms (i.e., reduced appetite, sleep disturbance, fatigue, and increased aches and pains) and fewer mood symptoms.

Behavioral medicine encompasses the study of how behavior impacts the onset, progression, and recovery from illness and the application of those behavioral factors to promoting health and preventing illness. Due to national demographic trends and the prevalence of chronic diseases in late life, older adults are emerging as a critical target group for behavioral medicine interventions. In the older adult age group, behavioral factors relating to comorbid depression, excess disability, pain, and sleep are of particular importance for behavioral medicine.

## **1. APPLYING BEHAVIORAL MEDICINE TO OLDER ADULTS**

Due to the twin factors of increased life expectancy and reduced birth rates, most nations throughout the world are experiencing a dramatic shift toward an older adult population. It is projected that by 2030, 70 million Americans (20% of the U.S. population) will be 65 years or older, compared to approximately 35 million (14%) in 2000. This aging revolution has led to increased public and scientific interest in how behavioral and psychological factors influence the aging process. In addition, with the great strides made in the successful treatment of acute illness in the past century, people are living longer but at the same time are accruing more age-related chronic illnesses. In response, the focus of the U.S. health care system is shifting toward managing chronic illness and disability, with an emphasis on maximizing independence and quality of life in late adulthood. The goal is to help people live well rather than simply live longer.

By necessity, the field of behavioral medicine will play an increasingly important role in the health care system of the future. Behavioral medicine encompasses the study of how behavior impacts the onset, progression, and recovery from illness and the application of those behavioral factors to promoting health and preventing illness.

Older adults are in many ways the ideal age group to apply behavioral medicine principles and interventions. First, they are at much greater risk for chronic medical conditions, such as heart disease, lung disease, diabetes, cancer, and arthritis, and they often have multiple chronic conditions. Interventions aimed at preventing the onset or progression of such conditions are most cost-effective when they are designed for and delivered to this age group. In-home exercise interventions designed specifically for older adults, for example, have been shown to yield substantial benefits to patients with cardiovascular disease, partly due to the fact that older adults are more vulnerable to diminished cardiovascular capacity when they do not engage in systematic exercise.

Second, each chronic medical condition typically requires several different medications, multiplying the risk for side effects (e.g., memory impairment) and drug interactions. Accordingly, behavioral interventions that can manage symptoms without the use of medications are particularly important for this population.

Last, and perhaps most important, is the fact that physical and psychological conditions become more

interdependent and reactive to each other with advancing age. Therefore, interventions aimed at psychological conditions, such as depression and anxiety, as well as the enhancement of psychosocial coping resources can have a significant impact on the physical health of older adults. Mind-body group interventions that teach a range of skills (e.g., relaxation training, problem solving, and cognitive approaches to managing stress) have been shown to be effective in enhancing coping and quality of life in older adults with chronic illness.

## **2. DEPRESSION AND HEALTH IN OLDER ADULTS**

The greater interplay between psyche and soma (body) in older adults is perhaps most evident in research on depression and health. When older adults have a depression overlaid on a medical condition, they are at much higher risk for excess disability. Excess disability is defined as a substantially lower than expected level of physical functioning for a given medical condition, attributable to behavioral factors (i.e., noncompliance and lack of activity). After a stroke or hip fracture, for example, when depression is detected an aggressive psychological intervention is required to prevent poor physical rehabilitation outcomes.

Even mortality has been shown to be affected by depression in older adults. In the Cardiovascular Health Study, depressive symptoms were related to 6-year mortality in a large sample of adults older than age 65, after controlling for medical and demographic variables known to affect mortality. The hypothesized mechanism by which depression affected mortality was decreased motivation to initiate and sustain cardiac health behaviors.

Unfortunately, medical professionals often fail to detect the presence of depression in older adults. This failure in detection is partly accounted for by the atypical presentation of depression in many older adults. One such presentation is as masked depression, which is characterized by the presence of more physical symptoms (i.e., reduced appetite, sleep disturbance, fatigue, and increased aches and pains) and fewer mood symptoms. Furthermore, depressed older adults are far less likely to have a negative view of themselves (e.g., "I am a failure" and "I am worthless"), which is a hallmark of depression in other stages in life. Another presentation, called pseudodementia, is characterized by reversible symptoms of dementia

(i.e., memory, attention, and other cognitive deficits) that are caused by an underlying depression.

Fortunately, cognitive-behavioral and other psychotherapy treatment approaches have been found to be as effective as medication in the treatment of depression; moreover, older adults are as likely to benefit from these treatments as younger adults. Furthermore, there is no evidence to suggest that when chronic illness coexists with depression that psychotherapy treatment for depression is any less effective, although the situation is likely to contribute to the range of cognitive errors and distortions that need to be addressed in cognitive-behavioral therapy (Table 1).

A positive development is a growing trend toward improved screening and treatment of depression in older medical patients as well as a greater presence of mental health professionals in primary care settings. This trend is particularly important given the fact that older Caucasian males are at the highest risk of any group for suicide and as many as 70% of suicide victims visit their primary care physicians prior to committing suicide. Tragically, these older male patients were unable to communicate their depression to their physicians and/or their physicians failed to ask the appropriate questions. Once depression is detected, psychotherapy that is provided on-site in a primary care office is often more acceptable to older adults. This detection treatment model highlights the importance of integrating mental health and medical treatment for older adults in particular.

TABLE 1

**Common Cognitive Errors and Distortions in Depressed Chronically Ill Older Adults**

---

Irrational belief: disability = inevitable misery	“Anyone would be depressed in my shoes.”
All-or-none thinking	“If I can’t do everything that I used to do, in the ways I was accustomed to, I might as well not do anything . . . because it won’t be satisfying.”
Negative filtering	“I’m nothing but a burden to my family.”
Negative forecasting	“If I continue to have all these medical problems, my friends won’t want to be around me anymore.”
Mind reading	“The medical staff is getting tired of working with me.”

---

### 3. BEHAVIORAL SLEEP MEDICINE AND AGING

Sleep has emerged as an important area in behavioral medicine. Treatment of insomnia, in particular, has been an important area of research and intervention because it is frequently caused or made worse by behavioral factors that are amenable to cognitive-behavioral intervention. In this case, again, the older adult population may be the ideal target group for these interventions.

There are numerous sleep disorders that increase in frequency with age, including sleep apnea, restless legs syndrome, and periodic limb movement disorder, but chronic insomnia is the most prevalent in late life. The risk for developing chronic insomnia doubles between the ages of 55 and 85 and is present in 20–30% of the population older than age 65. It usually manifests itself as a combination of different problems, including difficulty falling asleep, excessive time spent in bed awake, more frequent and longer awakenings after sleep onset, and fewer total hours of sleep. Furthermore, there is an increasing recognition that sleep disturbance at night leads to daytime fatigue and/or sleepiness, which is associated with decreased quality of life in older adults.

During the past 30 years, several different behavioral interventions have been developed and shown to be highly effective. In fact, it has been noted that behavioral interventions for insomnia are probably the most effective and durable interventions in the behavioral medicine repertoire, with typical success rates after only six to eight sessions of treatment of 60% or greater. In addition, most of these treatment effects have been measured and sustained over periods of as long as 2 years.

The most effective cognitive-behavioral interventions tested to date include all or most of the following components: sleep hygiene instruction, relaxation training, sleep restriction, stimulus control, and cognitive modification. Sleep hygiene involves basic education on how the sleep environment, caffeine, alcohol, nicotine, food, and exercise affect sleep. Several different types of relaxation training (e.g., progressive muscle relaxation, deep breathing, and visual imagery) have been provided under the premise that mental and physical arousal that accompanies stress and anxiety will inhibit and disrupt sleep. If patients can train themselves to relax before sleep or at night after an awakening, they are more capable of falling asleep and staying asleep.

### 3.1. Sleep Restriction

Sleep restriction is the most difficult to adhere to but is probably the most essential component of treatment of insomnia in older adults. Its aim is to counter the sleep fragmentation and daytime sleeping that is commonly observed in older adults by restricting the time allowed to sleep to a specific “sleep window.” This sleep window is initially set for a very short period of time, determined by the average amount of time that the individual slept per night (based on a sleep diary that is kept throughout treatment). For many individuals, this period is 5 or 6 h long, so they are required to be awake for as many as 19 h each day prior to going to bed. During the first few weeks the individual experiences a high level of sleep deprivation, compounded by the fact that he or she is not yet capable of sleeping all the time during the sleep window. This prolonged sleep deprivation creates “sleep pressure,” which makes it easier for patients to train themselves to sleep without interruption during the sleep window.

After each week of sleep restriction treatment, the patient’s sleep efficiency (percentage of time slept during the sleep window) is calculated and 30 min is added to the sleep window if the patient achieves a sleep efficiency of 85% or greater. The eventual goal is to give the patient enough time in bed to meet his or her particular sleep needs (i.e., enough sleep to feel refreshed and not sleepy during the day) while maintaining a high level of sleep efficiency.

### 3.2. Stimulus Control

The stimulus control method is designed to increase the association between being in bed and being asleep. Predicated on operant learning theory, it involves instructing individuals to greatly limit their activities in the bed and bedroom. Basically, the only activities allowed in bed are sleep and sex. They are not permitted to do such activities in bed as watching television, reading, eating, worrying, listening to the radio, working, or simply resting. In addition, they are not allowed to be awake in bed for longer than 15–20 min at any time during the night, whether they are trying to fall asleep or trying to get back to sleep after awakening in the middle of the night or early morning. If they cannot fall asleep or back to sleep after 15–20 min, they are to get out of bed and engage in a nonstimulating activity until they become sleepy enough to try again. If they are unable to fall asleep again within 15–20 min, they are to repeat the same procedure. In addition, to facilitate the functioning of the body’s

circadian clock system, they are instructed to use an alarm clock and get up at the same time every morning.

The final component of treatment is cognitive modification. Similar to cognitive restructuring for anxiety or depression, this approach emphasizes changing unrealistic beliefs and irrational fears involving sleep or sleep loss. Research shows that older adults with insomnia endorse stronger beliefs about the negative consequences of insomnia, express more fear of losing control of their sleep, and express more helplessness about its unpredictability compared to older adults without insomnia.

Dysfunctional cognitions about sleep create a vicious cycle by increasing performance anxiety about sleep, which in turn leads to arousal and greater difficulty falling asleep. Common misconceptions about sleep include the following: blaming daytime impairments, such as fatigue and inability to concentrate, exclusively on sleep loss; the fear that insomnia will worsen a medical condition; the belief that not getting 7 or 8 h a sleep per night will threaten one’s health; and the belief that sleeping late the next morning or napping is a good way to compensate for poor sleep at night. Finally, the belief that a bad night of sleep results in a “lost day” leads to cancellation or avoidance of activities, such as socializing and physical activities, that are important to maintaining health and promoting sleepiness at night.

Initially, the focus of research on behavioral treatment for insomnia was on younger adults and later it moved to older adults who had no other medical conditions. Until recently, individuals with chronic illness were excluded from these studies under the presumption that insomnia in these individuals was likely caused by a medical condition or a medication and therefore needed to be treated by directly managing the medical condition. However, it is now understood that insomnia in virtually all cases, regardless of coexisting medical conditions, is made worse by behavioral factors that are responsive to behavioral treatment. Indeed, two studies have demonstrated that older adults with multiple medical conditions are as responsive to treatment as individuals without medical conditions.

Insomnia in older adults occurs mostly in those with age-related chronic diseases, such as cardiac disease, arthritis, and lung disease. When insomnia coexists with chronic disease, these patients have significantly reduced quality of life, experience more pain, use more medical services, and are more likely to develop depression. Thus, providing behavioral treatment to ameliorate insomnia in these individuals has numerous benefits that go beyond improvement of sleep, leading to improved daytime mood, alertness, and physical functioning.

#### 4. CHRONIC PAIN INTERVENTIONS FOR OLDER ADULTS

Another important application of behavioral medicine to the older adult population is the management of chronic pain. Pain is a constant factor in the lives of a great many older adults. Due to the greater prevalence of conditions that cause pain (e.g., arthritis and diabetes), older adults have more chronic pain and are more likely to have multiple sites of pain than younger adults. Pain in older adults has been strongly linked to excess disability, insomnia, depression, and increased health care utilization. Older adults with a combination of pain and depression have been shown to be particularly vulnerable to physical disability.

The two types of behavioral treatments for chronic pain that have been applied to older adults are cognitive-behavioral therapy (CBT) and operant conditioning. CBT methods focus on modifying beliefs, thoughts, attitudes, and skills that affect different aspects of the pain experience. They also employ relaxation training to reduce the sympathetic nervous system component of the pain experience. Operant conditioning focuses on modifying observable behaviors, such as activity level, medication usage, and pain signaling (i.e., grimacing, limping, and complaining). This method uses operant conditioning principles to reward activity, change environmental contingencies, and shift medication usage from an as-needed to a time-contingent approach. Both treatment methods use learning techniques such as consistent self-monitoring, graded practice, and homework. CBT and operant techniques are compatible and are often used together in clinical practice.

Contrary to traditional bias among professionals that older adults may not benefit from CBT interventions due to the introspection that is involved, studies demonstrate that older adults obtain equal benefits from CBT pain interventions compared to younger adults. In addition, they are no less likely to drop out of treatment. Older adults have also been shown to be as capable of learning relaxation procedures as younger adults. Nonetheless, research exploring age bias in inpatient and outpatient pain treatment programs has found that rehabilitation professionals remain less confident that older adults will benefit from their program. This bias is confirmed by the much smaller than expected percentage of older adults enrolled in these programs, which means they are being underserved by these invaluable treatments.

Older adults may also have some unique, age-related strengths and resources that help them cope better

with pain. The majority of studies suggest that older adults have a higher pain threshold compared to younger adults. In addition, older adults may be more tolerant of pain due to the fact that they are more willing to accept certain levels of pain and discomfort as a part of growing older. This is consistent with studies showing that older adults report diminished emotional responses to pain, such as depression, anxiety, anger, or fear. Similarly, according to the life span theory of control, as individuals move into late adulthood there is a shift from a focus on primary control (control over the external world) toward an emphasis on secondary control (control of the internal mental and emotional life). This may equip older adults with a greater skill to cope with stressors that are irreversible or cannot be easily controlled, such as chronic pain.

#### 5. BEHAVIORAL MEDICINE ISSUES IN THE FRAIL OLD

Among the old-old (age 85+ years) with significant medical illnesses, sometimes called the frail old, behavioral medicine has perhaps an even greater role to play due to the even stronger interplay between psyche and soma. When the frail old develop new medical problems that compound existing ones, they often experience the onset of new psychological symptoms, such as delirium, depression, or paranoia. Similarly, older adults residing in nursing homes have been shown to experience increased mortality and medical illness, not accounted for by medical factors, when they are required to relocate from one institutional setting to another. Even “timing of death” research has shown that, remarkably, the old-old are more likely to die of natural causes after, rather than before, a holiday or important family event. This suggests that the psyche exerts some degree of control over the timing of death.

When frail older individuals are hospitalized they often experience depressed psychophysiological functioning in the form of increased incontinence, confusion, agitation, and loss of appetite. These complications lead to further medical interventions, such as catheterization for incontinence, the administration of sedating medications or use of restraints for agitation, or tube feeding in the case of failure to eat. These treatments, in turn, lead to iatrogenic effects (i.e., complications from treatment) such as bladder infections (from catheterization), blood clots (from being restrained to the bed), and pneumonia (from tube feeding).

## 6. SUMMARY

Due to the increased risk of chronic illness and the increased interaction between psyche and soma in late life, behavioral medicine interventions that target older adults are among the most cost-effective and essential. Interventions that reduce anxiety and depression or enhance coping in older adults with chronic diseases have much potential to yield crossover effects on medical variables, such as functional status, symptom severity, and cognitive functioning. These factors combined with the aging revolution in the United States will lead to an increased focus within behavioral medicine on providing services to older adults.

### See Also the Following Articles

Aging, Cognition, and Medication Adherence ■ Depression in Late Life ■ Elder Caregiving ■ End of Life Issues

## Further Reading

- Rybarczyk, B., Gallagher-Thompson, D., Rodman, J., Zeiss, A., Gantz, F., & Yesavage, J. (1992). Applying cognitive-behavioral psychotherapy to the chronically ill elderly: Treatment issues and case illustration. *International Psychogeriatrics*, 4(1), 127–140.
- Schulz, R., Beach, S. R., Ives, D. G., Martire, L. M., Ariyo, A. A., & Kop, W. J. (2002). Association between depression and mortality in older adults: The Cardiovascular Health Study. *Archives of Internal Medicine*, 160(12), 1761–1768.
- Stepanski, E., Rybarczyk, B., Lopez, M., & Stevens, S. (2003). Assessment and treatment of sleep disorders in older adults: An overview for rehabilitation psychologists. *Rehabilitation Psychology*, 48, 23–36.
- Yonan, C. A., & Wegener, S. T. (2003). Assessment and management of pain in the older adult. *Rehabilitation Psychology*, 48, 4–13.



# Behavioral Observation in Schools

John M. Hintze

University of Massachusetts, Amherst, Massachusetts, USA

1. Introduction
  2. General Methods of Behavioral Observation
  3. Measuring and Recording Behavior Systematically
  4. Coding Schemes
  5. Observational Instruments
  6. Summarizing Behavioral Observation Data
  7. Reliability and Validity of Behavioral Observation
  8. Summary
- Further Reading

## GLOSSARY

**frequency or event recording** Involves counting the number of times a behavior is observed during a specified time period.

**interobserver agreement** Technique in which the consistency of recording across two or more observers is compared to assess the reliability of measurement.

**interval recording** Divides the observational session into a number of equal intervals and simply records the presence or absence of specified behaviors during each interval.

**narrative recording** Refers to the observation and collection of information on student behavior in naturally occurring arrangements with little constraints placed on the observer for how and/or what to observe and record.

**partial-interval recording** An interval recording technique in which the interval is scored if the target behavior is present during any part of the interval.

**systematic direct observation** Refers to the observation of behavior that is explicitly defined under predetermined settings.

**whole-interval recording** An interval recording technique in which the interval is scored if the target behavior is present throughout the entire interval.

Behavioral observation refers to a method of assessment whereby human observers objectively record the ongoing behavior of a person or persons in specific environmental circumstances.

## 1. INTRODUCTION

Direct observation is one of the most widely used individual assessment procedures by school-based professionals. In a survey of more than 1000 school-based professionals, of the 26 different types of individual assessment instruments listed across seven different assessment categories (e.g., aptitude, social-emotional, and personality), behavioral observation methods ranked highest in terms of frequency of use. Overall, respondents indicated that on average they conduct approximately 15 behavioral observations during the course of a typical month.

## 2. GENERAL METHODS OF BEHAVIORAL OBSERVATION

When conducting behavioral observations, school-based professionals generally rely on one of two basic approaches to gathering information. One form, narrative recording, refers to the observation and collection of information on student behavior in naturally

occurring arrangements with little constraints placed on the observer for how and/or what to observe and record. With narrative recording, the observer generally describes the observed situation making notes of specific behaviors as they occur. The method(s) of observation, the type of information noted, and the manner in which the information is summarized are left to the discretion of the observer. The strength of such procedures lies in the flexibility that the observer has in choosing when and how to observe and the minimization of obtrusiveness or reactivity that may occur as a result of the presence of the observer. For interpretative purposes, narrative recording is often used to help develop hypotheses about the various behavioral and environmental factors that may be worthy of further observation and analysis. This flexibility, however, is also one of the main weaknesses of narrative recording. Because the observer has great autonomy in choosing how and what to observe, judgments about the worth of such reports are inextricably bound to the subjective judgments of the observer. As such, it would be highly unlikely for two independent observations collected at the same time to appear identical in the information provided. Therefore, narrative recording is generally best used as a precursor to more specific, objective accounts of behavior.

In contrast to narrative recording, systematic direct observation refers to the observation of behavior that is explicitly defined under predetermined settings. Although this approach is also concerned with observing behavior under naturally occurring environmental contexts, the aim is to define beforehand the behaviors of interest, choose specific recording strategies, and have observers record whenever behavior corresponding to the predefined operational definitions occurs. In particular, systematic direct observation is distinguished by five characteristics: (i) The goal of observation is to record specific behaviors, (ii) the behaviors being observed have been operationally defined a priori in a precise manner, (iii) observations are conducted using standardized procedures and are highly objective in nature, (iv) the times and settings for observation are carefully selected and specified, and (v) scoring and summarizing of data are standardized and do not vary from one observer to another. In comparison to narrative recording, a main objective of systematic direct observation is to have independent observers agree on what is observed and recorded, assuming that they have observed the same stream of behavior.

### **3. MEASURING AND RECORDING BEHAVIOR SYSTEMATICALLY**

Various types of data can be collected during systematic direct observation. A workable definition of a target behavior is one that provides an accurate description of the behavior and clearly defines the boundaries of its existence and nonexistence. Therefore, constructs and reifications do not lend themselves well to direct forms of observation. For example, raising one's hand to be called on is an observable and measurable behavior. Behaving "off the wall" is not something that can be directly observed (although operational definitions of the behaviors that constitute "off the wall" could be developed). Once behavior is defined, the calibration of the operational definition is determined by the nature of the data—the frequency of its occurrence and the particular interests of the observer. In addition, practical considerations, such as the availability of observers, the amount of time a student is accessible, or any combination of these factors, all dictate the type of data collected.

### **4. CODING SCHEMES**

When observing behavior systematically, observers generally use any one of a number or a combination of approaches to recording behavior. The more common approaches usually involve counting the number of times a behavior occurs during a specified time period and/or noting the presence or absence of behavior at specific time intervals of an observational session. Frequency or event recording involves counting the number of times a behavior is observed during a specified time period. When the time periods vary across multiple observational sessions, frequencies are converted to rate of behavior per unit time. For example, an observer may report that a child got out of his or her seat at an average rate of one time per minute during three separate observations conducted over the course of 3 days, even though the actual duration of each observation period varied. By using rate of behavior rather than frequency, comparisons can be made across observational sessions that differ with respect to time. Frequency recording is most useful for observing behaviors that have a discrete beginning and ending and are relatively consistent in the length of time that they take to occur. For example, "tantruming" may not lend itself well to frequency recording because the beginning and end of each tantrum might be difficult

to discern and the length of time of each tantrum might vary from a few minutes to hours. Moreover, frequency recording is usually better suited for behaviors that occur at lower rather than higher rates. With higher rates of behavior, accurately detecting each instance of behavior can be challenging.

In addition to frequency recording, another commonly used recording schedule is interval recording. In comparison to frequency recording, which notes each occurrence of behavior, interval recording divides the observational session into a number of equal intervals and simply records the presence or absence of specified behaviors during each interval. For example, a 20 min observation session could be broken down into 120 10 s intervals. Moreover, unlike frequency recording, in which each instance of behavior is noted, interval recording is only concerned with whether the targeted behavior occurs during the interval. As such, distinctions are not made regarding how many behaviors were observed during the interval. For example, an interval in which there were nine instances of out-of-seat behavior would be coded the same as an interval in which only one instance of out-of-seat behavior occurred.

In practice, interval-recording data are generally collected using one of three coding techniques. When using whole-interval recording, an interval is scored when the target behavior was present throughout the entire interval. Since the behavior must be present for the entire interval, whole-interval recording is well suited to behaviors that are continuous or intervals of short duration. One of the drawbacks of whole-interval recording is that it tends to underestimate the presence of the behavior. For example, if “off-task” behavior were the target behavior of interest and it was observed to have occurred for 8 s of a 10 s interval, the interval would not be scored for the presence of off-task since it did not occur for the entire 10 s. In contrast to whole-interval recording, partial-interval recording codes the occurrence of a behavior if it occurs during any part of the interval. Partial-interval recording is a good choice for behaviors that occur at a relatively low rate or for behaviors of somewhat inconsistent duration. However, because the interval is scored for any presence of the target behavior, partial-interval recording tends to overestimate the actual occurrence of behavior. Lastly, momentary time sampling codes the presence or absence of the target behavior at only one predefined instant of the interval (e.g., usually the instant in which the interval begins). As such, the target behavior is coded as occurring or not on the basis of a fractional observation of the total interval. Once the

target behavior is coded, no other behaviors are noted for the balance of the interval. Although momentary time sampling appears counterintuitive because it is based on the smallest sample of behavior, it actually provides the least biased estimate of behavior, generally providing neither over- nor underestimates of actual behavior. In addition to frequency and interval recording, other less commonly used coding methods include duration (i.e., the actual amount of time that a behavior occurs), latency (i.e., the time it takes for a behavior to be initiated following a prompt or directive), and intensity (i.e., the amplitude of behavior).

## 5. OBSERVATIONAL INSTRUMENTS

In addition to observer-designed coding schemes, observational instruments have been developed to assess a specific range of behaviors in specific environmental circumstances. For example, an observer might choose to use an observational instrument designed specifically to quantify the percentage of time a student is academically engaged in or on-task, the frequency with which the student interacts with other students on the playground, or the number of times a teacher provides directives, opportunities to respond, or positive reinforcement. In comparison to observer-developed coding schemes, observational instruments can be somewhat more limited in their flexibility since they are usually developed with specific standardized operational definitions and recording schedules. On the other hand, since observational instruments are usually developed with a specific purpose in mind, they often provide a more detailed account of the student’s behavior under the specific environmental context of interest.

## 6. SUMMARIZING BEHAVIORAL OBSERVATION DATA

Once behavioral observation data are coded they are usually summarized across the entire observational period to provide a description of behavior. As previously indicated, frequency data are reported as a function of rate (i.e., number of behaviors noted divided by the amount of time observed). By doing so, summary data collected across observational sessions of varying time can be compared. Interval data are expressed as a percentage of the intervals that the target behavior occurred compared to the total number of intervals observed. In summarizing

and reporting interval recording results, it is important to note that the behavior was coded for a certain percentage of intervals observed and not time (e.g., “Off-task behavior was noted for 60% of the intervals observed,” not “Off-task behavior was noted 60% of the time”). This subtle distinction highlights that the coding scheme used was interval recording, and that intervals and not actual time was the unit of analysis. Duration recording is generally reported as the cumulative amount of time that a behavior was observed. However, it is not uncommon to record the duration of each instance of observed behavior separately and then sum the durations for a final cumulative total. In addition to providing total duration, the latter procedure can also provide the average duration per instance of behavior. Latency recording is summarized similarly to duration, with either cumulative latency or average latency per behavior generally noted. Lastly, intensity is usually summarized using some form of subjective ordinal scale of analysis (e.g., spoke in a voice loud enough to be heard by a person sitting next to him or her, spoke in a voice loud enough to be heard by a person in the room, spoke in a voice loud enough to be heard by a person outside the room, etc.) or by some mechanical recording device specifically designed to measure the dimension of interest (e.g., amplimeter).

## **7. RELIABILITY AND VALIDITY OF BEHAVIORAL OBSERVATION**

The applicability of reliability and validity to behavioral observation has been equivocal, with some suggesting that classical psychometric concepts based on differences between persons are irrelevant to an assessment methodology that focuses on behavior and its variation within individuals, and others suggesting that the differences between traditional and behavioral assessment are primarily conceptual, not methodological, and as such reliability and validity considerations apply.

Traditionally, behavioral assessors and researchers have approached the issue of reliability using accuracy and interobserver agreement as substitutes. Specifically, accuracy refers to the extent to which observed values of behavior approximate the “true” state of the behavior as it actually occurs. When the true state of the behavior is known, comparing the observed values to the true values derives accuracy of measurement. In this case, the behavioral notion of accuracy is the same as reliability in the classic measurement case in which reliability is represented by the extent to which an observed

score represents the true score. As such, in order for a measurement system to be accurate it must be sensitive to the occurrence of true behavior, its repeated occurrence, and its occurrence in multiple settings.

The problem with accuracy, however, is obtaining measures of behavior that can legitimately be considered true values. In particular, when systematic direct observation by humans are used to collect data, developing an incontrovertible index against which to compare scores generated from such an observation system can prove difficult. When possible, some form of technological reproduction is typically used to generate the true values of the behavior. For example, sample sessions of behavior may be videotaped and studied carefully in order to determine the true values for the target behavior. The ability to replay the tape as many times as necessary or to use stop action and reverse options helps avoid error, although it can never be proven that error is completely absent. In this manner, observed behavioral values can be compared to the true value template and accuracy assessed.

The problems in assessing accuracy are obvious. First, the practicality and feasibility of obtaining true values or an incontrovertible index of behavior are low. Simply, most behavioral assessors and many researchers are not equipped with the type of technological assistance that is required to capture the true essence of the target behavior. Second, the issue that one can never fully know the true value behavior because it is impossible to separate what we know from how we know it is ever present. Therefore, accuracy is highly time and context dependent and only represents point estimates of reliable measurement. For these reasons, accuracy is rarely considered or reported in the behavioral assessment literature.

In addition to accuracy, a second commonly used proxy for reliability is that of interobserver agreement. Here, concurrent observations are made independently by two observers and the degree of association between the observers is assessed. It is important to note, however, that interobserver agreement provides no information with respect to accuracy and reliability. Reliable or accurate observations are those that demonstrate a consistent relationship to the behavior as it actually occurs; interobserver agreement provides no such information. Simply, the fact that two observers agree on the number of times a target behavior occurred says nothing about whether their observations were accurate or reliable. For example, two observers may both note the occurrence of 10 instances of the target behavior. The level of their agreement would appear to be 100%. However, if 20 instances of the behavior actually

occurred, and observer A witnessed only the odd-numbered occurrences and observer B noted only the even-numbered occurrences, their actual agreement would be 0%. As such, although complete agreement between two observers might seem like a comforting piece of information, the only conclusion that can be made is that their total values (not the actual behaviors) for the session were in agreement.

As can be seen, both interobserver agreement and accuracy involve comparing an observer's data to some other source. However, they differ considerably in the extent to which the source of comparisons can be entrusted to reflect the true behavior as it actually occurred. Accuracy, which makes a direct comparison to an incontrovertible index of the real behavior, is obviously the more reliable (and valid) of the two; however, it is difficult to estimate in applied research settings. Interobserver agreement, although easier to acquire and often used interchangeably with reliability, suffers from an inability to demonstrate accuracy and makes obvious the problems with using consensus as a replacement for reliability.

## 8. SUMMARY

Behavioral observation is one of the most widely used assessment strategies in schools. Given its flexibility and ease of use, behavioral observation procedures can be used to collect a range of data that provide helpful information and are useful for making a variety of psychoeducational decisions. Because of its direct nature, behavioral observation is particularly well suited for

everyday life settings and can provide a systematic record of behavior that can be used in preliminary evaluation, intervention planning and design, the documentation of changes over time, and as part of a multi-method-multisource evaluation that integrates other forms of assessment (e.g., interviews and rating scales) and sources (e.g., teachers, parents, and children).

## See Also the Following Articles

Effective Classroom Instruction ■ Emotional and Behavioral Problems, Students with ■ School Discipline and Behavior Management ■ School Environments

## Further Reading

- Alessi, G. (1988). Direct observation methods for emotional/behavioral problems. In E. S. Shapiro, & T. R. Kratochwill (Eds.), *Behavior assessment in schools: Conceptual foundations and practical applications* (pp. 14–75). New York: Guilford.
- Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology—IV* (Vol. 2, pp. 993–1006). Washington, DC: National Association of School Psychologists.
- Merrell, K. W. (1999). *Behavioral, social, and emotional assessment of children & adolescents*. Mahwah, NJ: Erlbaum.
- Skinner, C. H., Dittmer, K. I., & Howell, L. A. (2000). Direct observation in school settings: Theoretical issues. In E. S. Shapiro, & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations* (2nd ed., pp. 19–45). New York: Guilford.



# Boredom

William K. Balzer, Patricia C. Smith and Jennifer L. Burnfield

Bowling Green State University, Bowling Green, Ohio, USA

1. Introduction
  2. Definition of Boredom
  3. Models of Boredom
  4. Measuring Boredom
  5. Causes and Correlates of Boredom
  6. Consequences of Boredom at Work
  7. Prevention and Alleviation of Boredom
  8. Conclusion
- Further Reading

## GLOSSARY

**affect circumplex** A two-dimensional model of affect, with orthogonal dimensions of pleasure–displeasure and high–low activation (arousal).

**flow** Experiences in which people are highly engaged and focused on what they are doing. When in a flow state, one can complete tasks well and with ease.

**monotony** Perceived repetitiveness or tedium of a task.

**output curve** Plot of the rate of output produced during a specified time interval (e.g., a workday).

**vigilance** The ability to maintain attention and respond to stimuli over an extended period of time.

Boredom is an unpleasant, short-lived, affective state in which a person feels a pervasive lack of interest in, and difficulty attending to, a task or activity. Despite the absence of a single model or theory to explain boredom at work, research has made some progress on

identifying its causes and a wide range of interventions to prevent or alleviate its occurrence.

## 1. INTRODUCTION

Interest in boredom at work can be traced to the early 1900s. In 1913, Munsterberg, the father of applied psychology, identified boredom as a valuable area of inquiry in the workplace. The first scientific study, “The Physical and Mental Effects of Monotony in Modern Industry,” was conducted more than a decade later by Davies in 1926. Since these early studies, the research on boredom has been sporadic and fragmented, resulting in a nascent understanding of a construct that is thought to have significant performance outcomes and personal consequences for employers and employees. Smith’s 1981 article succinctly summarizes the then—and current—state of research on boredom at work: “The amount of research devoted to the topic of boredom by psychologists and psychiatrists is astonishingly small when compared to literary treatments and to the acknowledged importance of [boredom at work]” (p. 338). Although the focus of this article is on boredom at work, it is important to recognize that other areas of applied psychology have also investigated the causes and consequences of boredom, including education, occupational health, ergonomics, military psychology, mental health, and interpersonal relationships.

In fact, the recent research on boredom has not been conducted in the domain of work. Much of the recent work has been in the areas of clinical psychology and personality.

## 2. DEFINITION OF BOREDOM

A single clear definition of boredom has been elusive. Boredom has often been used interchangeably with other related but not identical constructs (monotony, repetitiveness, ennui, and tedium) or its observed or hypothesized phenomenological correlates (fatigue, irritation, depression, dissatisfaction, hopelessness, and stress). Others have approached boredom by focusing on methods of alleviating boredom: novelty, interest, complexity, curiosity, enrichment, and “flow” (i.e., experiences in which people are highly engaged, focused, and able to do well). Progress has been made toward a single definition of boredom, complicated by the fact that differing models or theories of boredom (e.g., psychological, psychophysiological, and motivational) necessarily affect its definition. Commonalities across definitions lead to the working definition of boredom used here, as proposed by Fisher in 1993: “Boredom is an unpleasant, transient, affective state in which an individual feels a pervasive lack of interest in and difficulty concentrating on the current activity.” By definition, boredom is viewed as a negative reaction in response to a personal assessment of a work situation that is more short-lived than attitudes such as job satisfaction. The relation of boredom to other affective constructs in a larger nomological net has been acknowledged. In the framework provided by the “affect circumplex,” “boredom” is located on the negative ends of the orthogonal dimensions of pleasantness and activation, along with “tired” and “drowsy.”

## 3. MODELS OF BOREDOM

Despite 90 years of investigation, there is no single, agreed upon model of boredom at work. Smith’s 1981 research provides a good historical overview of the development of boredom at work, including some models of the causes and consequences of boredom: conflict models, psychoanalytic models, signal detection models, cognitive models, and drive models. Four models serve here as exemplars of the existing approaches to the understanding of boredom at work:

the psychodynamic model, the cultural normative model, the psychophysical model, and the cognitive–affect model.

### 3.1. Psychodynamic Model

Sigmund Freud’s psychodynamic approach to the understanding of underlying psychological forces has been applied to explain boredom at work. Boredom at work is thought to reflect an inability to identify and ascribe emotional meaning to one’s work. In this view, workers avoid or deny finding meaning in their work because they expect that individual work is not meant or designed to provide personal meaning. The meaninglessness of their work is experienced as boredom, which may be hidden from awareness and understanding through repression, denial, and defense mechanisms, resulting in tension (between what a person wants and what a person gets from work) that manifests itself in psychosomatic complaints, pathology, and so forth.

### 3.2. Cultural Normative Model

This model proposes that prior socialization makes employees highly dependent on authority figures to ascribe meaning to their work. Workplace norms develop that lead workers to accept without question boredom at work. In this workplace, boredom is denied and suppressed. Individual workers are not free to share or discuss their perceptions or emotions due to these normative pressures. Instead, they are relegated to chronically boring and meaningless jobs.

### 3.3. Psychophysical Model

The psychophysical model of boredom at work holds that the human nervous system is an active searcher, rather than a passive recipient, of environmental stimulation. Boredom results from exposure to monotonous or repetitious sensory stimulation. Habituation or reduced arousal to the task will result unless the worker voluntarily exerts additional attention and effort to the task. This increased overall arousal offsets the declining sensory stimulation provided by the monotonous task while maintaining a level of arousal adequate to satisfactorily perform the task. Overall, a psychophysical model suggests that if individuals cannot compensate for inadequate arousal from a task by increasing their level of effort, boredom will result.

### 3.4. Cognitive–Affect Model

The cognitive–affect model proposed by Hill and Perkins in 1985 posits that boredom is a psychological (as opposed to psychophysical) phenomenon. If the sensory stimulation from work fails to meet the needs or expectations of the individual employee, and the employee is unable to supplement the task through additional or alternative stimulation, frustration (negative affect) will result, leading to the overall experience of boredom. In this model, it is the subjective interpretation of the cognitive components of the task that is critical, rather than the actual sensory stimulation. If the task is perceived as monotonous and unable to meet the instrumental needs of the employee engaged in the task, it will lead to an affective disturbance and the experience of boredom.

## 4. MEASURING BOREDOM

Early assessments of boredom in the workplace were based primarily on the observation of workers, unstructured self-reports, or the inference of boredom based on patterns of responses to related items on structured surveys. For instance, early studies used subjective perceptions of output as a criterion of boredom. In this method, employees indicate the time of the day in which they think they work better and have higher production rates while also indicating when they felt bored. Studies of production workers found higher levels of boredom in the morning and attributed the increased effort and interest as the day progressed to the impending goal of reaching the end of the workday. This pattern of findings supports Wyatt's 1929 rich metaphor for boredom at work—a long hike:

When activities provide little or no interest, the prospects are depressing and discouraging, and the effect seems to bear some relation to the magnitude of the task to be accomplished. To a person who dislikes walking, the prospects of a 20-mile tramp over uninteresting country may be almost painful, and although he starts with a certain amount of forced effort and enthusiasm, this soon disappears and leaves him in a darkened mood. At this stage the magnitude of the task may be appalling, but as the time for lunch and rest draws near interest is awakened and depression recedes. The distance still to be traversed now shrinks considerably, and bears little resemblance to the endless trail conceived a few hours before. A new swing

and liveliness characterize (sic) his step as he again takes to the road, and although these may diminish after a few more miles, they reappear with increased intensity as he enters the lap which separates him from his goal. (p. 162)

Several researchers sought an objective index of boredom in the curve of piece-rate output. This curve is a plot of the total number of pieces completed for some fixed period of time (e.g., 1 h intervals) throughout the day. Work output showed U-shaped curves with a sag in the middle of the morning and again in the middle of the afternoon. The declines in performance were attributed to slower rates of performance when machine operators were bored, with higher rates at the beginning of work and in anticipation of the end of the workday. However, when researchers actually observed and timed workers, the drop in production proved to represent voluntary rest pauses, including rest room and smoking breaks, which may be needed or scheduled at predictable times throughout the workday. Output rate was not slower when the operators were actually working at their machines. Overall, although the hiking metaphor may explain what is happening psychologically, observations of workers that experimentally control for employees who leave their workstations (e.g., a smoking break) do not necessarily support this assumed pattern of motivation and performance.

Following the era in which boredom in the workplace was measured through unstructured employee self-reports of boredom or objective measures of work output over time, a number of self-report rating instruments have been developed. For example, Lee's 1986 Job Boredom Scale measures boredom through employee responses to 17 items concerning satisfaction, interest, and connectedness to the job. Measures also exist that assess workers' susceptibility or proneness to experience boredom. The most common measures are Farmer and Sundberg's 1986 Boredom Proneness Scale and Zuckerman, Eysenck, and Eysenck's 1978 Sensation Seeking Scale (e.g., measuring aversion to repetition and dull people). Unfortunately, there is no single source reference that provides descriptions and psychometric evaluations (e.g., does the instrument measure boredom at work in a consistent and conceptually appropriate manner?) of available measures. College and university research libraries and Internet search engines may be the best sources for finding well-constructed measures for predicting or assessing boredom at work.

## 5. CAUSES AND CORRELATES OF BOREDOM

Despite the absence of a single unified model or theory to explain boredom at work, the state of research has progressed on identifying its causes. The hypothesized causes of boredom can be grouped into three general categories: the task, the work environment, and the person.

### 5.1. Task as a Cause of Boredom

The traditional view of boredom at work assumes that boredom resides in the task itself. Specifically, tasks that provide low stimulation (i.e., are repetitive and monotonous) have been thought to result in boredom. Fisher's 1993 article offers a conceptual framework of work tasks that may lead to boredom: quantitative underload (jobs that provide the worker with nothing to do or jobs with a high degree of workload variability, e.g., periods of inactivity after high levels of activity), qualitative underload (jobs that are simple, repetitive, limited in mental demand, unchallenging, or do not use the individual's skills), and qualitative overload (jobs that are too challenging, difficult, or incomprehensible).

### 5.2. Work Environment as a Cause of Boredom

The work environment may be seen as a lens through which feelings of boredom are reduced or intensified. The mere presence of coworkers can enhance arousal directly through conversation or interdependent work. However, dull, unhappy, or uncommunicative colleagues may actually contribute to feelings of boredom. From a social influence perspective, fellow employees may magnify or reduce boredom through statements or actions that make work seem challenging and interesting or boring and repetitive. Organizational policies and practices may also contribute to the presence or absence of boredom. Rules that prohibit talking, limit work breaks, or prescribe precise work procedure may be so constraining as to induce feelings of boredom.

### 5.3. Person as a Cause of Boredom

Individual differences among employees may also contribute directly to boredom at work. Research suggests that males tend to be more prone to boredom than females,

particularly when there is a perceived lack of stimulation. However, this gender finding may be culture-specific. Individuals with high cognitive ability or capacity may be more likely to experience boredom on tasks that underutilize their abilities. Conversely, the more intelligent worker may perceive work differently and find a higher degree of variety in a seemingly homogeneous and repetitive task. Personality differences may also contribute to feelings of boredom. Extraverted people may require more stimulation to keep them aroused and engaged at work. The expectations and alternatives available to the person may also affect the perception of interest versus boredom. If a job is the only available job without the hope of other prospects, it may be perceived as less boring. The psychiatric literature has also identified pathologically bored individuals who are unable to perceive the stimulation inherent in work activities or are unable to focus their attention at work as do their better adjusted colleagues.

It is also probable that the task, environment, or person characteristics that cause feelings of boredom interact to accentuate or attenuate these feelings. For example, an employee with limited cognitive capacity performing a challenging and demanding task (i.e., qualitative overload) may experience strong feelings of boredom, whereas an employee with a greater intellectual capacity would not. Conversely, the high-capacity employee may feel more bored on a simpler task (i.e., qualitative underload) than would a person of more limited capacity. As another example, employees who do not perceive any value or instrumentality in their participation at work are likely to feel more bored regardless of whether their work tasks are repetitive or challenging. Finally, individuals who have more complex approaches to the world (i.e., high in schema complexity) may be able to perceive more nuances and details in their work, making them more engaging and less boring. They also may be more likely to engage in goal setting, enhancing the meaning and motivation from their work. Overall, the causes of boredom are likely to be a combination of the direct and interactive impact of task, environment, and person characteristics.

## 6. CONSEQUENCES OF BOREDOM AT WORK

Boredom at work has negative consequences for workers, their organizations, and society as a whole. Research

suggests that the predicted or demonstrated effects of boredom for individual employees include work stress, job dissatisfaction, physical and mental health problems, workplace hostility, increased risk taking or thrill seeking, horseplay and sabotage, alcohol and drug problems, absenteeism, work-related injuries among adolescent employees, and poor performance. There is also evidence that boredom leads to the perception that time passes more slowly, although individuals' objective estimates of time passed seemed to be unaffected by high levels of boredom. The consequences of boredom for businesses and organizations include higher employment costs (due to employee turnover, absences, and health care costs), performance problems (work quality and work quantity), and reduced organizational effectiveness (including profitability). Finally, boredom at work may also have consequences for society as a whole, including lost productivity, reduced quality of work life, and reduced consumer safety. Numerous studies of vigilance—the ability to maintain attention and respond to stimuli over an extended period of time—reveal that an observer's ability to detect critical signals declines over time. This may be exhibited by a reduction in accuracy or increased response time. High levels of boredom experienced in vigilance tasks have been associated with stress and potentially more serious outcomes. Especially disconcerting is the case of a commercial airliner that overflew its destination by more than 100 miles before the sleeping flight crew was awakened by air traffic controllers. Clearly, boredom at work has significant consequences.

## **7. PREVENTION AND ALLEVIATION OF BOREDOM**

Repetitive assembly, continuous manual control, and inspection and monitoring tasks may be most likely to lead to boredom at work. A wide range of interventions have been suggested to prevent or alleviate boredom at work. One would suspect that the intervention would be most effective if it corresponded to the cause of boredom (e.g., if boredom is caused by the task, the intervention should also focus on changing the task or the perception of the task). Consistent with the framework for categorizing the causes of boredom at work, interventions to prevent or alleviate workplace boredom are therefore organized around the task, the work environment, and the person.

### **7.1. Task Interventions to Prevent and/or Alleviate Boredom at Work**

Although some work tasks may be inherently boring, interventions have been identified that can help minimize boredom on even these tasks. One category of intervention is to design work that is stimulating and psychologically rewarding rather than monotonous and fractionated. Morgeson and Campion described a number of conceptual models for redesigning work to reduce the effects of stressors such as workload and repetitive work. Such interventions can be focused on the motivational aspect of the task (e.g., job enrichment to enhance task meaningfulness) or a reduction in exposure to the stressor that induces boredom or its effects (e.g., ergonomic changes in the work environment to reduce the effects of repetitive work).

In addition, for repetitive tasks, grouping output into recognizable “lots” or batches and providing a system for tracking the number of lots completed may be a useful intervention. Job rotation, or changing work assignments throughout the day, may help employees stay alert and challenged at work. Finally, the use of performance goals (either self-set goals or the acceptance of company-set goals) and feedback on goal success can help prevent or alleviate boredom at work.

### **7.2. Work Environment Interventions to Prevent and/or Alleviate Boredom at Work**

Employee selection and placement systems are an underused means for preventing boredom. Matching individuals to particular jobs may be particularly effective given differences in individual perceptions and reactions to different work tasks. The use of performance-based reward systems, including piece rate payment systems, may also create an incentive that increases employee effort and may prevent or reduce boredom for those who are prone to boredom. However, workers with a low proneness to boredom may benefit from jobs and work environments that emphasize intrinsic rewards such as job enrichment. Companies can also provide rest periods, which have been found to neutralize the unpleasant and unfavorable effects of repeated activity. Providing work that is employee paced rather than externally paced (e.g., a conveyor system) can allow the employees either individually or as a group to adapt their workload to cyclical levels of attention and effort throughout the day. Communicating the importance of the employee's work

contribution to the overall effectiveness of the work unit or company can increase interest in and attention to work. Finally, designing the workplace so that it allows for conversations and social interactions among employees can help prevent or alleviate boredom at work.

### 7.3. Person Interventions to Prevent and/or Alleviate Boredom at Work

Employees can take steps to reduce boredom at work. Fisher's 1998 article demonstrated that brief interruptions by both internal (e.g., mind wandering) and external (e.g., mobile phones and e-mail) sources may alleviate boredom, depending on the level of attention required to perform the task. In terms of internal mechanisms, forcing oneself to refocus attention or effort on the task can help minimize boredom, particularly at the outset of unstimulating tasks or in constraining work environments. Employees can also seek additional stimulation in an unstimulating work environment by playing mental games, assisting coworkers, and so forth. For well-learned repetitive tasks, employees might use daydreams to escape a tedious and monotonous task. Individual employees may also cope with boredom by learning to integrate their rhythm with that of machines, allowing them to accomplish their tasks through more automatic attentional processes. External interruptions have also been found to reduce boredom on simple low-attention tasks, perhaps by providing variety and additional stimulation to a monotonous work environment. However, such interruptions may be problematic in other situations or for other outcomes.

## 8. CONCLUSION

Given the individual, organizational, and societal consequences of boredom at work, it is surprising that more attention and effort have not been devoted to theory development, psychological measurement, and scientific investigation. The available research, spanning 90 years of work by practitioners and scientists, has provided a good foundation for expanding our knowledge and understanding of boredom at work. How will current and future changes in the workplace affect boredom at work? The growing use of robots and computers to complete repetitive and monotonous tasks could potentially reduce boredom at work. The migration of simple manufacturing and assembly jobs to other countries may, on average, reduce the level of boredom in the U.S. workplace—if

replacement jobs offer more stimulating tasks or work environments. Changes in the workplace that enhance workers' control over their jobs (i.e., meaningful assignments, goal setting, and feedback) and meet their social needs at work may reduce the level of boredom for future employees. Additional research is needed to ensure the prevention and treatment of boredom in today's workplace and the changing workplace of the future.

### See Also the Following Articles

Competence at Work ■ Cooperation at Work ■ Work Adjustment ■ Work Environments ■ Work Motivation ■ Work Role, Values Sought in the ■ Work Teams

### Further Reading

- Davies, A. H. (1926). The physical and mental effect of monotony in modern industry. *British Medical Journal*, 2, 472–479.
- Farmer, R., & Sundberg, N. D. (1986). Boredom proneness: The development and correlates of a new scale. *Journal of Personality Assessment*, 50(1), 4–17.
- Fisher, C. D. (1993). Boredom at work: A neglected concept. *Human Relations*, 46(3), 395–417.
- Fisher, C. D. (1998). Effects of external and internal disruptions on boredom at work: Two studies. *Journal of Organizational Behavior*, 19, 503–522.
- Gemmill, G., & Oakley, J. (1992). The meaning of boredom in organizational life. *Group & Organization Management*, 17(4), 358–369.
- Hill, A. B., & Perkins, R. E. (1985). Towards a model of boredom. *British Journal of Psychology*, 76, 235–240.
- Kass, S. J., Vodanovich, S. J., & Callender, A. (2001). State-trait boredom: Relationship to absenteeism, tenure, and job satisfaction. *Journal of Business and Psychology*, 16, 317–327.
- Lee, T. W. (1986). Toward the development and validation of a measure of job boredom. *Manhattan College Journal of Business*, 15, 22–28.
- O'Hanlon, J. F. (1981). Boredom: Practical consequences and a theory. *Acta Psychologica*, 49, 53–82.
- Morgeson, F. P., & Campion, M. A. (2003). Work design. *Handbook of Psychology: Industrial and Organizational Psychology*, 12, 423–452.
- Smith, R. P. (1981). Boredom: A review. *Human Factors*, 23(3), 329–340.
- Sommers, J., & Vodanovich, S. J. (2000). Boredom proneness: Its relationship to psychological- and physical-health symptoms. *Journal of Clinical Psychology*, 56, 149–155.
- Wyatt, S. (1929). Boredom in industry. *Personnel Journal*, 8(3), 161–171.
- Zuckerman, M., Eysenck, S., & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology*, 46, 139–149.



# Bullying and Abuse on School Campuses

Michael J. Furlong, Alicia C. Soliz, Jenne M. Simental, and Jennifer L. Greif

University of California, Santa Barbara, California, USA

1. Introduction
  2. What Is Bullying?
  3. How Common Is Bullying on School Campuses?
  4. Who Are the Bullies?
  5. Who Are the Victims?
  6. What Are the Effects?
  7. What Can Be Done?
  8. The Special Role of Bystanders
  9. Implications
- Further Reading

Most adults can remember bullies and victims of bullying from when they were in school. The impact of these events does not stop when students graduate; these events can have long-lasting effects on bullies, victims, and even students who only watched bullying occur. The term bullying refers to a specific type of peer aggression that is intentional, repeated, and involves an imbalance of power between the bully and the victim. Suggestions for intervention approaches at the individual, classroom, and schoolwide level are provided in this article.

## GLOSSARY

**bullying** Aggression that is intentional and repeated over time, and there is an imbalance of power between the perpetrator and the victim. Bullying is a specific type of peer victimization that is categorized by these components, which indicate an ongoing relationship between the bully and the victim.

**direct bullying** When a bully is overtly aggressive toward a victim (e.g., repeatedly hits, teases, or steals from the victim).

**indirect bullying** When a bully expresses aggression toward a victim through a third party (e.g., spreads rumors and gossips). Although the bully and the victim do not actually interact with each other in instances of indirect bullying, this aggression still has an impact on the well-being of the victim.

## 1. INTRODUCTION

Bullying in school settings was once thought of as a normal, transient part of growing up. It was considered to have little lasting impact, and the best advice to the victim was to fight back or turn the other cheek. Beginning in the late 1970s and through the 1980s, pioneering research began to shed light on this behavior and to dispel common myths. Expanded studies during the 1990s started to examine bullying as a serious form of aggression and considered the possible impacts of this experience on the bully, his or her victim, and even bystanders. The results of these investigations have shown that not only is bullying a common occurrence but also it has lasting effects on all involved.

The pioneer work on bullying was conducted in the mid-1980s by Dan Olweus, a Norwegian psychologist commissioned to conduct a large-scale study on the topic. Norway's interest in bullying was driven by a string of suicides committed by adolescent boys who had been victims of severe bullying. These efforts were followed by a number of other investigations conducted by researchers in England (Smith), Canada (Peppler), and Australia (Slee, Rigby, and Griffiths). Detailed studies in the United States did not begin to focus seriously on bullying until later, primarily in response to several deadly school shootings that were perpetrated by students who were thought, at least in part, to be seeking revenge for past abuse by school bullies.

## **2. WHAT IS BULLYING?**

Bullying is defined as aggression between peers that has three essential elements: It is intentional, it is repeated over time, and there is an imbalance of power between the bully and the victim. This definition (developed by Olweus) distinguishes bullying from other types of peer aggression, such as one-time fights or friendly teasing between friends. It is these three defining characteristics that make bullying potentially so damaging. Victims are intimidated or hurt repeatedly by someone who is more powerful; consequently, they may learn to accept the victim role and come to distrust others. Ongoing victimization can have a serious impact on the development and well-being of children who are bullied. In addition, bullies who abuse their power (e.g., physical strength and popularity) to single out and intimidate peers potentially become involved in a pattern of using this social control strategy throughout their lives. As such, bullying is considered an exploitative relationship between peers rather than a one-time event.

Despite relative consensus among researchers as to what constitutes bullying, this clarity does not always extend into practice. Many of the most commonly used measures of bullying do not specify that, by definition, bullying is a series of intentional, repeated acts of aggression that occur in peer relationships in which there is an imbalance of power. The omission of a precise definition in research is problematic because, without these specifications, studies may overreport the prevalence of bullying when students are asked about general aggression. In addition, when educators and others think of bullying as a specific event, rather than a series of encounters, interventions may be more punitive and less focused on changing the process of interactions between peers.

In the past, bullying was considered to be physical acts of aggression that primarily occurred between boys. It is now recognized that bullying can take many forms, including ongoing physical abuse (e.g., hitting and pushing), verbal abuse (e.g., taunting and name-calling), social manipulation (e.g., rumor spreading and purposeful exclusion from activities), and attacks on property (e.g., breaking or stealing belongings). One common distinction is between direct bullying (i.e., physical aggression or chronic teasing that occur overtly between the bully and the victim) and indirect bullying (i.e., gossip or rumors targeting the victim that occur between the bully and a third party). Recognition that bullying can include these additional behaviors has broadened the definition and it is now more inclusive of types of aggression that occurs more frequently between girls (e.g., gossip and exclusion). Regardless of its specific form, all bullying involves the repeated abuse of a student by another more powerful student who wants to harm the victim. The bully, in fact, comes to enjoy exercising power over his or her selected target. It is a myth that bullies abuse others to compensate for low self-esteem. In fact, they often have positive self-esteem, may be popular with their peers, and inappropriately come to enjoy having power over others.

## **3. HOW COMMON IS BULLYING ON SCHOOL CAMPUSES?**

The incidence of bullying varies from country to country and school to school within countries. However, it is certain that bullying occurs at all comprehensive schools across all grade levels. Students consistently report that most bullying incidents take place away from an adult presence at the school, including on school playgrounds, during or between class, and walking to and from school. In addition, bullying is often unreported by the parties involved, primarily because the bullies wish to avoid punishment and the victims wish to avoid further aggravating the bully or prompting retaliation—bullies intimidate their victims. Current research practices for collecting information utilize a combination of teacher and/or parent reports as well as student self-reports and student identification or nomination. Because researchers use various methods, a definitive estimate of the prevalence of bullying in schools is difficult to obtain. It is also probable that the available statistics underestimate the extent and scope of this problem.

A 1998 survey of 15,686 students in the United States conducted by the World Health Organization found that 16% of boys and 11% of girls felt bullied, with 23% of boys and 11% of girls admitting to bullying other students. Research has also found an association between the frequency of short-term and long-term victimization. For example, in a 2003 study of Norwegian victims of bullying, Solberg and Olweus found that bullying that persisted over a period of 1 month or more was more frequently reported than short-term bullying events. In other words, bullying is often a chronic experience for its victims.

There are some developmental differences in bullying. A 1993 study by Whitney and Smith found that the occurrence of bullying decreases as students get older, with the highest rates occurring in the upper elementary and middle schools and the lowest rates in secondary schools. This is suspected to be due, in part, to increased social and physical maturity. By the time a child reaches secondary school, there is a decreased presence of older, more powerful peers who can single out and pick on smaller, weaker peers.

In addition to age differences, there are also gender differences in bullying trends. As mentioned previously, boys are more likely to report both bullying and victimization; however, girls are more likely than boys to experience indirect bullying as opposed to direct bullying. Although the consequences of social rejection and isolation associated with indirect bullying may not be immediately obvious, it may produce negative long-term effects on the victim's social development. Finally, female victims are more likely to tell a female peer about their experience, whereas males are more likely to turn to a trusted adult.

#### **4. WHO ARE THE BULLIES?**

Despite the increased awareness of female bullying, several researchers have found that most bullies tend to be boys or groups of boys. The results of studies designed to identify the etiology of bullying behavior have been mixed. Some experts suggest that bullies act to exert power over others with the primary purpose of gaining more power. Alternatively, other experts draw upon a social learning model in which bullying is a reaction to being bullied by other children or a reaction to a challenging home environment with associated authoritarian discipline experiences.

In cases in which bullies appear to be the popular children with no history of victimization, the bully behavior may be somewhat puzzling. In these cases, experts have suggested that bullying is a reaction to

some underlying need for power. Bullies are rewarded for exerting power over others by gaining greater power, which reinforces this negative behavior.

Alternatively, the social learning perspective suggests that children who have suffered bullying may begin to bully others less powerful than themselves either as a form of retaliation or as a way of gaining some sense of empowerment. These children have been categorized as bully/victims. Similar to these bully/victims, children from abusive family environments may experience bullying at the hands of family members and, in turn, become bullies themselves. For example, Roberts and Morotti suggest that bullying behavior stems from abusive, chaotic home environments in which families bully the child, who in turn bullies a peer (i.e., a "kick-the-dog" phenomenon). In this type of environment a child learns to normalize and eventually utilize negative social interactions, such as bullying. This idea is supported by the identification of common characteristics of children known to suffer from abusive family experiences. For example, children who bully often display behaviors indicating low social competence, antisocial behavior, a lack of empathy, and high levels of aggression (physical, verbal, and relational). In addition, when interviewed, bullies often describe home environments that are more negative and authoritarian in style than those of children without histories of bullying.

#### **5. WHO ARE THE VICTIMS?**

Identified victims of bullying share many similar characteristics. In one pattern, victims are classified as either passive (never behaving aggressively) or highly aggressive (often provoking peers). Most victims are children who spend more time alone than typical children, often falling into the socially rejected category when rated by their peers. They often have poor or less than desirable physical characteristics (e.g., smaller in height, weight, and strength), are younger than the majority of their peers, display low social competence, and emotional maladjustment including depression. In general, victims of bullying are those youth whose physical, psychological, or social status is diminished compared to that of school bullies, thus making them possible targets of repeated aggression.

#### **6. WHAT ARE THE EFFECTS?**

The commonly cited short-term effects of bullying to the victim include risks of diminished self-esteem and

depressed mood. Several physical symptoms, such as stomachaches and headaches, difficulty concentrating, and disruption of sleeping and eating habits, have been present among victims of bullying. Long-term effects can include depression, social anxiety, increased levels of aggression, lowered self-esteem, increased risk of suicide, decreased coping skills, and the presence of posttraumatic stress disorder symptoms. As a result of being traumatized by incidents of bullying, victims may carry these negative consequences into adulthood, leading to decreased levels of societal functioning.

Less well known are the effects of bullying on the bully himself or herself. It appears that the long-term effects include risks of school failure; further antisocial, criminal, and delinquent behavior; and an increase in depression, suicidal ideation, and global negative self-evaluations. In 1993, Olweus found that of the Norwegian children identified as bullies in grades 6–9, 60% were convicted of at least one crime by the age of 24. Surprisingly, of this 60%, 35–40% had three or more convictions. In addition, incidents of drug abuse in conjunction with depression and suicidal tendencies have often been high among individuals identified as bullies, especially in comparison to individuals with no history of bullying. Finally, bullying is a behavior that, once engaged in, becomes a pattern that is difficult to stop. The bully may become accustomed to and rewarded by getting what he or she wants through asserting power over other individuals. This behavior may continue throughout life and affect social interactions and interpersonal relationships. In short, someone who is abusive in one aspect of his or her life tends to be abusive in various other aspects of life, which can lead to an aggressive and typically frustrating pattern of social behavior.

## **7. WHAT CAN BE DONE?**

Bully intervention programs implemented at schools seek to eliminate existing bully problems, prevent the development of new bully problems, and achieve better peer relations at school. Other goals of bully intervention programs are to create a positive school climate and increase caring behaviors toward bully victims. These goals are accomplished at the schoolwide level, the classroom level, and the individual level.

### **7.1. Schoolwide Responses**

Before implementing a bully prevention program, the following must be considered:

1. Staff training: What type of staff training is required? Who will provide training? How will training occur?
2. Program funding: How much funding is available? How will funds be allocated?
3. Time commitments: How much time will be required by the trained individuals outside of normal duties? How often will the program be implemented and for how long?
4. Clear behavioral objectives: What will be expected of students after the intervention? Will the school rules be changed?
5. Alternative programs: Which programs will be available for students who are not affected by the intervention?
6. Program leadership: Who will take the lead? How will staff be chosen to participate? What qualifications and training are necessary?

In order to equip schools with appropriate tools to deal with bullying, training should include an awareness component through which the school staff is informed about the extent of bullying on campus. Staff should also be asked to examine their own beliefs about bullying. Training on bully prevention and intervention should include specific strategies as well as an overarching conceptual understanding of the nature of bullying as described in previous sections.

Because bullying often happens in unsupervised contexts, such as on the playground and in the lunchroom, where paraprofessionals and community volunteers often provide supervision, bus drivers, maintenance staff, and cafeteria workers should also receive training. Support staff should be aware of the school code and the school's general bully response procedures.

Bully intervention programs at the school level often restructure the school environment to remove the consequences and negative impacts of bullying. Comprehensive programs typically create specific school-site committees that coordinate aspects of the program, administer questionnaires to determine the nature and extent of the bully problem, hold meetings to raise community awareness, improve supervision and the outdoor environment, and involve parents.

### **7.2. Classroom-Level Responses**

As part of an overall schoolwide plan, teachers implement many bully prevention programs within their classrooms. These programs ideally include a set of clear classroom rules and routines (preferably shared

across classrooms at the same grade level), comprehensive academic lesson preparation, ample classroom structure, opportunities for students to develop empathy for one another, a warm classroom environment, cooperative instructional groups, and shared responsibilities. In addition to basic preparation and creating a positive classroom environment, teachers should never overlook bullying incidents because in doing so they may inadvertently contribute to a broader “culture of bullying” on the school campus. It is important to create conditions that encourage both victims and bystanders to report bullying when it occurs. It is also important to implement programs that address the unique needs and provide assistance to all involved—bullies, victims, and bystanders. Many teachers also implement programs to develop student skills for resolving conflict in order to provide opportunities for “in-the-moment” social skill learning.

### 7.3. Individual-Level Responses

It is important to determine the function that bullying serves in specific incidents because this will influence the intervention approach. Most bullies know that what they are doing is wrong, but they often have limited empathy for the victim and may have a personal need to dominate others. Merely punishing the bully (by using power over him or her) may only reinforce the notion that power is an effective social tool. Immediately following a bullying incident, a teacher or staff member should have a serious talk with the bully to determine what steps should be taken next. The talk should include a documentation of the incident, a message that bullying is not acceptable, a referral to the school’s code of conduct, and a reminder that the bully’s behavior will be closely monitored in the future. Individual responses to bullying also include continued one-on-one discussions, counseling, parent involvement, follow-up discussions, and, in extreme cases, the change of class or school for the bully or the victim.

In addition to talking to the bully, the victim should be interviewed as soon as possible, separately from the bully. It is important to obtain information about the duration and frequency of the bullying incident(s), if the victim has had similar experiences in the past, and what the victim has done to try to stop it. During this talk, the incident and nature of discussion should also be documented, the victim should receive information about the plan of action for the bully, and the victim should be taught how to report any future incidents. Assessing the social skills and status of the victim should

be carried out as part of an effort to provide social skills training to enhance his or her sense of personal efficacy to cope effectively with any future bullying.

Because peers play different roles in the bully process, it is important that they be considered as a possible resource in the intervention process. Training for students should also contain an awareness component in which they are taught about the nature of bullying specific to their school and the resources available to them. It is important that their consciousness is raised about the general processes of bullying and the individual’s responsibility as a bystander. Training should focus on increasing empathy for the victims as well as provide peers with strategies to encourage them to withstand the pressures of their peer groups. Despite the common occurrence of bullying on school campuses, most students are not victims or perpetrators. They can have a powerful effect on the school climate as it relates to bullying. In part, bullying occurs on school campuses because it can. If bystanders recognize their important role in setting school standards, then its occurrence can be significantly reduced.

## 8. THE SPECIAL ROLE OF BYSTANDERS

Because peers are such an important part of the bullying process, they merit additional in-depth examination. By its nature, bullying is a social interaction that is embedded in a broader school context. The aggression that occurs between bullies and victims is rarely restricted in its impact to the bully–victim dyad. Rather, peers play an integral part in bullying and intervening. Olweus conceptualized student involvement in an acute bullying situation as being part of a continuum (or “bullying circle”) (Fig. 1). At one end of the continuum are students who take an active part in bullying, support the bully, or follow the bully. At the other end of the continuum are students who defend the victim and dislike the bully’s actions. Many students fit in roles that fall somewhere in between—supporting the bully but not taking an active part, supporting the victim but not helping in his or her defense, or as disengaged onlookers.

Simply by observing the bullying behavior, peers reinforce bullies who often gain social prestige through their aggression. In one study of school bullying, O’Connell, Pepler, and Craig found that when a higher number of peers were observing an incident of

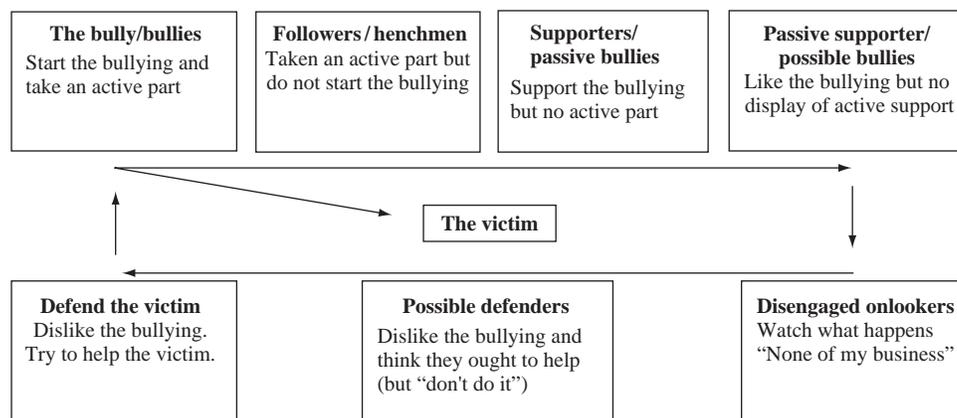


FIGURE 1 Continuum of student reactions in a situation involving acute bullying. Adapted from Olweus (2001).

bullying, the duration of the incident was longer. The presence of peers served to encourage the bully by giving him or her more peer attention. At least for some bullying behavior, one of the primary reasons that bullying occurs is because the bully gains social prestige through this peer attention. When peers are asked for their opinions of bullies and victims, they usually say that the bully is more likeable than the victim (this is thought to be true because the popularity and aggression of bullies are more appealing than the perceived weakness of victims).

Although the influence of peers encourages bullies, the intervention of peers is also one of the most powerful ways to stop bullying. In one investigation, bullying stopped in three-fourths of cases when a peer intervened on behalf of the victim. Because adults are often unaware of instances of bullying, the impact of this peer intervention should not go underestimated. Nonetheless, although 41% of students said that they would like to help the victim, in reality only 25% intervened. The question of what leads peers to action or inaction in the face of a bullying situation is an important consideration that should be discussed openly in every school.

Students with higher social status are more likely to intervene and help the victim than students who are less popular and perhaps fear that they will also be bullied. Another reason for failing to help the victim may be related to the fact that many acts of bullying occur in front of a group. When several students together observe a peer being bullied, they may feel a lack of individual responsibility to help the victim (social psychologists call this the bystander effect). Even when students would like to intervene, they

may not have the skills or know how to help the victim in a way that will be effective. These are topics that are addressed in several intervention programs that educate peers and involve the entire school community.

## 9. IMPLICATIONS

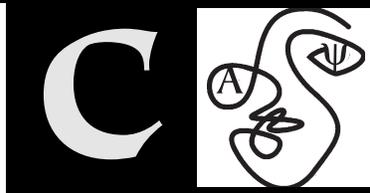
Research has identified the “culture of bullying” on school campuses. Students are less likely to report witnessing bullying or being personally victimized by a bully if they also perceive that in the past teachers at their school have not intervened to stop bullying. Thus, although many prevention and intervention efforts attempt to decrease the aggression of the bully and to enhance the social skills of victims, it is now recognized that if a school has a laissez-faire, inferred acceptance of, or a blind unawareness of bullying, then students are more likely to accept its occurrence as a “normal” part of the school experience. These developments in bullying research show that schools cannot afford to let students work out their conflicts. To do so may send a message that behaviors such as bullying are tacitly tolerated. Bullying is the largest violence/aggression problem in schools worldwide. As such, it is of the utmost importance that schools study their particular circumstances and take action to prevent its occurrence and to intervene when it occurs.

### See Also the Following Articles

Aggression and Culture ■ Emotional and Behavioral Problems, Students with ■ School Discipline and Behavior Management ■ School Violence Prevention

### Further Reading

- Espelage, D. L., & Swearer, S. M. (2003). Research on school bullying and victimization: What have we learned and where do we need to go? *School Psychology Review, 32*, 365–383.
- Juvonen, J., & Graham, S. (Eds.). (2001). *Peer harassment in school: The plight of the vulnerable and victimized*. New York: Guilford.
- Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among U.S. youth: Prevalence and association with psychosocial adjustment. *Journal of the American Medical Association, 285*, 2094–2100.
- O'Connell, P., Pepler, D., & Craig, W. (1999). Peer involvement in bullying: Insights and challenges for intervention. *Journal of Adolescence, 22*, 437–452.
- Olweus, D. (2001). Peer harassment: A critical analysis and some important issues. In J. Juvonen, & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 3–20). New York: Guilford.
- Olweus, D., Limber, S., & Mihalic, S. (1999). *Blueprints for violence prevention series: Book nine: Bullying Prevention Program (BPP)*. Boulder: University of Colorado, Center for the Prevention of Violence. (Information available at [www.colorado.edu/cspv/blueprints](http://www.colorado.edu/cspv/blueprints)).
- Rigby, K. (2004). What it takes to stop bullying in schools: An examination of the rationale and effectiveness of school-based interventions. In M. J. Furlong, M. P. Bates, D. C. Smith, & P. Kingery (Eds.), *Appraisal and prediction of school violence: Methods, issues, and contexts*. Hauppauge, NY: Nova.
- Roberts, W. B., Jr., & Morotti, A. A. (2000). The bully as victim: Understanding bully behaviors to increase the effectiveness of interventions in the bully–victim dyad. *Professional School Counseling, 4*, 148–155.
- Smith, P. K., et al. (2002). Definitions of bullying: A comparison of terms used, and age and gender differences, in a fourteen-country international comparison. *Child Development, 73*(4), 1119–1133.
- Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus bully/victim questionnaire. *Aggressive Behavior, 29*, 239–268.
- Unnever, J. D., & Cornell, D. G. (2003). The culture of bullying in middle school. *Journal of School Violence, 2*(2), 5–27.
- Whitney, I., & Smith, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary schools. *Educational Research, 35*, 3–25.



# Career Counseling

Mary J. Heppner and P. Paul Heppner

University of Missouri, Columbia, Missouri, USA

1. The Evolution of the Definition
  2. History of the Field of Career Counseling
  3. The Role of Frank Parsons
  4. Theories of Career Development
  5. Research in Career Counseling
- Further Reading

## GLOSSARY

**career choice** The act of choosing a career path; it is often first done during adolescence and early adulthood but is often modified as changes in the self, the world, or work change in the course of an individual's life.

**contextual factors in career planning** Factors in an individual's environment that may influence his or her career development; these factors might include the person's social class, race/ethnicity, family structure, and so forth.

**self-assessment** The process of gaining information about oneself (e.g., personality, values, skills, interests) so as to make an informed career decision; it can be conducted using inventories, using computerized career information systems, or having discussions with a career counselor.

**world of work information** Information that often includes job duties, education and training required, salaries and other benefits, and employment outlook; gathering world of work information is often an integral part of the career planning process.

Career counseling is ongoing face-to-face interaction performed by individuals who have specialized

training in the field to assist people in obtaining a clear understanding of themselves (e.g., interests, skills, values, personality traits) and to obtain an equally clear picture of the world of work so as to make choices that lead to satisfying work lives. Career counselors help clients within the context of a psychological relationship with issues such as making career choices and adjustments, dealing with career transitions, overcoming career barriers, and optimizing clients' work lives across the life span. Career counselors are cognizant of the many contextual factors present in the lives of their clients and of the ways in which social and emotional issues interplay with career issues.

## 1. THE EVOLUTION OF THE DEFINITION

The current field of career counseling has been referred to by many different terms such as vocational counseling, career guidance, and vocational guidance. In fact, its definition has evolved to meet the changing context and needs of individuals within society. During the early writings about career counseling, it was described as a process of "matching men and jobs," reflecting the basic function of helping primarily male workers find jobs. As the field progressed, its purview was widened to involve helping both men and women with their career development more broadly and including issues such as decision making, career adjustment, and career

change. There have also been refinements of the definition, including an emphasis on career counseling being done over the life span as opposed to an activity primarily restricted to one's initial job choice. In addition, there has been growing acknowledgment of the intertwined nature of career counseling and social/emotional counseling, with some research indicating that career counseling may help individuals not only with their career adjustment but also with their personal and psychological adjustment. With the growing diversity in society, there has also been particular attention paid to the importance of the cultural context of an individual's life, including factors such as race/ethnicity, social class, age, ability level, sexual orientation, and religion. Thus, the definition and purview of career counseling continues to be dynamic, addressing the needs of a changing society.

## 2. HISTORY OF THE FIELD OF CAREER COUNSELING

The origins of career counseling, broadly defined, can be found during early Greek and Roman times, but the roots of the field as is practiced today can be traced to the last half of the 1800s and the early 1900s in the United States. This was a time when the origins of the human sciences were forming, particularly in experimental psychology, with the work of Wilhelm Wundt. It was also a time when "mental testing" was born, with James Cattell first using this term in an article in the journal *Mind*. Cattell was most interested in using mental tests to examine individual differences, particularly in measuring student achievement. Economically, this was a time of great turmoil and demographic change because the industrial revolution was under way and the country was moving from a primarily agrarian economy to an industrial-based economy. At this time, there was also an influx of immigrants, primarily from Europe, entering the country. Often, these individuals faced a host of social and employment problems in finding employment. Even once they secured jobs, they often faced unhealthy working conditions, overt discrimination, and physical and psychological abuse. There was a strong need for services to help these new workers find their place in the employment structure and to provide a range of social services and social advocacy on their behalf.

## 3. THE ROLE OF FRANK PARSONS

The need for a strong social advocate for these immigrants was found in the person of Frank Parsons, who is often referred to as the "Father of Career Development." Parsons developed what was termed the Breadwinners Institute under the Civic Service House in Boston. The institute was designed to provide a range of educational services for immigrants. It was in this context that Parsons developed a systematic way of helping individuals to find appropriate work that still has much influence on the way in which career counseling is conducted today. Parsons' book *Choosing a Vocation*, which outlined his decision-making process, was published in 1909 after his death. Parsons theorized that there were three broad decision-making factors: (a) a clear understanding of oneself, including one's aptitudes, abilities, interests, and limitations; (b) a knowledge of the requirements, advantages, disadvantages, and prospects of jobs; and (c) the ability to reason regarding the relation of these two sets of facts. These three factors have had an enormous impact on how career counseling has been practiced during the century since it was originally published.

## 4. THEORIES OF CAREER DEVELOPMENT

Career development theories provide foundational knowledge from which to draw useful concepts to explain behavior. They offer a framework within which individual career behavior can be explained and hypotheses can be formed about the possible meaning of behavior. As such, they help career counselors to identify and understand clients' goals and problems within a theoretical framework. Since the inception of the field, career theory has been an important guiding force for the practice of career counseling. Early theories, such as those of Ginzberg, Bordin, Roe, and Super, all helped to provide theoretical approaches to career development. Ginzberg conceptualized the process of career development as being lifelong and subject to compromise, Bordin's conceptualization of career development was informed by a psychoanalytic view of human behavior. Roe built a theory based on Maslow's hierarchy of needs and personality types. Super conceptualized career development as a developmental process and not a one-time choice. All of these early theoretical contributions still have impacts on the way in which career counselors understand human behavior today.

Arguably, the theory that has spurred the most research, and has been used the most in practice, is that of Holland. Holland articulated his comprehensive trait-factor theory in his book *Choosing a Vocation*, which was published in 1973 and revised in 1985 and 1997. Holland's theory proposed that individuals' personalities and work environments can be categorized into one of six types: Realistic, Investigative, Artistic, Social, Enterprising, or Conventional. This RIASEC theory of career development maintains that individuals will search for environments that best fit their personalities and will derive more satisfaction in finding this fit. Holland's theory has been used extensively in the United States and abroad to provide a framework for career counseling.

More recently, a number of additional theories of career development have been proposed, including Lofquist and Dawis's theory of work adjustment, Krumboltz's social learning theory of career decision making, Gottfredson's theory of circumscription and compromise in career choice, and Lent Brown and Hackett's social-cognitive theory.

## 5. RESEARCH IN CAREER COUNSELING

The field of career development has been one of the most actively and rigorously researched fields within counseling psychology. The field has a great deal of information available indicating that career development services, such as workshops, classes, groups, computerized career information, and guidance systems, tend to be highly effective in helping individuals to become clearer and more confident about their career goals. Considerably less information is available specifically about the efficacy of career counseling. Even less is known about what specific aspects of the career counseling process lead to effective outcomes. A recent meta-analytic study conducted by Brown and Krane analyzed 62 studies and approximately 8000 participants in attempting to determine which interventions led to the greatest change. From these data, five specific components of career counseling were shown to contribute significantly to the effect size: written exercises, individualized interpretations of information, the provision of information on the world of work, opportunities for modeling, and the building of support for career choices in individuals' social networks. The authors noted that these five interventions collectively seemed to be associated with remarkable,

nearly linear increases in career choice effect size. Thus, as the methodological tools available to researchers are becoming more sophisticated, we are learning more about what specific interventions are most helpful in the career counseling process.

### See Also the Following Articles

Holland's Theory (Vocational Personality Types) ■ Indecision, Vocational ■ Vocational Assessment in Schools ■ Vocational Interests ■ Vocational Psychology, Overview

### Further Reading

- Blustein, D. L., & Spengler, P. M. (1995). Personal adjustment: Career counseling and psychotherapy. In W. B. Walsh, & S. H. Osipow (Eds.), *Handbook of vocational psychology* (2nd ed., pp. 295–330). Hillsdale, NJ: Lawrence Erlbaum.
- Brown, S. D., & Krane, N. E. R. (2000). Four (or five) sessions and a cloud of dust: Old assumptions and new observations about career counseling. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed., pp. 740–766). New York: John Wiley.
- Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology*, 28, 545–579.
- Gysbers, N. C., Heppner, M. J., & Johnston, J. A. (2003). *Career counseling: Process, issues, and techniques*. Boston: Allyn & Bacon.
- Heppner, M. J., & Heppner, P. P. (2003). Identifying process variables in career counseling: A research agenda. *Journal of Vocational Behavior*, 62, 429–452.
- Holland, J. L. (1997). *Making vocational choices*. Odessa, FL: Psychological Assessment Resources.
- Krumboltz, J. D. (1998). Serendipity is not serendipitous. *Journal of Counseling Psychology*, 45, 390–392.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, 45, 79–122.
- Leong, F. L. (Ed.). (1995). *Career development and vocational behavior of racial and ethnic minorities*. Hillsdale, NJ: Lawrence Erlbaum.
- Super, D. E. (1990). A life-span, life-space approach to career development. In D. Brown, L. Brooks, & Associates (Eds.), *Career choice and development* (2nd ed., pp. 197–261). San Francisco: Jossey-Bass.
- Swanson, J. L. (1995). The process and outcome of career counseling. In W. B. Walsh, & S. H. Osipow (Eds.), *Handbook of vocational psychology* (2nd ed., pp. 217–259). Hillsdale, NJ: Lawrence Erlbaum.



# Cheating in Academics

Gregory J. Cizek

University of North Carolina, Chapel Hill, North Carolina, USA

1. Introduction
2. Definitions and Examples
3. Cheating and Validity
4. Professional Guidelines on Cheating
5. Detecting Cheating
6. Preventing Cheating  
Further Reading

## GLOSSARY

**achievement test** A procedure designed to measure an examinee's attainment of knowledge or skills.

**cheating** Any intentional action or behavior that violates the established rules governing the administration of a test or the completion of an assignment; cheating gives one student an unfair advantage over other students on a test or an assignment and decreases the accuracy of the intended inferences arising from a student's performance on a test or an assignment.

**crib sheet** A piece of paper or other material on which inappropriate or unauthorized material is written for use during a test; it is also known as a cheat sheet.

**high-stakes test** A test for which serious consequences are associated with performance.

**honor code** A formal statement or policy concerning the rights and responsibilities of students and teachers in academic settings as well as the penalties for failing to uphold those responsibilities; the statement may be limited to matters of academic integrity but may also deal with honesty, respect, and citizenship in the academic community more broadly.

**plagiarism** Representing, in written or oral expression, the words, thoughts, or ideas of another person without appropriate citation or referencing, usually for the purpose of intentionally misrepresenting such words, thoughts, or ideas as one's own.

**proctoring** The process of monitoring test takers during an examination; a proctor (usually the instructor in a classroom) observes examinees during a test to discourage cheating and document irregularities or suspicions of cheating, but the proctor can also provide legitimate assistance to test takers if special needs, questions, and the like arise during the test.

**razoring** The removal of material from a printed source document for the purpose of preventing others from gaining access to the source; the term comes from the use of a razor blade to excise pages from, for example, a book or journal held in a library reserve section.

**surrogate** One who takes a test in place of another.

**validity** The degree to which the conclusions yielded by any sample of behavior (e.g., test, assignment, quiz, term paper, observation, interview) are meaningful, accurate, and useful; it is the degree to which a student's performance results in decisions about the student that are "correct" or to which inferences about the student's knowledge, skill, or ability that are "on target."

Cheating in academics refers to actions on the part of students that violate the explicit rules or commonly accepted norms for examinations or assignments. Cheating occurs at all levels, from the primary grades though testing for licensure or certification in a profession. Although most research on the topic of

cheating has focused on inappropriate behavior on the part of those who take tests, cheating in academics can also include cheating by those who give tests. Reasonably accurate methods exist for detecting some types of cheating. Many strategies have also been proven to be effective for organizing instruction and assessment so that cheating can be prevented.

## 1. INTRODUCTION

Sound assessment practice involves the development and administration of high-quality instruments and procedures to measure student learning. Assessment instruments can include graded tests and assignments. Information that results from these instruments can be helpful to students themselves or to others who use the information for educational decision making such as teachers, administrators, and policymakers. Increasingly in education contexts, consequences are also associated with performance on assessments. When the associated consequences are relatively serious, the assessment is said to involve high stakes.

The stakes associated with an assessment can be serious or mild, and they can be positive or negative. For example, awarding scholarship money to a student who performed at an exceptional level on a college admissions test would be considered a serious positive consequence. Assigning an elementary student to a remedial reading program due to a low test score could be either a positive or a negative consequence, depending on whether or not the remedial program was effective and other factors. Withholding a diploma because a student failed a graduation examination would be a fairly serious negative consequence. As the importance of decision making increases, and as the stakes associated with assessment rise, the problem of cheating becomes more pronounced.

Studies of cheating by students indicate that the incidence of the behavior increases as students pass through the grade levels, with the greatest frequency at the high school and college levels. Research evidence indicates that approximately 90% of students admit to having cheated on a test or an assignment, and it is estimated that 3 to 5% of students cheat on a given test or assignment.

## 2. DEFINITIONS AND EXAMPLES

This article focuses on the kinds of cheating that occur in classroom situations. Cheating can occur on tests,

for example, when one test taker copies test answers from another test taker. Cheating may also occur on assignments, such as on term papers or projects, when a student inappropriately uses material taken from another source for the assignment.

Regardless of the situation, a general definition of cheating can be developed. Cheating can be defined as any intentional action or behavior that violates the established rules governing the administration of a test or the completion of an assignment, gives one student an unfair advantage over other students on a test or an assignment, or decreases the accuracy of the intended inferences arising from a student's performance on a test or an assignment.

Varieties of cheating by test takers can range from simple copying of another student's answers to more elaborate schemes. An examinee may use unauthorized materials, such as a cheat sheet, or may take advantage of the testing situation by requesting testing accommodations that are not necessary. Cheating on assignments may involve impermissible books, notes, or persons when the use of such aids is contrary to guidelines provided for an out-of-class assignment. Academic cheating also may include plagiarism, in which the original words, thoughts, or ideas of another person are used without appropriate citation or referencing (usually for the purpose of intentionally misrepresenting such words, thoughts, or ideas as one's own). Many Internet sources exist from which students can, either at no charge or for a fee, copy or download complete term papers or smaller portions of existing works. Finally, other forms of cheating exist in testing situations (e.g., when a person does not actually take a test himself or herself but rather relies on another person, called a confederate or surrogate, to do so) or on assignments when cheating may take the form of razing (which involves the removal of material from a printed source document for the purpose of preventing others from gaining access to the source). More extensive lists of cheating methods were provided by Cizek in 1999 and 2003.

## 3. CHEATING AND VALIDITY

Validity is the single greatest concern in any assessment situation. The concept refers to the accuracy of the interpretations about examinees based on their performance on a test or an assignment. In technical terms, validity is the degree to which evidence supports the inferences made about a person's knowledge,

skill, or ability based on his or her observed performance. By definition, inferences are based on a less than ideal amount of information such as a sample of a student's knowledge or skill obtained via a test or term paper. Because it is often too costly or impractical to gather more information, inferences must be based on samples of behavior. Consequently, it is necessary to consider the accuracy of inferences based on the available evidence (e.g., test performance), that is, to consider validity. Any factor that attenuates the ability to make accurate inferences from the sample of performance threatens validity and jeopardizes the meaningfulness of conclusions based on the sample at hand. When cheating occurs, inaccurate inferences result.

#### 4. PROFESSIONAL GUIDELINES ON CHEATING

Many educational institutions have developed honor codes to formalize expectations about academic integrity; to clearly define the types of actions that constitute cheating on tests, assignments, or other academic work; and to outline penalties for cheating. Test publishers usually produce carefully scripted directions for administering their tests and provide clear guidelines for what kind of behavior is permissible and what kind is not. Acceptable behavior and unacceptable behavior are also sometimes codified in states' administrative regulations or statutes. Assessment specialists, through their professional associations, have developed professional guidelines to inform test takers and test administrators regarding inappropriate practices. The most authoritative of these guidelines are found in the *Standards for Educational and Psychological Testing*, a document that focuses on testing. Among their provisions, the standards indicate that those involved in testing programs should

- protect the security of tests (Standard 11.7);
- inform examinees that it is inappropriate for them to have someone else take the test for them, disclose secure test materials, or engage in any other form of cheating (Standard 8.7);
- ensure that individuals who administer and score tests are proficient in administration procedures and understand the importance of adhering to directions provided by the test developer (Standard 13.10);
- ensure that test preparation activities and materials provided to students will not adversely affect the

validity of test score inferences (Standard 13.11); and

- maintain the integrity of test results by eliminating practices designed to raise test scores without improving students' real knowledge, skills, or abilities in the area tested (Standard 15.9).

#### 5. DETECTING CHEATING

Attempts to detect cheating can be performed by observational/judgmental or statistical methods. Observational/Judgmental methods have the potential to detect cheating on academic assignments and tests. Statistical methods have been developed to detect a particular type of cheating (copying) on tests.

Observational/Judgmental methods rely more heavily on subjective human perceptions. For example, a student might enlist the aid of a surrogate to take an examination in his or her place. Human judgment is involved in detecting this form of cheating; exam proctors must be alert for test takers who do not appear familiar to them, must scrutinize examinees' identifications prior to the test, and so on. Judgment is also involved when handwriting samples from the student are compared with those of the surrogate to make a determination of whose handwriting appears on the test materials or when a student obtains an unusually large score gain from one test administration to another.

Judgment is also involved in the detecting of plagiarism on written assignments such as term papers. A number of techniques can be used to ascertain whether a student's written work is original or may have been plagiarized. Table I lists some common observational/judgmental methods for detecting this form of cheating. In addition to observational/judgmental methods, statistical methods and computer technology can be invoked to aid in the detection of cheating.

Basic search engines (e.g., Google, Yahoo) can be used to locate Internet sources from which plagiarized material may have been drawn. This method relies on an instructor entering a unique word (or, better yet, a unique phrase or string of words) from a student's paper to identify Internet sources containing the same string. Commercial (i.e., fee-based) Web sites devoted to the detection of plagiarism also exist. Most such sites require student work to be submitted electronically. Then, students' papers are compared with those in databases of student work maintained by the host of

**TABLE I**  
**Observational Indications of Potential Plagiarism**

1. The writing style, language, vocabulary, tone, grammar, and/or other features of a student's writing are not commensurate with what the student usually produces.
2. There are awkward changes in verb tense, pronouns, structure, and/or organization of the paper. The writing is choppy. Sections or sentences in the paper seem out of place or do not relate to the overall topic of the paper.
3. Strange text, such as a date in the footer of a paper or a URL (Web site address) in a header, appears at the top or bottom of printed pages.
4. There is unusual or inconsistent formatting such as page layout, line spacing, margins, and/or pagination. The paper contains changes in font, pitch, color, and the like, suggesting that the paper may have been downloaded from the Internet or cut-and-pasted from another source.
5. The paper contains references to graphs, charts, accompanying material, citations, chapters, footnotes, and/or additional text that are not provided with the paper.
6. The paper includes allusions to ideas, persons, settings, and/or other elements with which the student is not likely familiar.
7. The paper contains citations or references to sources or materials to which the student is not likely to have access.
8. Citations or references are missing, inaccurate, or incomplete.
9. The student has difficulty summarizing his or her paper or in responding to simple questions about what he or she wrote.

*Source.* Adapted from Bates and Fain (2002, p. 1) and Cizek (2003a, chap. 4).

the site. An overall probability is produced for each student's work, indicating the overall degree of similarity between the student's paper and all of the other sources. Software for local use that essentially performs a similar function of searching for common phrasings can also be purchased.

Statistical methods are also available to gauge the probability of copying on multiple-choice tests. Information on these methods, including details regarding how to compute statistical indexes of cheating, was provided by Cizek in 1999 and by Sotaridona in 2003.

## 6. PREVENTING CHEATING

Many strategies exist for deterring cheating in academics. Many of these strategies are unique to the

particular type of cheating to be deterred (i.e., whether on a test or on a written assignment). Other strategies aim to foster higher levels of academic integrity in general. Table II provides a list of specific ways in which to prevent cheating on tests and written assignments.

In addition, at the college and university levels, many institutions have found that inappropriate behavior can be prevented by developing and disseminating explicit expectations about academic integrity. Such expectations, formalized in what are referred to as honor codes, help to promote a campus culture and community characterized by the valuing of ethical conduct. An excerpt of one such honor code is reproduced in Table III. The excerpt shown is limited to the portion of the code related to cheating, and a source reference is provided for accessing the full code.

**TABLE II**  
**Strategies for Preventing Cheating**

1. Limit the amount of tests and assignments that "count" toward student grades. Foster an assessment climate in which students apprehend the value of accurate information on their strengths and weaknesses. Avoid using grades punitively.
2. Introduce students to the requirements of academic integrity explicitly and often. Provide concrete illustrations of inappropriate actions. Teach appropriate test preparation and citation practices.
3. Avoid assessment formats that make cheating easier. For example, reduce reliance on multiple-choice items in favor of constructed-response formats. Use alternate forms of tests in which test item positions are scrambled. Avoid "take home" tests. Avoid repeating term paper topics or test questions from year to year.
4. Keep test materials secure in advance of testing. Follow appropriate steps during testing such as requiring adequate spacing between test takers. Remain vigilant while proctoring.
5. Be fair and open about assessment. Provide students with adequate notice and description of examinations and adequate time to complete written assignments.
6. Inform students that statistical methods for detecting answer copying and technological aids for detecting plagiarism will be used.
7. Institute an honor code. Develop and disseminate policies on cheating. Implement penalties for academic dishonesty.

**TABLE III**  
**Sample Honor Code**

**A. General Responsibilities**

It shall be the responsibility of every student at the University of North Carolina at Chapel Hill to

1. Obey and support the enforcement of the Honor Code;
2. refrain from lying, cheating, or stealing;
3. conduct themselves so as not to impair significantly the welfare or the educational opportunities of others in the university community; and
4. Refrain from conduct that impairs or may impair the capacity of university and associated personnel to perform their duties, manage resources, protect the safety and welfare of members of the university community, and maintain the integrity of the university.

Offenses proscribed by this section include, but shall not be limited to, those set out in Sections II.B and II.C.

**B. Academic Dishonesty**

It shall be the responsibility of every student enrolled at the University of North Carolina to support the principles of academic integrity and to refrain from all forms of academic dishonesty, including but not limited to the following:

1. Plagiarism in the form of deliberate or reckless representation of another's words, thoughts, or ideas as one's own without attribution in connection with submission of academic work, whether graded or otherwise
2. Falsification, fabrication, or misrepresentation of data, other information, or citations in connection with an academic assignment, whether graded or otherwise
3. Unauthorized assistance or unauthorized collaboration in connection with academic work, whether graded or otherwise
4. Cheating on examinations or other academic assignments, whether graded or otherwise, including but not limited to the following: (a) using unauthorized materials and methods (notes, books, electronic information, telephonic or other forms of electronic communication, or other sources or methods); (b) violating or subverting requirements governing administration of examinations or other academic assignments; (c) compromising the security of examinations or academic assignments; (d) representing another's work as one's own; or (e) engaging in other actions that compromise the integrity of the grading or evaluation process
5. Deliberately furnishing false information to members of the university community in connection with their efforts to prevent, investigate, or enforce university requirements regarding academic dishonesty
6. Forging, falsifying, or misusing university documents, records, identification cards, computers, or other resources so as to violate requirements regarding academic dishonesty
7. Violating other university policies that are designed to ensure that academic work conforms to requirements relating to academic integrity
8. Assisting or aiding another to engage in acts of academic dishonesty prohibited by Section II.B.

*Source.* University of North Carolina at Chapel Hill (2003).

### See Also the Following Articles

Achievement Tests ■ Educational Achievement and Culture  
 ■ Educational and Child Assessment ■ Psychological  
 Assessment, Standards and Guidelines for

### Further Reading

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.  
 Bates, P., & Fain, M. (2002). *Detecting plagiarized papers*. [www.coastal.edu](http://www.coastal.edu).

Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.

Cizek, G. J. (2003a). *Detecting and preventing classroom cheating: Promoting integrity in schools*. Thousand Oaks, CA: Corwin.

Cizek, G. J. (2003b). Educational testing integrity. In J. E. Wall, & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 363–386). Greensboro, NC: ERIC Counseling and Student Services Clearinghouse.

Sotaridona, L. S. (2003). *Statistical methods for the detection of answer copying on achievement tests*. Enschede, Netherlands: Twente University Press.

University of North Carolina at Chapel Hill. (2003). *Instrument of student judicial governance*. Chapel Hill, NC: Author. (Available: <http://instrument.unc.edu>)



# Cheating in Sport

Glyn C. Roberts, Yngvar Ommundsen, Pierre-Nicolas Lemyre and Blake W. Miller

Norwegian University of Sport Science, Oslo, Norway

1. Introduction
  2. The Context of Sport
  3. What Is Cheating?
  4. The Motivation to Cheat: Achievement Goals as a Determinant of Moral Action
  5. Dispositional Achievement Goals
  6. Perceived Motivational Climate and Determinants of Cheating
  7. What Can We Do about Cheating?
- Further Reading

Recent research reveals that understanding the achievement goal of a person may help to explain cheating in sport. Achievement goals have dispositional elements as well as situational determinants, and both have been associated with cheating and moral action in sport. The more ego involved the person, the more the person cheats. Reducing cheating may be achieved through coach education and by deemphasizing normative comparisons.

## GLOSSARY

- achievement goals** Within an achievement context, the goals that give meaning to achievement striving.
- cheating** A quest to provide an unfair advantage over the opponent; cheating may be “unsportsmanlike” aggressive behavior or inappropriate moral action, but always with the intent to gain an advantage.
- dispositional goals** Achievement goals that are within a person, that is, an individual difference variable.
- ego involvement** When the goal of a person is to demonstrate superiority to peers.
- moral atmosphere** The collective norms of the team about whether or not to condone cheating.
- motivational climate** The perceived criteria of success and failure that a person perceives are manifest within an achievement context.
- task involvement** When the goal of a person is to achieve mastery or learning.

## 1. INTRODUCTION

Cheating in sports is endemic. Each week brings new revelations about cheating, from sprinters who take banned substances (e.g., Dwain Chambers, Carl Lewis, Kelly White) to endurance athletes who take performance-enhancing drugs (e.g., Roberta Jacobs, Richard Virenque, Pamela Chepchumba). But cheating is not limited to individual athletes; prestigious sport organizations have also been caught trying to bend the rules for their own advantage. Some of the most blatant examples of recent cases include the following. Six athletes from the Finnish Ski Federation were discovered to have taken performance-enhancing drugs with the blessing of the medical staff and were disqualified at the Lahti World Cup in 2001. The South Africa, Pakistan, and India cricket teams were found to have fixed matches in 2000. The Welsh Rugby Association recruited illegal players to play for Wales in 2000 by “discovering” fictitious Welsh grandparents for the players. Cyclists on sponsored teams

(e.g., the Festina cycling team) in the Tour de France were caught using banned substances in 2001 and 2002. Cheating has even polluted children's sports, with the most recent visible case being that of Little League pitcher Danny Almonte, who was 14 years of age when he was the winning pitcher for the Little League World Series (restricted to children 9–12 years of age) in 2001. Danny's father has been indicted for the fraud.

At the time of this writing, a huge cheating scandal involving track and field athletes was breaking in the United States. On October 16, 2003, the U.S. Anti-Doping Agency (USADA) released a statement describing the discovery of a new "designer steroid" that had been deliberately manufactured by BALCO laboratories in the United States to avoid detection by current testing procedures. The new steroid (tetrahydrogestrinone or THG) was discovered only after a prominent coach sent a syringe containing the substance to the USADA and named several prominent athletes who were using the steroid. The anti-doping laboratory at the University of California, Los Angeles, identified and developed a test for the substance, and the USADA retrospectively tested the samples from the U.S. Outdoor Track and Field Championships in 2003. Several positive sample results have emerged, and it has been decided that all of the samples from the 2003 World Championships in Paris are to be tested. Terry Madden, the USADA chief executive, stated that this is a case of intentional doping of the "worst sort" and that it involved a conspiracy involving chemists, coaches, and athletes in a deliberate attempt to defraud fellow competitors and the world public. The testing of both "A" samples and, if positive, subsequent "B" samples was ongoing at the time of this writing. In addition, Victor Conte, the founder of the BALCO laboratory that is at the heart of the scandal, was indicted to appear before a grand jury investigating the illegal distribution of a controlled substance. The scandal could involve many athletes and will tarnish many reputations and the sport. Among the athletes named thus far are U.S. shot put champion Kevin Roth, hammer thrower John McEwan, women's hammer throw champion Melissa Price, and middle distance runner Regina Jacobs, who all tested positive at the 2003 U.S. championships. The European sprint champion, the United Kingdom's Dwain Chambers, has also tested positive. Among other "clients" of BALCO laboratories who have been implicated are sprinters Marion Jones and Tim Montgomery, baseball star Barry Bonds, and the NFL's Bill Romanowski.

Not all cheating is as obvious. Each sport has its own manifestation of illegal or inappropriate activity. As an

example, professional soccer has what is termed the "professional foul" where players will sometimes pull an opponent's shirt to prevent or hinder progress with the ball. This is now so common that referees do not call it unless it is very blatant. Football players also "dive" when challenged for the ball, trying to draw a penalty. Baseball players cork their bats. The world's number one golfer, Tiger Woods, has accused many of his fellow professionals of using illegal drivers that give a "trampoline effect" to the ball. Cricket players tamper with the ball to make it "swing" more through the air. And so on,

Why do people cheat in sport? Many explanations have been given, and they are mainly economic (the rewards of being a successful elite athlete are huge), sociological (a breakdown in the moral fabric of society), and psychological (the focus of the current article). However, it is the act of competing that has been indicated the most. It is argued that competition, whether on the sport field, in the classroom, or in the boardroom, is the culprit. When people compete, especially in important contexts, people will cheat to achieve success. When winning is everything, they will do anything to win. But so many people who compete do not cheat, so it might not be competition per se that is the problem. An interesting question may then be asked: What are the determinants of cheating, and how are these manifested in sport contexts? The current authors' research over the past few years has investigated this issue and, like so many other researchers investigating cheating in sport, has looked at children's sport. What are some of the determinants of children engaging in inappropriate behavior in sport? First, one must look at the context of sport, and what it means for children, and then define what is meant by "cheating" before discussing the manifestation of cheating in sport.

## 2. THE CONTEXT OF SPORT

Performance in sport is clearly important for adult athletes; the rewards of competence can be significant. In addition, public interest in sport is high and is reflected in the column inches devoted to sport in the popular newspapers as well as in television airtime. It was estimated that one in four persons worldwide watched the final World Cup soccer match between Brazil and Germany on television in 2002. Thus, the entertainment value of sport is increasing for adults. The context is also important to children for different reasons.

Performing in sport contexts is assumed to be important in the socialization process of children toward

the development of appropriate moral behavior. In play, games, and sport, children are brought into contact with social order and the values inherent in society and are provided a context within which desirable social behavior is developed. The psychosocial and moral development of young participants is fostered when peer status, peer acceptance, and self-worth can be established and developed, and the adoption of various perspectives is enhanced. Sport is also assumed to provide a vehicle for learning to cooperate with teammates, negotiate and offer solutions to moral conflicts, develop self-control, display courage, and learn virtues such as fairness, team loyalty, persistence, and teamwork. Despite popular beliefs that sport builds character, this notion has been questioned. Research has shown that competition may promote antisocial behavior and reduce prosocial behavior.

The context of sports is becoming an increasingly important one for modern-day children. With the demise of children's game-playing culture, children are more and more likely to be involved with adult organized sport competition, even as young as 4 years of age (e.g., motocross in Belgium). In addition, research has demonstrated that the domain of competitive sport is a particularly important context for psychosocial development in that peer status, peer acceptance, and self-worth are established and developed. These social attributes are based on many factors, but one way in which a child can gain peer acceptance and status is to demonstrate competence in an activity valued by other children. One area of competence that is highly valued by children is sporting ability. In fact, because of the modern-day visibility of sport, being a good sport player appears to be a strong social asset for a child, especially in the case of boys. Thus, the context is an important one and is one where cheating to gain advantage or gain acceptance with one's peers can take place.

### 3. WHAT IS CHEATING?

In the area of sport moral action, especially with children, many variables have been investigated. Not all of them may be defined as cheating *per se*. For example, moral atmosphere has been included and refers to the cultural norms developed within a team about whether cheating is condoned or not. However, all of the variables may be defined as inappropriate behavior, at the very least, and range from poor "sportspersonship" to outright aggression so as to achieve a competitive advantage. Because cheating is a difficult concept to

pinpoint and define universally, it may be helpful to view cheating as a product or combination of several more or less moral concepts that are referred to often in the sport psychology literature. For example, cheating in sport has been viewed as a quest to provide an unfair advantage over the opponent. According to this definition, a number of concepts fall into the category of cheating in sport, for example, unsportspersonlike aggressive behavior, inappropriate moral functioning, and a team moral atmosphere where cheating is condoned.

Of all the virtues that sport supposedly fosters, sportspersonship is perhaps the most frequently cited. The virtue of sportspersonship is oriented toward maximizing the enjoyable experience of all participants. Although most people believe that they know what sportspersonship is, the development and understanding of the concept has suffered from the lack of a precise definition and an overreliance on broad theoretical approaches. In essence, a sport participant manifests sportspersonship when he or she tries to play well and strive for victory, avoids taking an unfair advantage over the opponent, and reacts graciously following either victory or defeat.

In an effort to generate a much-needed conceptual base to promote research, recent work from Vallerand and colleagues has contributed to a better understanding of the sportspersonship concept. Vallerand has adopted a social psychological view of sportspersonship that separates the latter from aggression and assumes a multidimensional definition consisting of five clear and practical dimensions: full commitment toward sport participation, respect for social conventions, respect and concern for the rules and officials, true respect and concern for the opponent, and negative approach toward sportspersonship. This definition and instrument has been used in sport to investigate sportspersonship. Research has used the scale; however, the negative dimension has never really worked out and is frequently not included. Therefore, when children respect the social conventions (e.g., shaking hands after a game), respect the rules and officials (e.g., not violating the rules or arguing with the referee), and respect the opponent (e.g., helping an opposing player up from the ground in football), they are considered to be high in sportspersonship.

Sportspersonship is but one dimension of social moral functioning in sport, and one model used to more fully investigate the issue is a model suggested by Rest in 1984. Rest proposed a four-component interactive model of sociomoral action that seems well suited when examining sociomoral aspects in

sport. The first component of the model deals with interpreting the sport situation by recognizing possible courses of moral action. The second component encompasses forming a moral judgment involving judgments about both the social legitimacy and moral legitimacy of inappropriate sport behavior. The third component involves deciding what one intends to do as a solution to the dilemma, and the fourth involves executing and implementing one's intended behavior. Behavior might include incidences of nonmoral actions (e.g., aggression) as well as sportpersonship behavior.

The process of making a moral decision may be influenced by many factors, including motivational factors, whereas actual behavior may be affected by distraction, fatigue, and factors that physically prevent someone from carrying out a plan of action. The interactive nature of the four processes means that factors proposed to act primarily on one process also influence the others indirectly. As an example, when a child is in a competitive context, many kinds of behavior are possible, including whether to cheat or not. A situation may develop where one could cheat to stop a player from scoring, but is it appropriate to do so? If doing so is judged as appropriate, does one intend to cheat to stop the player from scoring? Finally, does one actually cheat to stop the player from scoring? In this way, players go through the components of Rest's model.

In 1985, Shields and Bredemeier argued that a major factor influencing the construction of a moral judgment, and consequently moral behavior, in sport is the moral atmosphere of the team. To capture the perceived judgments by significant others as either approving or disapproving of one's moral actions, research has often included how the players perceived the moral atmosphere in terms of social moral team norms. Sport teams, like all groups, develop a moral atmosphere composed of collective norms that help to shape the moral actions of each group member. An example again is the professional foul in soccer. Sociomoral atmosphere, or perceived sociomoral team norms, has been measured by means of a questionnaire.

Another variable has been the perceived legitimacy of intentionally injurious acts. Participants are asked to imagine themselves competing in an important competitive context. Various intentionally injurious acts are described, and participants are asked to indicate to what degree they agree or disagree with the legitimacy of the actions described.

This is how moral functioning and cheating have been defined in research in sport, especially with children. But why do people cheat? One avenue for

exploring this issue has been to look at the motivation to cheat and has used concepts from motivation theory and research. The basic arguments behind this line of research are considered next.

#### **4. THE MOTIVATION TO CHEAT: ACHIEVEMENT GOALS AS A DETERMINANT OF MORAL ACTION**

The most promising avenue to investigate moral functioning and action in sport has been to use achievement goal theory. This framework assumes that achievement goals govern achievement beliefs and guide subsequent decision making and behavior in achievement contexts. It is argued that to understand the motivation of individuals, the function and meaning of the achievement behavior to the individual must be taken into account and the goal of action must be understood. It is clear that there may be multiple goals of action rather than just one. Thus, variation of behavior might not be the manifestation of high or low motivation per se; instead, it might be the expression of different perceptions of appropriate goals. An individual's investment of personal resources, such as effort, talent, and time in an activity as well as moral action, may be dependent on the achievement goal of the individual in that activity.

The goal of action in achievement goal theory is assumed to be the demonstration of competence. In 1989, Nicholls argued that two conceptions of ability manifest themselves in achievement contexts: an undifferentiated concept of ability (where ability and effort are perceived as the same concept by the individual or he or she chooses not to differentiate) and a differentiated concept of ability (where ability and effort are seen as independent concepts). Nicholls identified achievement behavior using the undifferentiated conception of ability as task involvement and identified achievement behavior using the differentiated conception of ability as ego involvement. The two conceptions of ability have different criteria by which individuals measure success. The goals of action are to meet those criteria by which success is assessed. When task involved, the goal of action is to develop mastery or improvement and the demonstration of ability is self-referenced. Success is realized when mastery or improvement has been attained. The goal of action for an ego-involved individual, on the other hand, is to demonstrate normative ability so as to outperform others. Success is realized when the performance of others is exceeded, especially when little effort is expended.

Whether one is engaged in a state of ego or task involvement is dependent on the dispositional orientation of the individual as well as situational factors. Consider the dispositional aspect first. It is assumed that individuals are predisposed to act in an ego- or task-involved manner. These predispositions are called achievement goal orientations. An individual who is task oriented is assumed to become task involved, or chooses to be task involved, so as to assess demonstrated competence in the achievement task. The individual evaluates personal performance to determine whether effort is expended and mastery is achieved; thus, the demonstration of ability is self-referenced and success is realized when mastery or improvement is demonstrated. In contrast, an individual who is ego oriented is assumed to become ego involved in the activity. The individual evaluates personal performance with reference to the performance of others; thus, the demonstration of ability is other-referenced and success is realized when the performance of others is exceeded, especially when little effort is expended.

Achievement goal theory holds that the state of motivational goal involvement that the individual adopts in a given achievement context is a function of both motivational dispositions and situational factors. An individual enters an achievement setting with the disposition tendency to be task and/or ego oriented (goal orientation), but the motivational dynamics of the context will also have a profound influence on the adopted goal of action, especially for children. If the sport context is characterized by a value placed on interpersonal competition and social comparison, the coach emphasizing winning and achieving outcomes, and a public recognition of the demonstration of ability, a performance climate prevails. This reinforces an individual's likelihood of being ego involved in that context. If, on the other hand, the context is characterized by learning and mastery of skills, trying hard to do one's best, and the coach using private evaluation of demonstrated ability, a mastery climate prevails. An individual is more likely to be task involved in that context. Therefore, being task or ego involved is the product of an interaction of personal dispositions and the perceived motivational climate. However, in the research literature, investigators typically look at one aspect or the other are rarely look at them in combination.

Shields and Bredemeier argued that situational influences may have a great effect on an athlete's moral action. Competitive ego-involving structures may focus the individual's attention on the self and, in the case of team sports, on those comprising the in-group as well. This may reduce the player's sensitivity to the welfare of

opposing players. Extensive involvement in competitive contexts may reduce the person's ability to show empathy, thereby reducing consideration for the needs of others faced with a moral dilemma. Indeed, several studies have shown that participation in competition is associated both with reduced sportpersonship and prosocial behavior and with increased antisocial behavior, hostility, and aggressiveness. However, it is argued here that it might not be competition in and of itself that induces sociomoral dysfunction on the individual and group levels. Rather, it might be the perceived motivational climate that may shape an athlete's moral functioning. Indeed, moral development theorists, such as Rest, agree that moral behavior is intentional motivated behavior. Thus, to predict sociomoral perceptions and actions, one must consider the motivational characteristics of the situation. First, dispositional influences on cheating are examined.

## 5. DISPOSITIONAL ACHIEVEMENT GOALS

When looking at the various important personal factors influencing people's behavior in sport, moral reasoning ability and achievement goal orientations have emerged as significant variables. An early research study found that moral reasoning in the context of sport is much more egocentric than moral reasoning in most situations in everyday life. Athletes, as compared with nonathletes, seem to change their basis for moral reasoning from non-sport- to sport-related activities. This type of reasoning seems to focus on self-interest and personal gain.

Research has shown that when athletes are task or ego oriented, they differ in various aspects of moral behavior as well as in views regarding what represents acceptable behavior in sport. For example, athletes who are primarily concerned with outperforming others (ego oriented) have been found to display less mature moral reasoning, tend to be low in sportpersonlike behavior, approve of cheating to win, and perceive intentionally injurious and aggressive acts in sport as legitimate. In contrast, among task-oriented athletes who tend to use self-referenced criteria to judge competence and feel successful when they have achieved mastery or improvement of the task, greater approval of sportpersonship, disagreement with cheating to win, and less acceptance of intentionally injurious acts have been observed.

Research has found that players who reported higher temptation to play unfairly and greater approval of

behavior designed to obtain an unfair advantage were more likely to be ego oriented. These players also perceived their coaches as being more ego involving than task involving. In addition, the athletes believed that teammates would play more unfairly and would have a higher rate of approval of behavior aimed at obtaining an unfair edge over the opponent. In a study investigating the influence of moral atmosphere (i.e., the collective team norms approving of cheating) on the moral reasoning of young female soccer players, players who were in a culture where aggression and unfair play were tolerated were more ego oriented. Furthermore, it has been found that ego-oriented players had lower levels of moral functioning, had greater approval of unsportsmanlike behavior, and judged that intentionally injurious acts are legitimate. Ego orientation has been related to unacceptable achievement strategies such as the use of aggression. Findings indicate that players high in ego orientation display more instrumental aggression than do participants with low ego orientation, suggesting that participants high in ego orientation may adopt a "win at all costs" attitude.

However, ego-involved individuals can also adopt adaptive achievement strategies. It is believed that when individuals are both ego involved and high in perceived ability, so long as the perception of high ability lasts, these individuals will seek challenging tasks, will revel in demonstrating their ability, and generally will not cheat. But when these individuals experience a change in their perception of ability, they are more likely to cheat so as to continue winning. Research has confirmed this in that high ego-, low task-oriented participants who are low in perceived ability are more likely to endorse cheating in sport than are high ego-, low task-oriented participants who are high in perceived ability. High task-oriented individuals are less likely to approve of cheating in sport than are any other sport participants, regardless of perceived ability.

An athlete high in ego orientation is assumed to believe that winning is the single most important thing when competing. When doubt about one's own ability exists, this may lead the young athlete to believe that it is harder to contribute to his or her team's success by using skills alone. The athlete can then be tempted to cheat so as to help the team win and gain peer acceptance. In some of the current authors' research over the years with young players in soccer, they have found that an ego achievement goal orientation is associated with poor sportpersonship and cheating when perception of ability is low. These findings suggest that young athletes are likely to cheat when they do not believe that they

can succeed by simply playing fair. How far one is willing to go to win is often seen as a positive attribute in the world of competitive sport. If one player is willing to get caught or be reprimanded by the officials while trying to intimidate or undermine the opponent, this athlete is likely to generate a positive response from his or her peers, parents, and coaches when playing on a team where winning is everything. Team loyalty can be demonstrated by going against the rules, particularly if the athlete does not get caught. The current authors have found that athletes with a high ego orientation show low respect for rules and officials as well as low respect for opponents.

When winning is everything and perception of ability is low, cheating and going against the rules may be judged to be acceptable so as to gain peer acceptance. This is true for both boys and girls. But boys cheat more. To illustrate this, two figures are included showing how boys and girls differ in degree if not in kind. These figures present data from the authors' own ongoing research involving young soccer players in the Norway Cup, the largest international competition for children in the world. The data show the impact of being high in ego orientation. In this particular example, the authors, using the Rest model to inform their questions, asked the players some questions about their experience with playing soccer. They wanted to see the relationships of the motivational variables on moral reasoning and moral action variables. [Figure 1](#) shows a canonical correlation chart of the responses of high-ego girls (ages 15–16 years). To be meaningful, the correlations should be greater than .30. As can be seen in the figure, when girls are ego involved, they judge inappropriate behavior as being appropriate, and the most important reason was negative; therefore, they were low in moral reasoning, they intended to cheat, and they judged injurious acts as being appropriate to win.

Not attaining significance, these girls had not cheated over the previous five games and considered the team atmosphere and the coaching climate as being relatively neutral. However, they were quite prepared to injure an opponent in the quest to win. Compare this with [Fig. 2](#), which shows a canonical correlation chart of the responses of high-ego boys. These boys judged inappropriate behavior as being appropriate, and the most important reason was clearly negative; therefore, they were low in moral reasoning, and all of the other variables are clearly positive.

This means the high-ego boys admitted that they had cheated over the previous five games, thought that the team atmosphere supported cheating to win, considered the coach as supporting cheating, and were quite

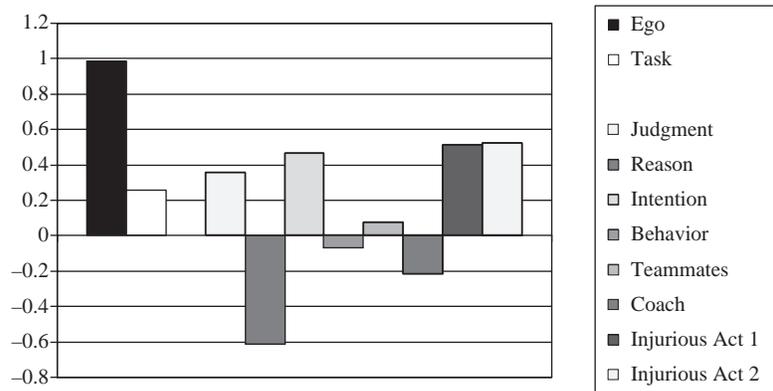


FIGURE 1 Girls: High ego, low task.

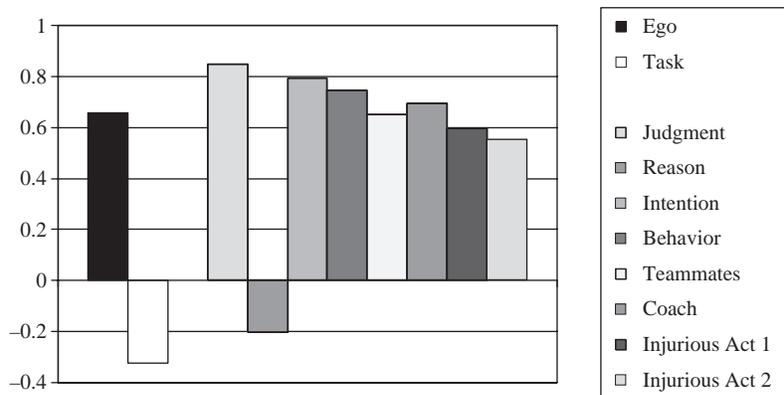


FIGURE 2 Boys: High ego, low task.

prepared to injure opponents in their desire to win. These data show clearly that high-ego players have more tolerance toward and support cheating to win, even to the extent of supporting injurious acts. Although the boys and girls thought and behaved in the same general way, boys are more extreme and do cheat more.

Clearly, being ego or task involved has implications for moral action in sport. The influence that the motivational climate has on moral thinking and action is considered next.

### 6. PERCEIVED MOTIVATIONAL CLIMATE AND DETERMINANTS OF CHEATING

Similar to the task- and ego-involving criteria to which an individual may be disposed, the situational aspects

of the competitive environment are also important. Although a number of significant others (e.g., teachers, parents, peers) can and do influence, to some extent, the type of motivational climate that is perceived, the coach is perhaps the most influential facilitator of task- or ego-involving criteria for the athlete. The bottom line here is that the criteria of success and failure that the coach brings into the situation are the criteria that the athlete adopts. The younger the athlete, the more likely it is that he or she will pay attention to the criteria of the coach. Indeed, there is evidence indicating that the criteria of what constitutes success and failure for the coach “washes out” the personal criteria of the young player. As such, the achievement goal that the coach imposes on the context becomes an important determinant of the behavior of the athlete.

In discussing the motivational climate, the terms “performance climate” (an ego-involving coach) and “mastery climate” (a task-involving coach) are used. If a performance climate prevails, the coach and the

athletes may come to view competition as a process of striving against others. The players may perceive pressure from the coach to perform well or be punished as well as pressure to outplay opponents and win so as to receive recognition and attention. Players may resort to cheating, violating rules, and behaving aggressively as a means of coping. In competition, a performance climate may generate a strong interteam rivalry. A hostile atmosphere toward opposing players may result, leading to the development of team norms or shared social moral perceptions among team members reflecting a derogatory and depersonalized picture of opposing players as mere obstacles to be overcome in the quest for victory by their own team.

The investigation of perceived motivational climate in relation to cheating is a relatively new research area among sport social scientists. The findings thus far indicate that perceived mastery and performance motivational climates created by coaches do affect cheating in systematic ways.

Players perceiving a strong mastery motivational climate reported that sportspersonlike behavior is important in competitive soccer. Players reported that it was important to value and respect the rules of the game and the officials who represented the rule structure. In addition, players noted the importance of respecting the social conventions found within the soccer environment when they perceived the coach as emphasizing mastery motivational criteria. In contrast, it has been shown that players who perceive a high performance climate indicate lower sportspersonship than do players who perceive a high mastery climate.

In general, the majority of research indicates that a perceived performance motivational climate is associated with cheating, whereas a perceived mastery motivational climate is associated with more positive

moral actions. This is true for college-level athletes as well as for younger athletes. In brief, perceptions of performance climate criteria were related to a lower level of moral understanding within the team, resulting in lower moral functioning illustrated by means of low moral judgment, a high intention to behave nonmorally, and low self-reported moral behavior.

Among high-level collegiate male and female basketball players, it has been found that players perceiving a performance motivational climate perceived a low moral atmosphere within their team that was related to lower moral functioning. Various perceptions of the motivational climate have been shown to be related to differing levels of moral functioning and perceptions of team moral atmosphere. Evidence illustrated in Fig. 3 shows that young players who perceive a strong mastery motivational climate indicated positive sportspersonship (i.e., had respect for opponents and officials and also respected social conventions), had positive moral reasoning, had a low intention to cheat, seldom engaged in cheating behavior, and considered the team norms and the atmosphere created by the coach to be anti-cheating.

In contrast, Fig. 4 shows a canonical correlation chart of the responses of girls in a performance-oriented climate. When girls are ego involved, they judged inappropriate behavior as being appropriate, and the most important reason was negative; therefore, they were low in moral reasoning, intended to cheat, had cheated during the previous five games, considered that the coach and teammates approved of cheating, and considered injurious acts appropriate to win. Again, the figure illustrates that even girls (and young female youth), when ego involved through a performance climate created by the coach, will increase their cheating.

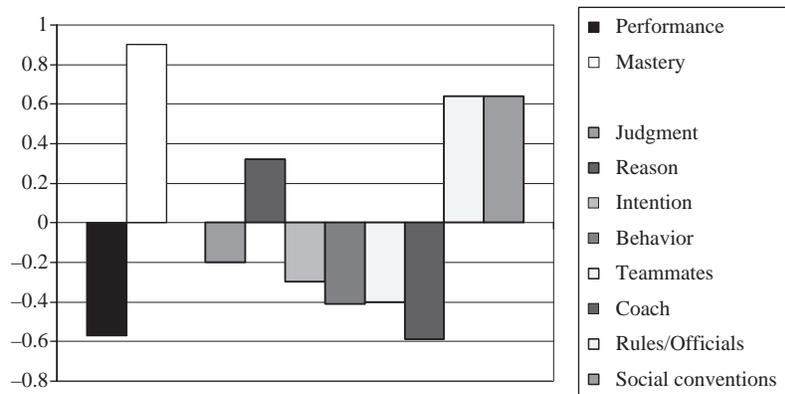


FIGURE 3 Boys: Low performance, high mastery.

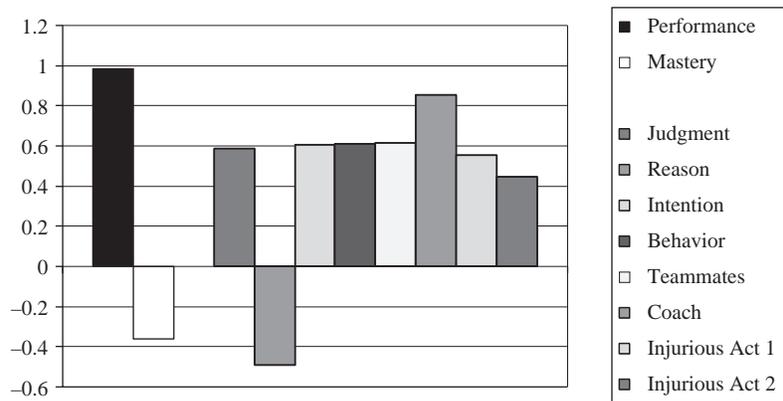


FIGURE 4 Girls: High performance, low mastery.

Gender differences have been found in the motivational climate literature consistently in that young male players react more strongly to performance climate criteria and are more prone to cheating compared with female players: Again, boys cheat more than do girls. However, the important point here is that regardless of gender, a strong coach-created performance motivational climate is related to low moral functioning, inappropriate behavior is judged to be appropriate, and players have lower moral reasoning, have a higher intention to cheat, and actually do cheat more. In addition, both male and female players identified their team moral atmosphere to be more supportive of cheating. These data show clearly that when players are ego or task involved, they think differently, have different ideas about cheating, and report different episodes of intention to cheat and actual cheating. Ego-involved players have more tolerance toward and support cheating to win, even to the extent of supporting injurious acts.

The research findings are quite consistent. When the coach emphasizes criteria of success and failure showing that he or she values winning above all else, athletes indicate that they tolerate cheating more and actually do cheat more than athletes who perceive their coach as emphasizing more mastery criteria of success and failure. This is true for elite athletes, at the professional and Olympic levels, as well as for child and youth sport participants. Clearly, the coach matters, but the major point here is that the coach does not have to endorse cheating per se; rather, it is simply that the coach who emphasizes winning above all else creates a motivational climate that the athletes interpret as being more accepting of cheating behavior.

## 7. WHAT CAN WE DO ABOUT CHEATING?

Some people deliberately cheat. There are enough cases in sport, in corporate business, and in everyday life demonstrating that some individuals and organizations simply cheat to gain an unfair advantage. There is little that can be done about these systematic cheats except to try to catch and punish them to the full extent possible. Whether that is through drug testing in sport, financial oversight in corporate business, and so on, regulations and oversight are needed to cull out the cheats. But cheating can be much more subtle and unintentional in many cases. It is this latter form of cheating that is addressed in these final remarks.

Competitive sport often places people in conflicting situations, where winning is emphasized and where fair play and justice are sometimes deemphasized. It would be wrong to attribute this to the competitive nature of sport. The research reported in this article suggests that the culprit is not competition in and of itself. Rather, the culprit may be the difference in salience of ego- and task-involving cues in the environment, whether through one's own disposition or through the cues emphasized by one's coach (or parent). It is these perceived criteria of success and failure that induce differential concern for moral action and cheating. A focus on winning may reinforce prejudice (i.e., "us against them") and lead to a more depersonalized view of opponents that, in turn, makes cheating more possible.

How can the development of good sportsmanship and sound moral reasoning be maximized, and how can cheating be prevented? Clearly, one way in which to help is to teach the coach to reinforce the importance of

task-involving achievement criteria in the competitive environment. This is important at all levels, especially in light of the degree of cheating going on in professional and Olympic sports. But this is more important for children's and youth sports. This does not mean that one cannot strive to win; rather, it means that we should apply different criteria to the meaning of winning and losing. By giving feedback to players on effort, hard work, and self-referenced increments in competence, more respect for rules, conventions, officials, opponents, and appropriate moral reasoning and action could be cultivated. One way in which to say this is that there are no problem athletes, just problem coaches (and parents). Coaches must be aware of the criteria of success and failure that they are emphasizing. If task-involving criteria are emphasized, cheating can be reduced and moral reasoning and action can be enhanced.

When focusing on children in the competitive sport experience, it is not just the coach who needs to be aware of the criteria being emphasized; parents are the most important source of information for the children. Indeed, it is often the parents who are most to blame in children's sports. However, coaching programs that address both parents and coaches now exist in most countries. The evidence is accumulating that strategies and instructional practices should be developed to facilitate the coach (and parents) in creating a task-involving coaching environment that is concerned with mastery criteria of success and failure. If there is concern about optimizing the motivation, psychosocial development, and moral functioning and action of children, the impact that teaching and coaching style have on these outcomes cannot be ignored.

However, coaches must realize that simply trying to enhance mastery criteria is often not enough. The ego-involving criteria are emphasized as well. Normative evaluation should not be used for athletes. Comparisons should be made with children's own past performance rather than with the performance of their peers. By giving feedback based on their own past performance and trying to give feedback to athletes privately rather than in an announcement to the whole team, coaches can do their best to maintain the motivation of the athletes and remove an important determinant of cheating in sport.

Will cheating ever be eliminated from sport? That is unlikely in elite sport when the rewards are so high. But in children's sport, attempting to reduce cheating is critical. Playing fair and recognizing justice for all should be a natural part of the sport experience. If the sport experience is the context in which children learn to relate to their peers, not only in sport but also for life, the criteria of success and failure inherent in the sport experience should be a critical concern.

### See Also the Following Articles

Fair Treatment and Discrimination in Sport ■ Goal Setting and Achievement Motivation in Sport ■ Intrinsic and Extrinsic Motivation in Sport

### Further Reading

- Dunn, J. G. H., & Dunn, J. C. (1999). Goal orientations, perception of aggression, and sportsmanship in elite male youth ice hockey players. *Journal of Sport and Exercise Psychology, 13*, 183–200.
- Kohn, A. (1986). *No contest: The case against competition*. Boston: Houghton Mifflin.
- Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Cambridge, MA: Harvard University Press.
- Ommundsen, Y., Roberts, G. C., Lemyre, P. N., & Treasure, D. (2003). Perceived motivational climate in male youth soccer: Relations to social-moral functioning, sportpersonship, and team norm perceptions. *Psychology of Sport and Exercise, 4*, 397–414.
- Rest, J. R. (1984). The major components of morality. In W. Kurtines, & J. Gewirtz (Eds.), *Morality, moral behaviour, and moral functioning* (pp. 356–429). New York: John Wiley.
- Roberts, G. C. (2001). Understanding the dynamics of motivation in physical activity: The influence of achievement goals on motivational processes. In G. C. Roberts (Ed.), *Advances in motivation in sport and exercise* (pp. 1–50). Champaign, IL: Human Kinetics.
- Shields, D. L., & Bredemeier, B. J. L. (1995). *Character development and physical activity*. Champaign, IL: Human Kinetics.
- Vallerand, R., & Losier, G. F. (1994). Self-determined motivation and sportsmanship orientations: An assessment of their temporal relationship. *Journal of Sport and Exercise Psychology, 16*, 229–245.



# Child Custody

David A. Martindale    Jonathan W. Gould  
Morristown, New Jersey, USA    Charlotte, North Carolina, USA

1. Psychology's Contribution to Understanding Best Interests
2. Psychological Assessment: What Can We Measure?
3. The Marriage of Clinical Skills and Forensic Precision
4. Dancing with Attorneys without Tripping
5. Following the Rainbow  
Further Reading

## GLOSSARY

*best interests of the child doctrine* The philosophy, now incorporated into the law in all 50 states, that in making decisions concerning custodial placement and parenting plans, the interests of the children should supersede those of the parents and/or others involved in the dispute.

*evaluation of comparative custodial suitability* A forensic psychological evaluation the primary focus of which is the relative strengths and deficiencies in the ability of parents to parent the specific children whose custodial placement is at issue.

*forensic role* Within the context of custody evaluations, a perspective and actions that are consistent with one's obligation to function as an impartial examiner and to gather information that bears upon specific issues before the court.

*pendente lite* A reference to a matter that one or both parties assert must be addressed by the court while the primary issues in litigation are still pending.

*therapeutic role* Within the context of custody evaluations, a perspective and actions that are consistent with one's obligation to assist those with whom one is professionally interacting.

This article begins with a historical overview of the manner in which child placement disputes have been adjudicated and a discussion of the best interests of the child standard. The outline of the factors that psychologists endeavor to assess includes a discussion of the methodological problems inherent in this form of assessment. The significant differences between therapeutic work with families and forensic assessment of families are explained. Because performing custody evaluations requires that psychologists work in a collaborative manner with attorneys and adapt to procedures with which mental health professionals are often unfamiliar, the adjustments that must be made are discussed. The article concludes by providing some perspectives on experts offered by two judges.

## 1. PSYCHOLOGY'S CONTRIBUTION TO UNDERSTANDING BEST INTERESTS

The historical context of custody evaluations begins with Roman law and the presumption of paternal preference. Children were viewed as the property of their fathers. A father had absolute power over his children. He was allowed to place his children into slave labor or sell them for profit. A child's mother had no legal rights. English common law also provided for absolute paternal power. Children were viewed as the property of the father and he had sole discretion over where the children lived, both during the

marriage and after a divorce. Mothers had restricted access to their children after divorce.

Though the reasoning varied, custodial placement decisions based primarily on generally accepted sex roles were the norm until 1970. In 1813, in deciding on the custodial placement of two children, ages 6 and 10, the Supreme Court of Pennsylvania declared: "It appears to us, that considering their tender age, they stand in need of that kind of assistance which can be afforded by none so well as a mother" [*Commonwealth v. Addicks and Lee*, 5 Binney's Rep. 520 (1813), at 521]. At approximately the same time, in Great Britain, what came to be referred to as the tender years doctrine dictated that children younger than age 7 required the care of their mothers. From age 7 on, the custodial responsibility for children would revert back to the father. As the work of fathers began to take them outside the home and as mothers became the primary caretakers, a preference for placement of children with their mothers developed. By the 1920s, a presumption in favor of the mother was clear in most American courts.

Comfort with the notion of a maternal preference was buttressed by Freud's then-popular theory that, among other things, stressed the unique role of mothers in the lives of children. Theories of bonding contributed further to the view that children needed to be with their mothers. Current research has demonstrated that infants form meaningful attachments to both parents by the middle of their first year; however, only recently has such research raised questions concerning the wisdom of a maternal preference.

In the 1960s, several social changes stimulated a reexamination of the maternal preference. These factors included an increase in the number of fathers wishing to maintain an active postdivorce role in the lives of their children, the entry into the workforce of large numbers of mothers, and an increased sensitivity to the discriminatory effects of placement decisions predicated on sex.

In 1970, the U.S. Congress passed the Uniform Marriage and Divorce Act, providing for a best interests of the child standard. The tide had turned from a focus on paternal or maternal preference or rights to a focus on what is best for the child. Simultaneously, psychology offered a revolutionary idea: The concept of the psychological parent was developed. Once society had decided that children, being the most innocent victims of divorce, should be protected as effectively as possible from its emotional consequences, the view emerged that mental health professionals could assist triers of fact in determining what parenting plan was most likely to best serve the interests of the children.

The best interests standard is not without its critics. It has been criticized as poorly defined, as too ambiguous, and as providing fodder for disputes and litigation. Mnookin has argued that the broad judicial discretion provided by the best interests standard leaves judges free to impose their personal biases and beliefs on their judicial determinations. The Supreme Court, in deciding *Troxel v. Granville* (a grandparents' visitation case), made note of the fact that the Washington Superior Court trial judge (whose decision was overturned) had deemed it appropriate to "look back at some personal experiences" [*Troxel v. Granville*, 530 U.S. 57 (2000), at 61].

The challenge is to develop a legal standard that is sensitive both to the social values of American culture and to the needs of the legal system. A legal standard addressing the best interests of the child should "offer effective and useful guidelines, so that similar cases are decided similarly without extralegal factors significantly affecting final dispositions" (Krauss & Sales, 2000, p. 845). A legally effective child custody standard would balance the needs of the family, the needs of society, the evolving social science research, the evolving case law precedents, and the changes in each family configuration while minimizing state intrusions into the functioning of the family.

In many states the criteria to be utilized in evaluating comparative custodial suitability are defined by statute. In states where criteria are not so defined, experts conducting evaluations of comparative custodial fitness bear an obligation to articulate the criteria that they will utilize in performing their assessments.

## 2. PSYCHOLOGICAL ASSESSMENT: WHAT CAN WE MEASURE?

In 1971, the Supreme Court handed down a decision in *Griggs et al. v. Duke Power Co.* (401 U.S. 424), a case that, on its face, had nothing whatsoever to do with children and custodial placement decisions. The *Griggs* case involved assessment in the selection, placement, and promotion of personnel. The court ruled that testing procedures must be demonstrably reasonable measures of (or predictors of) job performance. In the years since 1971, forensic psychologists involved in custody work have come to view the *Griggs* decision as an exhortation, urging us to focus our attention and our assessment efforts on functional abilities that bear directly on the issue before the court. Within the context of custody

work this suggests that we must endeavor to ascertain what attributes, behaviors, attitudes, and skills are reliably associated with effective parenting and confine ourselves to assessing those characteristics as they relate to the children who are the focus of the dispute.

Our problem is that the only assessment instruments with established reliability and validity provide data pertaining only to one or two of the criteria of potential interest to us. An instrument such as the MMPI-2, for example, may provide useful information concerning a test taker's general mental/emotional health, but that leaves much to be assessed about the individual's parenting. LaFortune and Carpenter (1998, p. 222) list the seven most frequently used assessment instruments that "focus on parenting skill and the parent-child relationship" and that are "touted by their authors as helpful in clinical determinations of parental fitness." LaFortune and Carpenter declare: "[T]he validity of these measures is unestablished at best and seriously flawed at worst" (p. 222). Although the hope is expressed that improvements in these instruments may make them useful in the future, the authors conclude that their use at present "cannot be recommended" (p. 222).

Although frequency of use by surveyed mental health professionals is not an acceptable measure of either the reliability or the validity of an assessment instrument, a disturbingly high proportion of custody evaluators use instruments that are psychometrically deficient but that are popular, largely because of successful marketing. Additionally, some clinicians have brought to forensic work psychodiagnostic assessment instruments generally considered useful in health care settings but ill suited to the evidentiary demands of forensic work. In a therapy setting, a psychodiagnostic assessment marks the beginning of an ongoing relationship in the course of which there will be opportunities for subsequent reassessment. In a forensic setting, the report in which one's assessment is described marks the end of a relationship. No opportunities to reassess are provided. Because of this critical difference between clinical assessment and forensic assessment, it cannot be presumed that instruments popular among clinicians are suitable in forensic work.

Consider the following, from the Standards for Educational and Psychological Testing (American Education Research Association *et al.*, 1999): Tests are to be accompanied by documentation that will provide test users with "the information needed to make sound judgments about the nature and quality of the test, the resulting scores, and the interpretations based on the test scores" (p. 67). Elsewhere, the

authors of the Standards opine: "The greater the potential impact on test takers, for good or ill, the greater the need to identify and satisfy the relevant standards" (p. 112).

The authors of the Standards have called attention to the fact that "[w]hen directions to examinees, testing conditions, and scoring procedures follow the same detailed procedures, the test is said to be standardized. Without such standardization, the accuracy and comparability of score interpretations would be reduced" (p. 61). Many of those currently performing evaluations of comparative custodial suitability came to this work via clinical psychology, where being helpful is not viewed as an impediment to assessment. Moreland, a former member of the NCS staff and the MMPI-2 restandardization team, observed that "in responding to questions, examiners may inadvertently hint at the nonpathological answer when the test taker is someone whose position they favor (whether it be a custody case, injury case, or whatever)." He added: "[N]either NCS nor the Eyde group has told test givers how to deal with questions from test takers" (e-mail from Kevin Moreland, August 22, 1996).

In the vast majority of custody disputes, it is presumed by all involved that both parents are suitable. The court's task, with advisory input from the mental health professional, is to decide whether, in addition to being capable, the two parents can coparent effectively. If they can, a joint custodial arrangement may be appropriate; if they cannot, one of these two suitable parents must be designated as the primary custodian. If a primary custodian must be designated, then the court must decide which of two presumably suitable parents is more suitable. Psychologists, and those who seek our advisory input in custodial placement disputes, must recognize that we have no way of evaluating the validity of our recommendations because we have no way of knowing what would have occurred if the nonfavored parent had, instead, been selected as the primary custodian. Otto *et al.* (2003) observed that "research in this area will always be constrained by the inability to use true experimental designs to address the most prominent questions related to custody" (p. 203).

The foregoing paragraphs have an unmistakably pessimistic tone. Some might argue that we do not do custody evaluations well and that if we cannot do them well, we should not do them at all. It is reasonable to presume that as increasingly more psychologists become involved in assessing comparative custodial fitness and share with their colleagues their ideas for improving our procedures, we will become increasingly

more proficient. Deciding not to do custody evaluations is not the answer.

### 3. THE MARRIAGE OF CLINICAL SKILLS AND FORENSIC PRECISION

Most psychologists currently practicing or preparing to practice forensic psychology obtained their original training in clinical psychology. Forensic work is built on solid clinical training and experience; however, to function effectively as a forensic psychologist, the therapeutic mind-set must be altered significantly. It is also important to recognize that there is, in the words of Greenberg and Shuman (1997), an "irreconcilable conflict between therapeutic and forensic roles" (p. 50). A psychologist who is serving or who has served in a therapeutic capacity with any of those involved in a custody dispute should not function as the evaluator. Similarly, an evaluator should not offer therapy or counseling services to those whom he or she has evaluated.

In a forensic setting we seek answers to questions that have been posed (either directly or implicitly) by the court. We are not attempting to determine the etiology of a litigant's problems, nor are we attempting to set therapeutic goals and formulate treatment plans. Instead of communicating warmth, we communicate skepticism. Although that may sound harsh, it is important that litigants be reminded that information offered by them will not simply be written down in our notes and accepted at face value. Reasonable attempts will be made by us to verify information given to us. Section VI.F of the *Specialty Guidelines for Forensic Psychologists* (Committee on Ethical Guidelines for Forensic Psychologists, 1991) reminds forensic psychologists that "[w]here circumstances reasonably permit, forensic psychologists seek to obtain independent and personal verification of data relied upon." Ceci and Hembrooke (1998, p. 4) have expressed the view that "the uniquely human qualities that compel us to want to help are the very qualities that can make us poor experts."

Some of the most serious problems that forensic psychologists create for themselves stem from an inability or unwillingness to control the impulse to think like and act like "helpers" when they are obligated, by virtue of their forensic task, to function as examiners. Shuman (1993, p. 298) has opined that the techniques customarily employed by mental health professionals to establish a therapeutic alliance "are entirely inappropriate in

a court-ordered examination." Pruett and Solnit (1998, p. 126) express the view that "empathic skills must be constrained in a manner unusual for clinical practice but essential for fair, ethical, evaluative interaction."

The prevailing presumption in clinical work is that those who appear before us honestly share with us their perceptions of themselves and those with whom they interact. There is a tendency to presume, further, that discrepancies between a client's descriptions and objective reality are attributable, primarily, to the operation of various perception- and memory-distorting phenomena of which the client is not consciously aware. Only limited thought is given to the possibility of deliberate, calculated lying.

Individuals being examined within the context of custody disputes are strongly motivated to present themselves as being better adjusted than they know themselves to be. An examination that might be deemed adequate in a clinical context might be viewed as inadequate in a forensic context. Regardless of the assessment instruments utilized, a forensic examiner must operate more like an investigator than a sympathetic and supportive listener. In particular, custody evaluators cannot be passive recipients of offered information. We must use clinical skills to elicit information and actively seek, both from the litigants and from appropriate collaterals, any information that might reasonably be viewed as pertinent. Data collected by Ackerman and Ackerman suggest that approximately 28% of the evaluative time expended by those responding to a survey was devoted to information verification (obtaining information from documents, disinterested collateral sources, and other nonparties). Bow and Quinnell found that 98% of evaluators review documents relevant to the case, 86% interview therapists, 78% interview teachers, and 52% interview physicians and/or pediatricians.

### 4. DANCING WITH ATTORNEYS WITHOUT TRIPPING

In the field of custody work, most of the missteps that occur in interactions between forensic psychologists and attorneys relate to conflicting obligations. Such conflicts are easy to understand when psychologists are endeavoring to function as impartial examiners. Clashes are often unanticipated when psychologists have been retained by attorneys to function as their consultants or to offer expert testimony. Even when functioning in the consultant role, however, psychologists' ethical

obligations differ from those of the attorneys who have retained them and choreography can still be a challenge.

It is not uncommon for judges and attorneys to ask that a psychologist functioning as impartial evaluator alter his or her customary data collection sequence in the name of expediency or offer interim recommendations to assist the court in resolving a *pendente lite* issue. With regard to data collection, there is a logical progression to the information-gathering aspect of a custodial fitness evaluation, and although some modifications are acceptable, others are not. Data obtained through the administration to the parties of psychological tests and information secured by means of various questionnaires inevitably suggest areas of inquiry. It is inadvisable to conclude one's evaluative sessions with the litigants without having in hand test data and the questionnaires.

Standard 9.01 (a) of the psychologists' ethics code (American Psychological Association, 2002) declares: "Psychologists base the opinions contained in their recommendations, reports, and diagnostic or evaluative statements, including forensic testimony, on information and techniques sufficient to substantiate their findings." In the absence of sufficient information, recommendations should not be offered. All involved should also be mindful of the fact that parenting plans set in place in response to specific and immediate concerns alter family dynamics in ways that are not readily predictable and that may not be in the best interests of the children.

Court appointment does not guarantee evaluator competence. Neither does competence assure findings and recommendations with which all involved will be pleased. Where an expert's opinions have not been formulated through the utilization of appropriate procedures and are not supported by reliable data, exposing these deficiencies is a critically important function of an opposing expert.

The *Specialty Guidelines for Forensic Psychologists* (Committee on Ethical Guidelines for Forensic Psychologists, 1991) closes with a reminder (contained in section VII.F) that "[f]orensic psychologists are aware that their essential role as expert to the court is to assist the trier of fact to understand the evidence or to determine a fact in issue." Psychologists retained as consultants to attorneys or retained to offer expert testimony are, despite their collaborative relationships with those who have employed them, obligated to promote understanding and to decline participation in partisan attempts to distort or misrepresent evidence. Just as those whom we evaluate should be provided with information concerning our policies and procedures, so, too,

should attorneys who retain us be provided with such information. Information provided in advance prevents problems.

## 5. FOLLOWING THE RAINBOW

Standard 2.01 (c) of the psychologists' ethics code states, in part: "Psychologists planning to provide services . . . new to them undertake relevant education, training, supervised experience, consultation, or study." Section 2.01 (e) states, in part: "In those emerging areas in which generally recognized standards for preparatory training do not yet exist, psychologists nevertheless take reasonable steps to ensure the competence of their work" (p. 1064). In 2002, the American Psychological Association formally designated forensic psychology as a specialty. The obligation of psychologists trained in clinical psychology to professionally prepare themselves for forensic work if they plan to offer forensic psychological services is clear yet occasionally ignored.

The performance of an evaluation of comparative custodial suitability is, by definition, a forensic psychological endeavor. In our view, such evaluations should be guided by the forensic model. Earlier, we alluded to the Supreme Court's expressed disapproval of a trial court judge's decision to "look back at some personal experiences" [*Troxel v. Granville*, 530 U.S. 57 (2000), at 61] in formulating his opinion in a visitation dispute. The opinion formulation process to which the judge made direct reference is, more often, unarticulated. There is reason for concern that psychologists, whose services have been sought because of their presumed expertise, occasionally reflect on personal experiences, values, and beliefs in formulating their opinions. What makes an opinion an expert opinion is not the professional credentials of the individual whose lips form the words or whose keyboard is used to type them. What makes an opinion an expert opinion is the manner in which the opinion was formulated. Professional credentials are necessary but not sufficient.

An expert opinion is formulated utilizing the accumulated knowledge and the recognized procedures of the expert's field. If the accumulated knowledge of the expert's field was not utilized, the opinion expressed is not an expert opinion. It is a personal opinion, albeit one being expressed by an expert. Judge Alex Kozinski has astutely observed that the task facing triers of fact, when expert testimony has been offered, "is to analyze not what the experts say, but what basis they have for saying

it" [*Daubert v. Merrell Dow Pharmaceuticals, Inc.* (on remand), 43 F.3d. 1311 (9th Cir. 1995), at 1316]. The Honorable Stephen Hjelt, the presiding Administrative Law Judge for the California Office of Administrative Hearings in San Diego, in what might best be described as an open letter to psychologists, opined that our "profession has strong roots as a discipline that has a foundation in the scientific method. However, some of you simply stopped using it" (Hjelt, 2000, p. 12). We close by urging psychologists, attorneys, and judges to consider the words of judges Kozinski and Hjelt. Psychologists should use their training in methodology, cross-examining attorneys should demand that testifying experts articulate the bases for their opinions, and judges should consider the data that form the basis for expert opinions and not be gullible consumers of recommendations offered by experts.

### See Also the Following Articles

Diverse Cultures, Dealing with Children and Families from  
 ■ Eyewitness Identification ■ Forensic Mental Health Assessment

### Further Reading

- Ackerman, M. J., & Ackerman, M. C. (1997). Custody evaluation practices: A survey of experienced professionals (revisited). *Professional Psychology: Research and Practice*, 28, 137–145.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073.
- Bow, J. N., & Quinnell, F. A. (2001). Psychologists' current practices and procedures in child custody evaluations: Five years after American Psychological Association Guidelines. *Professional Psychology: Research and Practice*, 32, 261–268.
- Ceci, S. J., & Hembrooke, H. (1998). *Expert witnesses in child abuse cases: What can and should be said in court*. Washington, DC: American Psychological Association.
- Committee on Ethical Guidelines for Forensic Psychologists. (1991). Specialty guidelines for forensic psychologists. *Law and Human Behavior*, 15(6), 655–665.
- Coons, J., Mnookin, R., & Sugarman, S. (1993). Deciding what's best for children. *Notre Dame Journal of Law and Public Policy*, 7, 465–490.
- Eyde, L. D., Robertson, G. J., Krug, S. E., Moreland, K. L., Robertson, A. G., Shewan, C. M., Harrison, P. L., Porch, B. E., Hammer, A. L., & Primoff, E. S. (1993). *Responsible test use: Case studies for assessing human behavior*. Washington, DC: American Psychological Association. [The authors are also referred to, collectively, as the Test user Training Work Group of the Joint Committee on Testing Practices.]
- Greenberg, S. A., & Shuman, D. W. (1997). Irreconcilable conflict between therapeutic and forensic roles. *Professional Psychology: Research and Practice*, 28(1), 50–57.
- Hjelt, S. (2000). Professional psychology: A view from the bench. *Register Report*, 26(1), 8–13.
- Kelly, J. B., & Lamb, M. E. (2000). Using child development research to make appropriate custody and access decisions for young children. *Family & Conciliation Courts Review*, 38(3), 297–311.
- Krauss, D. A., & Sales, B. D. (2000). Legal standards, expertise, and experts in the resolution of contested child custody cases. *Psychology, Public Policy, & Law*, 6(4), 843–879.
- LaFortune, K. A., & Carpenter, B. N. (1998). Custody evaluations: A survey of mental health professionals. *Behavioral Sciences and the Law*, 16, 207–224.
- Lamb, M. E., & Kelly, J. B. (2001). Using empirical literature to guide the development of parenting plans for young children: A rejoinder to Solomon and Biringen. *Family Court Review*, 39, 365–371.
- Martindale, D. A., & Gould, J. W. (2004). The forensic model: Ethics and scientific methodology applied to custody evaluations. *Journal of Child Custody: Research, Issues, and Practices*, 1, 1–22.
- Mnookin, R. (1975). Child custody adjudication and judicial function in the face of indeterminacy. *Law and Contemporary Problems*, 39, 226–293.
- Otto, R. K., Buffington-Vollum, J. K., and Edens, J. F. (2003). Child custody evaluation. In A. M. Goldstein (Ed.), *Handbook of psychology: Vol. 11, Forensic Psychology* (pp. 179–208). New York: Wiley.
- Pruett, K. D., & Solnit, A. J. (1998). Psychological and ethical considerations in the preparation of the mental health professional as expert witness. In S. J. Ceci, & H. Hembrooke (Eds.), *Expert witnesses in child abuse cases: What can and should be said in court* (pp. 123–136). Washington, DC: American Psychological Association.
- Shuman, D. W. (1993). The use of empathy in forensic examinations. *Ethics and Behavior*, 3, 223–302.
- Warshak, R. A. (2000). Blanket restrictions: Overnight contact between parents and young children. *Family and Conciliation Courts Review*, 38(4), 422–445.
- Whiteside, M. F. (1998). Custody for children age 5 and younger. *Family & Conciliation Courts Review*, 36(4), 479–502.



# Child Development and Culture

Cigdem Kagitcibasi

Koc University, Istanbul, Turkey

1. Study of Child Development in Psychology
  2. Development in Context
  3. Development of Competence
  4. Nature–Nurture Debate and Plasticity
  5. Implications for Application
  6. Conclusion
- Further Reading

## GLOSSARY

*cultures of relatedness* Collectivistic societies where interpersonal distances are small between connected selves.

*heritability* The portion of variability in a trait that is passed to offspring via genes; high heritability is often understood to mean high genetic causality.

*mechanistic model* A behaviorist approach that interprets “environment” in a very limited sense as proximal stimuli.

*organismic model* An approach that emphasizes biologically based maturation, ignoring the context of development.

*scaffolding* Temporary support and guidance in problem solving.

Human development occurs in a cultural context. Obvious as this statement may be, its realization in developmental science is far from adequate. Most mainstream research and theory that derives from the Western (particularly American) academic centers and informs the field tends to approach human development in and of its own, without much attention to its cultural aspects. This state of affairs has implications for how

child development is construed, how it is assessed, and what is done (if anything) to enhance it. This article presents the current knowledge regarding the interface of culture and child development, with the main processes and dynamics involved, and also provides a brief overview of the scholarship on the subject. The main theoretical conceptualizations regarding child development, as well as their implications in significant spheres of applied fields, are reviewed.

## 1. STUDY OF CHILD DEVELOPMENT IN PSYCHOLOGY

A main issue in the psychological study of child development is the scarcity of a contextual approach. The prominence of the mechanistic and organismic models in developmental psychology and the stress on the individual as the unit of analysis have worked as deterrents to contextual conceptualizations of child development. The mechanistic model, used by behaviorism, has construed “environment” in a very limited sense as proximal stimuli. The organismic model has stressed biologically based maturation, ignoring the context of development. These perspectives are of both historical and contemporary influence. Recent advances in cognitive neuroscience, genetic research, and the like have further strengthened the “biological shift.” The focus has been on the individual organism as a product of these two major influences—more recently, as a combination of the two. Thus, a vast

majority of current psychological research on child development is conducted in the laboratory and does not take the larger cultural context into account.

Notwithstanding this state of affairs, there is also a rich tradition of research and thinking in child development, particularly in anthropology, that takes culture seriously. Introducing culture into child development has been done, starting mainly during the second half of the 20th century, first in anthropology and then in ecological, cross-cultural, and cultural psychology. Although this work remained rather marginal for some decades, it is beginning to be better recognized, to increase in volume, and to leave its mark on mainstream research and conceptualizations in developmental psychology. One reason for this changing orientation is the fact that research informed by cultural context is now being done in modern nations, including those in the Western world, rather than in isolated small human groups, as was the case with classic anthropological work. So, it is seen as more relevant by psychologists. A second reason is the fact that Western societies, especially the United States where there is the greatest amount of research production, have become multicultural societies. The growing cultural diversity has again made culture more relevant in the psychological study of development.

### **1.1. Study of Child Development Within Culture**

Research on child development that takes culture seriously is being done within two different paradigms. One of these, known as cross-cultural psychology, uses a comparative approach in which theory testing is of high priority. Cross-cultural comparative methodology works as a corrective to the rather ethnocentric assumption of the pan-human generality of findings obtained in unicultural studies. The view here is that the findings may in fact have universal generality, but this has to be demonstrated rather than assumed. Indeed, the theoretical basis for a comparative methodology is the “universalism” of psychological processes, although their manifestations in behavior differ due to cultural factors. The methods/instruments used are similar to those used in unicultural psychological studies of development; however, the issue of equivalence in measurement emerges as much more significant in cross-cultural methodology.

The other paradigm is cultural psychology, which claims that cultural context gives meaning; thus, human behavior is context specific and should be

studied within culture, not comparatively. For example, topics such as the cultural organization of human development within family and society and the cultural construction of intelligence are studied. This is akin to an anthropological perspective stressing the uniqueness of each cultural context. Accordingly, research tends to be based on observations within context and descriptive qualitative analysis rather than on experimentation and standardized measures with a comparative orientation that are typical in cross-cultural research. For example, observational studies have been conducted extensively within the tradition of “everyday cognition” or “informal learning” by children of culturally relevant skills such as tailoring, woodcarving, weaving, pottery, and other handicrafts as well as street mathematics.

The two paradigms have also been called *etic* and *emic*, in analogy with phonetics and phonemics in linguistics. Phonetics studies general aspects of sounds in languages, whereas phonemics studies sounds in a particular language. The different perspectives underlying cultural and cross-cultural psychology and the *etic* and *emic* approaches are posited as conflicting; however, they both contribute to our understanding. They both have emerged as correctives to the “culture-blind” orientation of mainstream developmental psychology’s organismic and mechanistic paradigms. However, they both entail some risks when taken to an extreme. The comparative cross-cultural approach can claim “false universals,” whereas the cultural approach can assume “false uniqueness” (of each cultural context). The challenge is to combine the benefits of each in understanding the unity and diversity within the dynamics of culture and child development.

### **1.2. Life Stages and Culture**

Over the past few decades, the study of human development has moved from a focus on childhood to a focus on the life span. The span of development is seen to be continuous and to cover the whole spectrum, from prenatal period to old age and death. This approach renders culture even more relevant because cultural differences emerge quite significantly during the various phases of the life span. Thus, although there tends to be a great deal of commonality in the meaning attributed to, and the socially described role of the growing human during, infancy and early childhood, cultural variations start and increase progressively during the subsequent stages of middle and later childhood, adolescence, young and later adulthood, and old age. For example, early anthropological work showed that adolescence

and gender roles during adolescence in some preliterate societies are quite different from those in Western technological societies. Even when less diverse socioeconomic cultural contexts are compared, such as more affluent/urban and less affluent/rural ones, the roles of both adolescents and children in family and society are found to differ. In particular, a much greater proportion of children and adolescents work, carrying out household chores or family business, and are economically productive in less affluent/rural and more traditional society. For example, the fact that children in Africa spend less time in “play” and more time in “work,” as compared with children in the United States, can be attributed to many factors such as the necessities of less affluent lifestyles but also the cultural meaning of childhood and children’s work. Nsamenang claimed that child work in West Africa is a social learning process for adult roles and, therefore, an inherent part of development. From a Western human rights perspective, however, child work can be seen as child “labor”—and, therefore, child abuse—especially if it interferes with formal education. This is a case in point for the various cultural meanings accompanying different lifestyles that can be a source of debate or even conflict.

Thus, cultural meanings associated with different life stages reflect lifestyles, social norms, and social values. Old age is another life stage that is defined and understood variously. Whereas old people in Western technological societies tend to lose social status and experience some isolation from the community of young people, old people in many traditional societies remain very much integrated within family and community and are respected for their life experience, knowledge, and wisdom. They also tend to be more active in important domestic and communal roles involving imparting of traditions to the young, moderating dispute settlements, and the like.

Thus, culturally informed life span perspectives show that age does not have the same meaning across space and time. What is universal is possibly the tendency of all cultures to use age as a basis for defining the individual and assigning him or her with certain roles, responsibilities and privileges. With this in mind, this article focuses on the life stage of childhood in the cultural context.

## 2. DEVELOPMENT IN CONTEXT

The context of child development includes numerous levels of influences, all interrelated with and embedded within one another. One way of conceptualizing

contextual variables is in terms of their degree of comprehensiveness as encompassing systems. This ecological perspective has been influential in the study of child development in the cultural context. For example, Bronfenbrenner differentiated among four levels, increasing in complexity and comprehensiveness from micro-, to meso-, to exo-, to macrosystems. Determining which of these ecological systems is to figure in the study of child development at any particular time is an empirical issue. An important component of the macro-system of the ecology of child development is societal values regarding children that are reflected in parental beliefs, values, and child-rearing orientations.

The recognition of the significant implications of the ecological perspective for child-rearing orientations and the resultant child outcomes have led to concepts such as “ethnotheories” or “naive theories” (e.g., LeVine’s theory of parental goals). Super and Harkness proposed the concept of the “developmental niche,” consisting of the physical and social setting, child care and child rearing, and the psychology of the caretaker, to explain the cultural structuring of child development. Similarly, the “ecocultural niche” (suggested by Weisner), and the “developmental microniche” (suggested by Worthman) have been proposed as organizing concepts to model the relationships of biology, behavior, and culture in shaping human development. Other theoretical models have also entailed an ecological/cultural contextual framework, such as Berry’s “ecocultural framework,” further combined with ecological systems theory by Georgas and with the developmental niche by Dasen.

An ecological perspective also calls for situating parental beliefs and child-rearing orientations within the socioeconomic and social structural context. This is important to understand why certain belief systems show systematic variation across different parental populations. The answer often lies in the underlying functional relations. Ecological perspectives have also figured in other analyses. For example, bringing the temporal dimension into an ecological perspective in forming a sociological life events approach, Elder studied intergenerational relations and adolescent adjustment and development at various historical periods. Similarly, concepts of developmental pathways and “canalization” over an epigenetic landscape were used by Waddington to analyze variations and self-stabilizing constraints that restrain the variations in the organism–environment system as it changes and evolves over time.

These conceptualizations aim to deal systematically with the myriad spatial and temporal environmental influences that provide developmental pathways for

the individual. Nevertheless, the individual not only is acted on but also acts on the environment so as to create its own developmental pathway that is unique yet also has common characteristics with others. This view is stressed particularly by European-based theories of individualization.

### **2.1. The Value of Children and Family as Developmental Niche**

Study of parental values and child-rearing orientations can serve as an example of an ecological approach to development in the cultural context. Values attributed to children and how childhood is conceptualized provide important clues as to the place of children in the family and society. A cross-cultural study on the Value of Children (VOC), conducted during the 1970s and currently being replicated in several countries, points in particular to the economic/utilitarian and psychological values attributed to children. The economic/utilitarian value is stressed in sociocultural contexts where children's material contributions to the family are significant. This often takes the form of contributing to family income while young and providing elderly parents with old-age security during later years. Psychological VOC, on the other hand, is stressed in contexts where children entail more economic costs than economic assets. These dynamic relations among ecological factors, family/parental values, and child outcomes are dealt with by another ecological theory, namely the "family change model" (proposed by Kagitcibasi).

In sociocultural contexts of low affluence, rural/agrarian/low-socioeconomic status standing where children in fact contribute to the family's material well-being, the economic/utilitarian VOC is salient, child work is common both in and out of the house, and the family is characterized by intergenerational material interdependencies. This is the family model of "interdependence." In contrast, in more affluent contexts where children are in school and are costly to raise, the psychological VOC is salient, child work is negligible, and intergenerational material interdependencies are weak. This is the family model of "independence." The interdependent pattern is more prevalent in collectivistic cultures or "cultures of relatedness" with closely knit human relations. The independent pattern is more typical of the individualistic societies or "cultures of separateness." Thus, there is a general correspondence among societal conceptions of childhood, parental values, family patterns, and children's actual lifestyles.

In the family model of interdependence, there is obedience-oriented child rearing that does not promote the development of autonomy. This is because the independence of children is not valued and might even be seen as a threat to the family livelihood because independent offspring may look after their own self-interests rather than those of the family. Intergenerational interdependence is manifested through the family life cycle first as the dependence of the children on their parents, with this dependence being reversed later on as the dependence of the elderly parents on their adult offspring for their livelihood.

A contrasting pattern is seen in the family model of independence, characteristic of the Western (mostly American) middle-class nuclear family, at least in professed ideals. Here, child rearing is oriented toward engendering self-reliance and autonomy, and independence is also valued between generations, where objective social welfare systems and affluence render family interdependence unnecessary or even dysfunctional. Individuation-separation is considered a requisite for healthy human development and is also reinforced by (popular) psychological teaching.

It is commonly assumed (e.g., by modernization theory) that there is a shift from the model of interdependence to the model of independence with urbanization and economic development. However, recent research shows that the changes in family interaction patterns often do not follow such a linear route. In collectivistic cultures of relatedness where interpersonal distances are small between connected selves, individuals and families do not shift toward individualistic separateness with urbanization and economic development. Connectedness continues, although mainly in the realm of psychological interdependencies, while material interdependencies weaken with changing lifestyles. This leads to a third pattern proposed by Kagitcibasi, the "family model of psychological interdependence." In some ways, it is different from both the traditional (rural) family characterized by total interdependence and the individualistic (urban) family characterized by independence.

Child rearing involves control rather than permissiveness because the goal is not individualistic separation. However, autonomy also emerges in child rearing, mainly for two reasons. First, autonomy of the child is no longer seen as a danger because elderly parents are no longer dependent on their grown-up offspring for their material livelihood. Second, autonomy of the growing person is functional in urban lifestyles where, for example, decision making (rather than obedience) is required in school and in more specialized jobs.

## 2.2. Development of the Self

Thus, the family is the developmental niche for the self. Studying the family provides insights into understanding the self, in particular how and why the different types of self develop. For example, a distinction is made between the independent self and the interdependent self, and the cognitive and behavioral concomitants or consequences of these two types of self are examined. Thus, current work on culture and self focuses on cross-cultural variability, particularly along the dimension of independence–interdependence. However, why and how these different types of self develop is not well understood except in reference to macrosocietal characteristics such as individualism and collectivism.

Family is the significant mediating variable here that sheds light on the functional underpinnings of self development. In the family model of interdependence the “related self” develops and is interdependent with others. In the family model of independence, the “separate self” develops and is independent. The third family model of psychological interdependence is more complex because it entails both autonomy and connectedness in child rearing, with the resultant “autonomous-related self” (proposed by Kagitcibasi). The coexistence of autonomy with connectedness is important because the two are often seen to be antithetical in psychology. Under the influence of psychoanalytic thinking, particularly object relations theory and the separation–individuation hypothesis, separation is considered to be a requisite of autonomy in human development. Therefore, the implication is that connected selves cannot be adequately autonomous. This is a central debate, particularly in understanding adolescence. A psychoanalytically oriented individualistic perspective, such as that proposed by Steinberg and Silverberg, considers adolescence to be a “second separation–individuation process” where detachment from parents is seen as a requisite for the development of autonomy. Others, such as Ryan and colleagues, suggest that individuation during adolescence is facilitated not by detachment but rather by attachment. Recently, there also has been some recognition of the compatibility of relatedness and autonomy in the adolescent and adult attachment literature. The autonomous-related self challenges the individualistic assumption. It also goes beyond the dualistic conceptualization of the independent self versus the interdependent self that is prevalent in cross-cultural psychology.

Current research provides support for the compatibility of autonomy and relatedness, indeed pointing to their

combination as being a psychologically more healthy state that satisfies the two basic human needs for autonomy and connection (merging). The family model of psychological interdependence also finds research support in both the Western and non-Western contexts. For example, a more positive relationship is found between autonomy and relatedness than between autonomy and separateness in both Korean and American samples; positive links, rather than negative links, are found between relatedness to parents and autonomy in adolescents in the United States; combined autonomy and control orientation is noted among Chinese and Korean parents; parental autonomy goals do not imply separateness, and achievement values are associated with parental collectivism (rather than individualism), among Turkish parents in Germany; family interdependencies coexist with some individualistic values in Hong Kong; and Chinese and Chinese American parents are found to endorse both relatedness and autonomy, together with high control of and closeness with their children.

Parental orientations are of crucial importance in leading to diverse developmental outcomes. Systematic variations are noted even in parents’ orientations to infants, providing evidence that the various developmental pathways and their combinations may have their roots all the way back to infancy and are reinforced throughout the life span. Such research, informed by cultural and cross-cultural perspectives, opens up new vistas in the development of the self that can shed light on the interface of culture, parenting, and the individual through time.

## 3. DEVELOPMENT OF COMPETENCE

Child rearing is goal oriented, although this is mostly not made explicit. Often, the goal is competence in the sense that socialization, by definition, implies becoming a competent member of a social group. Competence in this perspective refers to what is culturally valued, thereby showing variation across cultures. This sphere of child development is the meeting ground for the development of the self and cognitive development.

In more traditional contexts, particularly with closely knit human ties and interdependent family systems, cultural conceptualizations of cognitive competence include a strong social component. Because much research from Africa has pointed to this, it has been called “African social intelligence”; however, it is much more widespread, particularly in traditional

collectivistic cultures of relatedness with closely knit human ties. It reflects cultural valuing of social responsibility and sensitivity to others' needs and expectations. Obedience orientation in child rearing, discussed earlier, is also involved here because it engenders dependent children who grow up to be loyal adult offspring. Independence and separation are not socialization goals.

Research also shows that, particularly in nonindustrialized rural/agrarian contexts of low levels of affluence, children develop practical skills that serve the family and reflect material dependence on children's work. For example, from early on, children do household chores, take care of animals and babies, and even cook food for the family—tasks in which most urban middle-class children would perform poorly. However, the same children often do poorly in simple cognitive tasks with which urban middle-class children have no difficulty. Clearly, children's cognitive competence in culturally valued domains gets promoted, whereas development in other domains lags behind if it is recognized at all. For example, Serpell showed that "folk" conceptions of intelligence in a Zambian village differed significantly from what is measured by intelligence tests, even culturally sensitive ones. Dasen found in Ivory Coast (West Africa) that intelligence involved mainly social and practical skills (e.g., "being good with hands").

### **3.1. Environmental Change**

An issue in the development of competence in cultural context is social change. The channeling of competence toward culturally valued domains is adaptive so long as there is stability in lifestyles and societal demands. A problem emerges, however, in contexts of social change when adaptive mechanisms get challenged by modifications in lifestyles. Such modifications often accompany social structural and economic changes, especially urbanization and migration. For example, rural-to-urban mobility in developing countries is of immense proportions, also feeding into international migration. Although the young population (10–19 years of age) in developing countries was mainly rural in 1990 (approximately 600 million rural vs 300 million urban), the rural and urban populations of youth are expected to become equal before 2015. Thus, this significant global human mobility calls for a better understanding of what is adaptive and what is maladaptive in the cultural construction of children's developmental niches in contexts of social change.

For example, research on ethnic minorities in North America and Europe, mostly migrants from other parts of the world, points to misfits between parental beliefs

and values and to new environmental demands emerging from urban living conditions, particularly schools. In the face of social change and mobility, the question of whether there is an optimal fit between (traditional) child-rearing orientations and children's developmental trajectories becomes relevant. Such a question challenges cultural relativism, which assumes that what is culturally valued is by definition the right channel of child socialization. The matter turns into an empirical issue open to investigation by research. Such research calls for the possibility of using common (minimal) standards of child development and competence. It also points to the potential of applied work involving intervention.

### **3.2. Cognitive Development: Piaget and Vygotsky**

A great deal of developmental research in the cultural context has focused on cognitive development. A general conclusion is that basic cognitive processes appear to have cross-cultural similarity; however, their manifestation in behavior shows variability given the diversities in cultural/environmental factors that may serve as constraints or reinforcers. The distinction between capacity and performance comes to the fore, as does the issue of culturally sensitive assessment of cognitive competence.

Most theory in psychology, and also in developmental psychology, claims universality, particularly with the weight of the organismic perspective mentioned previously. This is true for Piagetian theory, the generality of which has been subjected to a great deal of cross-cultural empirical testing with variable results. The most important cross-cultural support for the theory is given to the sequencing of stages of cognitive development. Although the particular ages at which the stages or substages are attained, or whether they are attained, show cross-cultural variability, reversal of stage sequence is never found. Beyond this, the first stage of sensorimotor intelligence is found to be a "strong" universal, shown in all studies with infants. The preoperational period is also considered to be a candidate for a strong universal. Even during these early periods, however, differences emerge, for example, in terms of whether objects or social stimuli are handled or attended to more (apparently due to the different types of stimuli, language use, and maternal responsiveness to which infants and young children are subjected in different cultures).

During the period of concrete operations, cultural differences become more important such that in some cultures, some children fail to engage in concrete

operations even when they reach adolescence. Dasen suggested that the qualitative characteristics of concrete operational development (reasoning types) are universal but that the rate of development in certain domains, particularly quantification and spatial reasoning, is influenced by environmental factors. Finally, individuals in many societies never achieve the formal operational stage of cognitive development. Schooling is the main factor that plays a role in whether formal operations are mastered. Actually, this factor is not confined to the nonindustrialized, non-Western contexts. Not all adults in Western technological society achieve this type of thinking. When confronted with this research evidence, Piaget accepted the role of experience and culture for the development of formal operational thinking.

Vygotsky's sociocultural theory of development takes human development out of the confines of the individual and into the cultural and historical realms. As such, it has been important for contextualist perspectives, particularly those of cultural psychologists. Vygotsky stressed the role played by culture, language, and the child's "zone of proximal development" in cognitive development. This is not a stage theory, but the evolution of speech through "social," "egocentric," and "inner" speech periods is sequential. Research on infant-mother interaction across several cultural groups, and informed by Vygotskian thinking, pointed to cultural differences, particularly in various types of maternal responsiveness and speech. Later, the child is aided by the adult through verbal interaction and instruction to reach beyond the current level of competence. The theory that proposes a continuous process of interaction between persons and sociocultural factors is used as a framework in "everyday learning" or "apprenticeship" situations involving "scaffolding" (i.e., temporary support and guidance in problem solving) by an adult. However, it is also criticized as vague and not conducive to measurement or hypothesis testing. As is also apparent from this glimpse of the two most significant theories of cognitive development, diversity and commonality across cultures challenge developmental theory.

#### 4. NATURE-NURTURE DEBATE AND PLASTICITY

Cross-cultural research can help to provide insights into some of the basic issues in child development. This is mainly because a cross-cultural developmental approach uncovers a greater range of variation than does a single-culture study. With increased coverage of diversity, a

wider perspective emerges according to which what is typical and what is atypical might need to be redefined. It also becomes more possible to distinguish between biological and environmental influences. That is, the greater the commonality found across varied cultural contexts, the greater the likelihood of a biological basis for a developmental process, whereas a finding of increased diversity across cultures implies environmental causality.

The perennial nature-nurture debate is a case in point. Although there is much greater sophistication in the treatment of the issue today compared with the earlier, rather simplistic comparisons, the issue stands unresolved. In several spheres, including temperament, personality, intelligence, and gender identity, the nature and malleability of child development can be better understood if cultural context is taken into consideration. "Heritability" refers to the portion of variability in a trait that results from heredity. As such, high heritability is often understood to mean high genetic causality and, thus, a negligible role for the environment. Yet even with extremely high heritability, the environment can still have a powerful influence. For example, a surge in the heights of second-generation Japanese children raised in the United States was reported. The heritability in this group was more than .90, yet the American-reared sons were substantially taller than they would have been if they had been reared in Japan. Similarly, it is reported that the heights of young adult males in Japan increased by approximately 3½ inches since World War II. If a trait as highly heritable as height can show such fluctuations in relatively short time periods and across societies, other traits with lower levels of heritability, such as intelligence and personality, can change as well.

A currently popular view stressing biological determination is Scarr's genotype-phenotype theory, which holds that each individual picks and chooses from a range of environments those that best fit his or her genetic potential/orientations. Thus, the theory claims that genetic endowment creates the environments that individuals choose for themselves. When cultural context is taken seriously, however, the limitations of the theory become obvious. The theory assumes that people have equal opportunity or options from which to choose, yet a great many people in the world are not in a position to choose much at all. Environmental opportunities vary greatly; particularly in deprived environments, the genetic human potentials might not be realized. Thus, heritability for intelligence, for example, would be expected to be lower in deprived conditions than the .50 estimated for the general population

and would be expected to be higher in “good” environments than in “poor” ones because the former provide the resources for the biological potential to be realized.

Indeed a great deal of research points both to lower levels of measured intellectual performance and to other indicators of poor competence such as school failure in socioeconomically deprived environments. Other evidence regarding the powerful influence of the context of development and the plasticity of human development comes from studies of adoption after privation and the effects of schooling.

An early study reported massive intelligence quotient (IQ) gains in children born to low-IQ mothers after they were given up for adoption into environmentally enriched middle-class homes. Although children’s IQs still correlated with those of their biological mothers, indicating genetic influence, they were much higher than the mothers’ IQs and even higher than those of the general population. This study was criticized for its methodology, but its findings were corroborated by later research. For example, recent work with Romanian adoptees from orphanages after the fall of the Ceausescu regime points to the crucial importance of both early environmental deprivation and plasticity of human development. It was found that severely malnourished children (below the third percentile in head circumference, body size, and developmental level) recovered to normal development by 4 years of age if they were adopted into English homes before 6 months of age. Those who were adopted later, however, did not fare as well.

There is debate concerning the long-term effects of early deprivation. Some consider the deleterious effects to be critical and long-lasting, even referring to the first 2 years as possibly a “critical period” in human development. Others are less certain due to the limited capacity of infants to process their experiences cognitively and/or due to plasticity, particularly with regard to catching up in physical growth. Nevertheless, there appears to be consensus that whatever the genetic potential, environment, particularly during the early years, has an important impact on human development and that even quite high values of heritability still leave plenty of room for environmental enhancement or reduction of human potential.

## **5. IMPLICATIONS FOR APPLICATION**

Beyond “saving” children from deprivation through adoption, there are other more common implications

for application of the significant cultural/environmental impact on human development. Some examples of implications for application, involving advocacy and program- and policy-relevant research, would help to put the issues and their possible solutions into perspective.

### **5.1. Developmental Norms**

Growth norms for infants and children have been established by pediatricians. Some of these have a great deal of cross-cultural validity. However, developmental norms paralleling physical growth norms are largely missing and are badly needed for assessment, screening, and detection of developmental risks. With better public health measures such as immunizations, significant increases have been achieved in the rates of infant and child survival globally. However, the circumstances that used to put infants at risk for death continue to put the survivors at risk for arrested or delayed development. Thus, the question of what happens to those who survive assumes great importance. With new policy shifts focusing on early childhood care and development, developmental psychologists are now being called on for their expertise. This is required especially for the cross-culturally valid conceptualization and assessment of human development.

Interactive conceptualization of nutrition, health, and psychosocial development, with each one acting on the other two, is proposed to better represent the dynamic processes of child development in the cultural context. For example, the effectiveness of health and nutrition programs can be increased by combining them with early childhood stimulation and education. The role of the proximal environment, particularly child care, is of importance here as a mediator of the macro-level influences. Thus, within the same poverty situations, some children thrive, whereas most others stay behind. Mother–child (or caretaker–child) interaction emerges as an important “shield” factor in promoting the child’s resilience, conceptualized as “positive deviance.” Here, supporting and training the mother or caretaker goes a long way toward leading to better child outcomes.

In dealing with disadvantaged contexts of child development, culturally valid assessment tools and developmental norms are crucially important. The challenge is to develop child development indicators that have cross-cultural validity to function as “standards” of development and also have cultural validity. Environmental indicators are also of crucial importance and are even less well established cross-culturally. For example, research points to the positive effects of maternal literacy

and education on child survival, longer intervals between pregnancies, better child physical growth, and more developmentally facilitative maternal child-rearing patterns. A concept such as “optimal parenting” becomes relevant here. When made applicable to large numbers of children and parents, such knowledge can be used for detection and screening of risk cases and for establishing and promoting the environmental factors conducive to healthy child development.

### 5.2. Schooling

Schooling is another case in point. A great deal of evidence points to the significant contributions of education, particularly schooling to societal development. Of greater relevance for psychology is the contribution of schooling for cognitive development. For example, Ceci showed that each additional month that a student remains in school may increase his or her IQ score above what would be expected if the student had dropped out. Other research provides similar evidence. Furthermore, a causal relationship is implicated from education to IQ, rather than the other way around, on the basis of higher-IQ children choosing to stay in school longer, as claimed by genotype–phenotype theory mentioned previously. There are different types of evidence supporting this causal relationship, including the effects of intermittent school attendance, delayed school entrance, length of school year, summer vacations, and early-year birth dates.

Given the overwhelming evidence regarding the societal and individual benefits of schooling, there needs to be strong advocacy for increased schooling for all, particularly for girls. Access to schooling is behind schedule in most of the world. For example, during the 1990s, when overall school enrollment reached 92% in North America, it remained at a mere 42% in Sub-Saharan Africa. Although it is developmental psychological research that reveals the positive effects of schooling on cognitive competence and human potential development, psychologists play only a marginal role in efforts to make “education for all” a reality in the world. At the same time, schools should be improved so as to entail better pedagogy and both culturally relevant and globally shared knowledge. Psychologists have a role to play here as well. For example, Stevenson and colleagues studied the cultural contexts of schooling and its cognitive products, focusing on the values and practices of Japanese, Chinese, and American schools and parents, with significantly different achievement outcomes for

children. The implications for educational policy and practice are far-reaching.

### 5.3. Early Enrichment

The importance of early experience for the development of competence is well recognized, notwithstanding the significance of organismic factors. Therefore, concerted efforts have been expended to provide deprived children with early enrichment in attempting to enhance their ability to benefit from formal schooling. “School readiness” is defined in terms of the child’s activity level, social competence, psychological preparedness, and cognitive abilities, including preliteracy and prenumeracy skills. It is also reflected in the positive orientation and support of the child’s family.

Research and applications during early childhood education abound in North America and Europe but are few and far between in developing countries. An important issue is whether a “center-based” or a “home/community-based” contextual approach is to be used (a related distinction is that between a child-focused and a parent-focused orientation). A general conclusion that can be drawn from research in various cultural contexts is that if the child’s family environment can provide the child with the necessary stimulation, a center-based, child-focused approach is effective; otherwise, parental involvement is needed to improve the home environment. A home-based contextual approach has the potential to create lasting effects that would continue to support the child even after the completion of the early enrichment program, as evidenced by longitudinal research. It also tends to have multiple goals and multiple effects in supporting and changing the home environment, particularly the parents, together with the child. Given the crucial importance of the proximal environment and particularly of parenting, a contextual and culturally sensitive early enrichment can go a long way toward increasing the possibility that children from disadvantaged backgrounds can enter adulthood on a more equal footing with those from relatively advantaged backgrounds.

### 5.4. Problems in Development

Various problems in child development arise from environments that fall far short of presenting an optimal fit with children’s developmental trajectories. Most of these emerge in social–cultural–economic deprivation (discussed earlier). Even more serious developmental problems may be encountered in especially difficult

circumstances such as war, family violence, child abuse and neglect, abject poverty involving malnutrition, debilitating and fatal morbidity (e.g., HIV/AIDS), and homelessness (e.g., street children). Under such difficult circumstances, beyond arrests in physical and cognitive/language development, emotional and social development also suffers greatly with possible psychopathology. Beyond medical and other research and interventions, there is a need for research that examines the cultural aspects involved such as societal values, parental views of children's needs, and parental behaviors and priorities for survival. Culturally sensitive assessment of the situation and interventions that are informed by cultural knowledge would be expected to have a greater chance for success. In particular, cultural strengths, such as closely knit human ties and social support mechanisms, could be identified and buttressed further, whereas vulnerabilities and weaknesses could be ameliorated. Thus, especially in the most difficult circumstances, a cultural/contextual approach is valuable. This is an important area of applied developmental psychology.

## 6. CONCLUSION

Child development occurs in culture and has to be studied in cultural context. Such a cultural/ecological perspective is necessary for both scientific and applied work. Understanding, assessing, and possibly enhancing child development requires that both a "whole child" perspective, attending to different spheres of development, and a contextual perspective be used. In so doing, both the unity and diversity of human development across cultures need to be recognized. Finally, developmental scientists should strive to go beyond understanding child development to use this understanding optimally in enhancing human development and well-being within and across cultures.

### See Also the Following Articles

Diverse Cultures, Dealing with Children and Families from

## Further Reading

- Berry, J., Dasen, P. R., & Saraswathi, T. S. (Eds.). (1997). *Handbook of cross-cultural psychology: Basic processes and human development* (Vol. 2). Boston: Allyn & Bacon.
- Bornstein, M. H. (Ed.). (1991). *Cultural approaches to parenting*. Hillsdale, NJ: Lawrence Erlbaum.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U., & Ceci, S. J. (1993). Heredity, environment, and the question "how": A first approximation. In R. Plomin, & G. E. McClearn (Eds.), *Nature-nurture and psychology* (pp. 313-324). Washington, DC: American Psychological Association.
- Ceci, S. J., & Williams, M. W. (Eds.). (1999). *The nature-nurture debate: Essential readings*. Cambridge, MA: Blackwell.
- Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Belknap Press of Harvard University Press.
- Gardiner, H., Mutter, J., & Kosmitzki, C. (1998). *Lives across cultures: Cross-cultural human development*. Boston: Allyn & Bacon.
- Greenfield, P., & Cocking, R. (1994). *Cross-cultural roots of minority child development*. Hillsdale, NJ: Lawrence Erlbaum.
- Kagitcibasi, C. (1996). *Family and human development across cultures: A view from the other side*. Hillsdale, NJ: Lawrence Erlbaum.
- LeVine, R. A., Miller, P., & West, M. (1988). *Parental behavior in diverse societies*. (New Directions for Child Development, No. 40). San Francisco: Jossey-Bass.
- Masten, A. S. (Ed.). (1999). *Cultural processes in child development*. Mahwah, NJ: Lawrence Erlbaum.
- Nsamenang, B. (1992). *Human development in cultural context: A Third World perspective*. Newbury Park, CA: Sage.
- Paris, S. C., & Wellman, H. (Eds.). (1998). *Global perspectives for education: Development, culture, and schooling*. Washington, DC: American Psychological Association.
- Ryan, R. M., Deci, E. L., & Grolnick, W. S. (1995). Autonomy, relatedness, and the self: Their relation to development and psychopathology. In D. Cicchetti, & D. J. Cohen (Eds.), *Developmental psychopathology* (Vol. 1, pp. 618-655). New York: John Wiley.
- Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioral Development*, 9, 545-570.



# Child Testimony

Jodi A. Quas and Brianne E. Beck

University of California, Irvine, California, USA

1. Factors That Influence Children's Testimony
  2. Children's Involvement in the Legal System
  3. Conclusions
- Further Reading

## GLOSSARY

**direct questions** Yes/No or short-answer questions that probe for particular details about an experienced or witnessed event; questions can range from nonleading to suggestive, depending on how the question is phrased and the type of information that is included.

**free recall questions** Vague general prompts requesting a narrative description of prior events; prompts provide minimal cues or information about what should be recalled.

**script** A general knowledge structure that represents information from a class of similar events; it is a description of the typical features of repeated events.

**structured interviews** Scripted interview prompts and questions that are designed to obtain as much narrative information from children and to be as minimally leading as possible.

**suggestibility** Susceptibility to errors when describing witnessed or experienced events; it can be caused by internal factors (e.g., biases) or external factors (e.g., questioning technique).

Since the 1980s, there has been a virtual explosion of studies concerning children's eyewitness abilities and

participation in legal proceedings. The studies, along with considerable public, legal, and legislative attention, were motivated by a number of phenomena. One was a growing awareness of the widespread prevalence of maltreatment, especially child sexual abuse. Large U.S. national surveys revealed that many more individuals were reporting unwanted sexual experiences as children than had been recognized previously, yet few perpetrators were successfully prosecuted criminally. In response, increased efforts were made to identify victims of maltreatment and facilitate the prosecution of perpetrators of crimes against children. A second contributing factor involved revisions in statutes (e.g., competency requirements) that reduced barriers to admitting children's testimony as evidence in court. By reducing such barriers, the number of child victims/witnesses becoming involved in legal cases increased dramatically. Unfortunately, little was known about how to elicit accurate information (e.g., about abuse) from children. Legal professionals and scientists, as well as the public, focused on the consequences of failing to identify a victimized child rather than on the potential for false reports of a crime. A third phenomenon during the 1980s, however, altered the public's and researchers' focus on false allegations. Specifically, several high-profile legal cases that involved scores of children alleging sexual assault in their day care facilities took place. The children's allegations included fantastic and impossible claims and often emerged only after the children had been subjected to repeated, highly suggestive interviews. These children's reports, and the

cases in general, raised serious concerns about the reliability of children's memory abilities. Questions about how to protect children from harm and ensure that those who commit crimes against children are punished, while at the same time guarding against false allegations of abuse, were brought to the forefront of public, legal, and scientific attention. Research on children's memory, suggestibility, and false memories, as well as on children's participation in legal cases, has sought to answer these questions. The present article focuses on this research. First, research concerning factors that influence children's memory accuracy and susceptibility to false suggestions is presented. Second, the effects of participation in legal proceedings on children's well-being are discussed. Third, a brief overview of new directions in the field is provided. The article is limited in scope to findings that are most applicable to legal cases, are particularly well established, and have received considerable attention. It is also limited to discussion of child victims/witnesses. A large and growing body of literature now focuses on child defendants, and it is not possible to review both areas in a single brief article.

## **1. FACTORS THAT INFLUENCE CHILDREN'S TESTIMONY**

Hundreds of studies have now been conducted concerning the reliability of children's eyewitness abilities. A number of factors have been identified that have implications for children's memory and suggestibility. Those discussed here can be heuristically divided into child, to-be-remembered event, and interview characteristics.

### **1.1. Child Characteristics**

Numerous characteristics of children can influence their eyewitness abilities. The most well-studied characteristic is age, and the most consistent finding across research involves age differences in performance. With advancing age, children remember more information, are more accurate, are less suggestible, and make fewer false event reports. Both cognitive and social factors underlie age differences. Cognitively, older children are better able to conduct memory searches, know what is important to recount, and know how to cue their own memories. They also have more experiences and greater general knowledge, and this helps them to encode, interpret, and later remember information. Socially, older

children are more aware that adults can make mistakes, and this reduces their willingness to report false information suggested by adults as having occurred.

Other characteristics, in addition to age, have also been studied as predictors of children's memory and suggestibility. These include cognitive factors (e.g., intelligence, language, inhibitory control) and socioemotional factors (e.g., temperament, emotional reactivity, attachment). With regard to cognitive factors, although findings have not always been consistent, higher intelligence, greater language understanding, and better inhibitory control are generally associated with improved memory and reduced suggestibility, even when children's age is statistically controlled. With regard to socioemotional factors, outgoing temperament, low emotional reactivity, and secure attachment are often related to better memory, especially for stressful experiences.

Although research concerning child characteristics is important theoretically and provides considerable insight into processes underlying accuracies and inaccuracies in children's memory and suggestibility, it is difficult to apply to legal contexts. In particular, it is not possible to use any of the aforementioned characteristics, alone or in combination, as a litmus test to determine whether to accept or reject a particular child's eyewitness account. Nonetheless, the research may help to identify children who need special precautions taken when they are questioned about forensically relevant matters. Armed with information about factors that predict performance in individual children, interviewers may be better able to alter their questioning procedures to ensure that each child is interviewed in the most appropriate manner possible.

### **1.2. To-Be-Remembered Event Characteristics**

Numerous factors associated with a to-be-remembered event may influence children's memory and suggestibility. Three factors that have received considerable attention are the stressfulness of the to-be-remembered event, the number of times that an event occurred, and children's level of involvement.

For some time, researchers have attempted to determine whether stress helps or hurts children's memory. Studies have focused on children's memory for mildly arousing laboratory-based events (e.g., fire alarms) as well as for highly distressing, salient, naturally occurring experiences (e.g., medical procedures, natural disasters). Although results have not revealed any

uniform positive or negative effect of stress on memory, some general conclusions are possible. First, children, like adults, remember emotional events better than they do neutral events. Thus, although the effects of varying levels of distress on memory remain unclear, stressful events appear to be more memorable than nonstressful neutral events. Second, memory for central details of an event (e.g., details directly associated with the cause of the stress) is often better than memory for peripheral details (e.g., information unrelated to the cause of the stress). It should be noted, however, that what is considered central to a child is not necessarily the same as what is considered central to an adult. For instance, a parent's departure during a medical procedure may be especially salient, and hence central, to a child even though the parent's departure is unrelated factually to the medical procedure. Third, the effects of stress on memory vary across children such that for some children (e.g., securely attached children) stress helps memory, whereas for other children (e.g., shy or insecurely attached children) stress inhibits memory. Researchers have begun to investigate interactions between stress and child characteristics as joint predictors of memory and suggestibility. The results of these new studies will continue to elucidate when and how stress affects children's performance.

The number of times an event is experienced can affect children's memory, although the direction of the effect varies depending on the type of to-be-remembered information. When an event is experienced repeatedly, children form scripts, which are mental representations of what "usually" happens or the common recurring features of the event. When asked to recount what happened during a repeated event, children can use these scripts to help them describe what usually happens. Thus, scripts enhance memory for the common repeated components. Unfortunately, children may fail to report nonrepeated details because these details are not included in the script of what usually occurs. Children may also have difficulty discriminating between repeated events and report details that are consistent with what usually occurs, even if those details did not occur during one particular incident.

A third event-relevant characteristic is children's participatory role. Children's memory is better for events in which they participated than for events that they merely witnessed. Active participation often results in more elaborate encoding and enhanced understanding of an event, both of which facilitate recall. Accordingly, children may often be better witnesses about events that they personally experienced than about events that

they only witnessed, although some witnessed events, such as those that are especially relevant personally or are particularly salient or stressful, can also be remembered quite well.

### 1.3. Interview Characteristics

Much of the research concerning children's eyewitness testimony has focused on how characteristics of the interview affect performance. Results have been applied directly to legal contexts. For instance, several criminal convictions have been overturned because the interviews with alleged child victims were conducted in a manner that has been shown, in empirical research, to lead to errors and entirely false event reports. Also, interviewers are now often taught appropriate interviewing tactics and which tactics to avoid when questioning alleged child victims/witnesses.

How questions are phrased can affect the completeness and accuracy of children's reports. Questions are often divided into "free recall" and "direct" questions. Free recall questions are vague general statements prompting a child to recount, in narrative form, a past event (e.g., "Tell me what happened when . . ."). Direct questions include nonleading, leading, and suggestive questions. Nonleading questions do not imply incorrect information or a particular response (e.g., "What color was the car?"). Leading questions imply a particular response (e.g., "Was the car red?"). Suggestive questions contain false embedded information (e.g., "What was the man wearing when he entered the room?" when in fact there was no man) or explicitly imply which answer should be given ("He entered the room first, didn't he?" when the man did not in fact enter first).

Free recall questions are less controversial than direct questions because children's free recall responses are typically quite accurate. That is, children rarely provide incorrect information when asked, in an open-ended manner, to recount a particular event. The major drawback to free recall, however, is that children's reports are often incomplete. Children might not know what is important to recount. They also might have difficulty in forming cogent narrative answers, even when they remember event details.

To obtain additional information, direct questions that probe for specific information are often needed. Such questions indeed increase the amount of information that children provide; however, the questions, especially those that are suggestive, also increase errors. Children may answer in a way that they believe the interviewer desires, they might not be comfortable

indicating that they do not know an answer, they may develop a “yeah”- or “nay”-saying response bias, or they may accept the interviewer’s suggestions as fact. Each of these increases errors, and in some circumstances it may lead children to claim that entirely false events occurred.

Importantly, although children’s free recall reports are generally accurate but incomplete and children’s responses to direct questions are less accurate, two exceptions to these trends must be mentioned. First, accuracy of free recall can be compromised if, prior to asking free recall questions, children are exposed to false suggestive information. For instance, before an interview, a parent or an interviewer might imply to a child that an alleged perpetrator did something bad or a child might learn from other children what happened during an alleged event. These preinterview sources of false suggestions lead to increased errors in children’s responses to both free recall and direct questions. Second, when no abuse occurred, children tend to be more accurate when answering abuse-related direct questions (e.g., “Did he take pictures of you when you were in the bathroom?,” “Did he touch your bottom?” [asked to children who merely experienced a play interaction with an adult]) than when answering non-abuse-related questions. Thus, children’s willingness to err in response to direct questions varies depending on the content of those questions.

Highly leading interview instructions can have a dramatic effect on children’s memory and suggestibility. Such instructions include asking children to imagine or pretend that a particular event occurred or introducing false stereotypic information about an event or individual(s). When these instructions are given before or during an interview, children will often alter their statements about an individual’s activities or change their interpretation of an event and recount information that is consistent with the leading instructions. In addition, when children are explicitly told that an event occurred and are given information about what happened, a sizable minority will repeat the information as having happened even if the event is false. Interestingly, children are less likely to assent that false negative events (e.g., falling off a tricycle and getting stitches) occurred than to assent that false positive events (e.g., going on a hot air balloon ride) occurred, although some children nonetheless assent to both positive and negative events. What remains of theoretical interest is the extent to which children believe that the false events took place versus the extent to which they are simply acquiescing to an interviewer’s suggestions.

Question repetition, both within and across interviews, has been especially controversial. Some claim

that simply repeating questions about untrue information causes false reports in children. However, such a phenomenon has not yet been confirmed empirically. Instead, the effects of repeated questions vary as a function of the type of question, whether questions are repeated within or across interviews, and preinterview instructions. Within an interview, particularly young (e.g., preschool-age) children often change their responses when asked repeated direct (e.g., yes/no) questions. They tend to interpret the repetition as an indication that their first response was incorrect. In contrast, when questions are repeated across interviews, at least when no highly leading instructions have been provided, earlier interviews can actually have a buffering effect, reminding children of an original event and reducing errors in later interviews. Finally, when repeated interviews are combined with highly leading interview instructions such as those described previously, a high percentage of children (often one-third or more) will make false reports that are consistent with the instructions.

#### **1.4. Summary**

A variety of factors have been studied as predictors of accuracies and inaccuracies in children’s memory and eyewitness abilities. Some findings (e.g., those concerning age differences in performance, the accuracy of children’s free recall reports, the deleterious effects of highly suggestive interview questions and instructions) have been consistently found in empirical research and are now met with minimal controversy when applied to legal settings. Other findings (e.g., those involving effects of stress on memory or interview repetition), however, have been more variable and are not amenable to broad straightforward conclusions or applications to legal contexts. That is, whether many factors (e.g., repeated interviews, stress, participation) help or hurt children’s eyewitness abilities depends on a number of considerations, such as how questions are asked and the type of memory being probed. These considerations need to be taken into account when attempting to evaluate child victims’/witnesses’ reports.

## **2. CHILDREN’S INVOLVEMENT IN THE LEGAL SYSTEM**

Although most research concerning children’s testimony has focused on the accuracy of children’s eyewitness

abilities, studies also have investigated the effects of legal involvement on children and the use of innovative practices to facilitate children's participation in legal cases. Some psychologists and legal scholars contend that children suffer "secondary" victimization by participating in a legal system that was not designed for, and can be quite intimidating to, children. Such individuals advocate altering standard courtroom procedures (e.g., allowing children to testify without a defendant present) to reduce children's distress. Other scholars, however, maintain that, because all witnesses can be distressed during a legal case, distress should not be used as the basis for abandoning defendants' fundamental rights or changing standard legal practice.

A few studies have investigated the effects of legal involvement on children. Most noteworthy are studies of the effects on child sexual abuse victims. Findings indicate that some children are adversely affected by legal involvement in terms of increased distress and behavior problems following their legal cases. Factors that increase a child's risk of adverse effects include testifying multiple times, testifying while facing the defendant, not having maternal support during the case, and not having corroborative evidence confirming the abuse allegation. However, not all children appear to be adversely affected, and adverse effects appear more often following criminal involvement than following juvenile or family court involvement. Furthermore, the extent to which adverse effects are maintained over long time periods has only recently begun to be investigated.

As mentioned previously, some advocate altering standard legal procedures to facilitate children's participation in legal cases. A related line of research has investigated the use of these alterations and their effects on children's distress and case outcomes. Innovative techniques include those that affect pre-court interviews with children (e.g., multidisciplinary centers, videotaping interviews) and those that affect in-court practices (e.g., admitting children's out-of-court statements [hearsay] as evidence, allowing children to testify via closed-circuit television testimony). Although there is support for the use of the innovations, and the constitutionality of their use has been upheld by some U.S. courts, their actual use appears to be limited. In particular, surveys of prosecutor and judicial behaviors indicate that techniques that alter in-court practices are rarely used, primarily due to concerns that decisions will be overturned. Techniques that alter how children are interviewed are also rated as too costly for widespread use. Despite the limited use of innovations, however, they appear to be effective

in reducing children's distress. Also, studies of jurors' decision making have revealed relatively few differences in case outcomes based on whether children testify live or children's statements are presented in some other format (e.g., hearsay is presented, children testify via closed-circuit television). Finally, the use of several innovations, such as videotaped testimony and closed-circuit television testimony, is standard in other countries (e.g., Great Britain). Perhaps as awareness of both the benefits of the innovations and the lack of adverse effects on case outcomes grows, the innovations will be used more frequently in the United States as they are in other countries.

### 3. CONCLUSIONS

Numerous characteristics, including those both intrinsic (e.g., age) and extrinsic (e.g., stress of a to-be-remembered event, interviewer demeanor, testifying in court) to children, have implications for children's ability to recount past experiences accurately and for their ability to cope with involvement in the legal system. Although this article has highlighted some of the better established and particularly noteworthy findings and controversies, it is by no means exhaustive. Furthermore, several exciting new directions in research continue to refine understanding of the conditions under which children's testimony is more likely to be accurate and how to facilitate children's participation in legal cases. These new directions focus on how characteristics of children interact with contextual characteristics to affect their memory and suggestibility and on the development and testing of formal interview protocols that can be used in forensic contexts to ensure that children are questioned in the most appropriate manner. With regard to interactions between children and contextual characteristics, it is likely that individual children respond very differently to particular events, such as stressful experiences, and these different responses may affect how individual children later remember the events. With regard to interview training protocols, simply providing forensic interviewers with background knowledge (e.g., about the hazards of leading questions) and general training does not improve how actual interviews are conducted. Thus, researchers have begun to develop specific guidelines and structured protocols that interviewers can use when questioning alleged child victims/witnesses. Although these new protocols, which are more widely accepted in other countries (e.g., Great Britain, Israel), are currently being tested in the United

States, they already appear to be effective in increasing the amount of narrative information provided (e.g., in response to free recall questions) and in helping interviewers to avoid the most problematic questioning tactics. These protocols may also decrease the number of times children need to be questioned, thereby reducing distress that can result from being interviewed multiple times about alleged crimes.

As new psychological research is conducted, and as previously reported findings are confirmed and extended, it will be imperative to continue to educate those involved in actual forensic contexts regarding abilities and limitations of child witnesses. Not only will this knowledge advance scientific understanding of children's memory capabilities and suggestibility, but it will also benefit both alleged victims/witnesses and defendants in legal cases involving children.

### See Also the Following Articles

Eyewitness Testimony ■ Interrogation and Interviewing

### Further Reading

- Ceci, S. J., & Bruck, M. (1998). Children's testimony: Applied and basic issues. In W. Damon, I. E. Sigel, & K. A. Renninger (Eds.), *Handbook of child psychology* (5th ed., Vol. 4, pp. 713–773). New York: John Wiley.
- Ceci, S. J., & Friedman, R. D. (2000). The suggestibility of children: Scientific research and legal implications. *Cornell Law Review*, 86, 33–109.
- Goodman, G. S., Emery, R., & Haugaard, J. (1998). Developmental psychology and the law: Divorce, child maltreatment, foster care, and adoption. In I. Sigel & A. Renninger (Eds.), *Handbook of child psychology*, Vol. 4: *Child psychology in practice* (5th ed.). New York: John Wiley.
- Goodman, G. S., Taub, E., Jones, D. P. H., England, P., Port, L. P., Rudy, L., & Prado, L. (1992). Emotional effects of criminal court testimony on child sexual assault victims. *Monographs of the Society for Research in Child Development*, 57 (Serial No. 229).
- Home Office Communication Directorate. (2001). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children*. London: Author.
- Lyon, T. D. (1999). The new wave of suggestibility research: A critique. *Cornell Law Review*, 84, 1004–1087.
- Pool, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association.
- Quas, J. A., Goodman, G. S., Ghetti, S., & Redlich, A. (2000). Questioning the child witness: What can we conclude from the research thus far? *Trauma, Abuse, and Violence*, 1, 223–249.



# Classical Conditioning

P. Ferrándiz

Catedrática, Universidad Complutense, Madrid, Spain

1. From Pavlov to the Present
  2. Basic Conditioning Procedure
  3. Basic Processes
  4. Experimental Procedures
  5. Excitatory Conditioning
  6. Inhibitory Conditioning
  7. What Influences Classical Conditioning?
  8. Applications
- Further Reading

## GLOSSARY

- behaviorism** The term originated with the work of Watson. Behaviorism was a movement in psychology that argued that mental events were impossible to observe and measure. It was claimed that psychology was only the study of observable behavior. Until the 1960s, this was the dominant approach to learning.
- cognitivism** An approach in psychology that was born in 1956. It began as a reaction to behaviorism. It proposed that human information and storing processes were very important in the process of learning.
- conditioned or conditional response (CR)** Any response that is learned by conditioning.
- conditioned or conditional stimulus (CS)** Any stimulus that, through conditioning, ultimately evokes a conditioned response. The conditioned response is originally a neutral stimulus.
- contiguity** The quality of two events occurring relatively closely in time. Conditioning will only occur if events or stimuli (CS-US) follow each other.
- contingency** The relationship between two events in which one depends on the other. Contingency refers to the

probability that when the CS occurs, the US is likely to follow.

**unconditioned or unconditional response (UR)** Any response elicited spontaneously by an unconditioned stimulus, without any prior learning or conditioning.

**unconditioned or unconditional stimulus (US)** Any stimulus that elicits a response without any previous learning.

The simplest associative mechanism, whereby organisms learn to produce new responses to stimuli and learn about relations between stimuli or events, is classical, or Pavlovian, conditioning. In classical conditioning, a neutral stimulus is repeatedly paired with another stimulus, called an unconditioned stimulus (US), that naturally elicits a certain response, called an unconditioned response (UR). After repeated trials, the neutral stimulus becomes a conditioned stimulus (CS) and evokes the same or a similar response, now called the conditioned response (CR). Most of the concepts presented in this article were developed by Pavlov. However, it also examines how original classical conditioning has been modified. Although the basic conditioning paradigm emerged 100 years ago, its comprehension and conceptualization are still developing.

## 1. FROM PAVLOV TO THE PRESENT

Russian physiologist I. Pavlov (1849–1936) was one of the pioneers in research on classical conditioning. His investigations, carried out at the end of the 19th

century, form the axis of associative learning. Pavlov was interested in the digestive system and, through these studies, he observed that dogs, the subjects of his experiments, anticipated the salivation response when they saw food. Subsequently, Pavlov presented a light or the ticking of a metronome [conditioned stimulus (CS)] for a number of seconds before the delivery of food [unconditioned stimulus (US)]. At first, the animal would show little reaction to the light but, as conditioning progressed, the dog salivated during the CS even when no food was delivered. This response was defined as the conditioned response. Hence, research on classical conditioning began. Soviet psychology has changed in many ways since the death of Pavlov in 1936, and many of these changes brought it closer to American psychology. One way in which the general theory of Pavlovian conditioning developed similarly in the two countries was via the incorporation of cognitive variables. The originally simple Pavlovian paradigm was expanded and reinterpreted through the efforts of many American researchers. American learning theorists developed alternative methods for estimating the strength of the association between a CS and a US. They also identified new Pavlovian phenomena and suggested new explanations of the basic mechanisms underlying the conditioning produced by use of Pavlovian procedures. Although Pavlov's writings were widely disseminated in the United States, most of the criticism of his ideas was expressed in 1950 by J. A. Konorski in Poland. Konorski and Miller, Polish physiologists, began the first cognitive analysis of classical conditioning, the forerunner of the work by R. A. Rescorla and A. R. Wagner (American psychologists) and A. Dickinson and H. J. Mackintosh (English psychologists).

In the 1950s, some clinical psychologists—such as J. Wolpe (in South Africa), H. J. Eysenck (in the United Kingdom), and B. F. Skinner (in the United States)—began to treat psychological disorders with a technology based on learning principles, aiming to substitute pathological behavior for a normal one. This movement, known as behavior modification, was based on both classical and operant conditioning and reinforced the popularity of these techniques among clinicians.

Significant changes on theoretical views upon this phenomenon were introduced by Rescorla and Wagner in 1972, whose model assumed that organisms, in their interaction with the environment, form causal expectations that allow them to predict relations between events. From this perspective, learning would consist

of the acquisition of information about the causal organization of the environment. In order to acquire this information, associations between elements are established. However, it is not sufficient for the elements to be contiguous so that an animal will associate them, as Pavlov proposed; it is also necessary for the elements to provide information about a causal relation. In fact, the central concept is the notion of contingency and the way in which it is represented in the animal's mind. This new "cognitive associationism" accepted the existence of both cognitive processes and associative mechanisms in the mind and rapidly gained wide support among learning experimentalists.

Although in some cases the organism seems to be conditioned only on grounds of CS-US "contiguity," in many other cases learning seems to allow the organism to predict US arrival from the CS appearance; this would be due to a "contingency relationship" that is expressed in terms of probability; if the probability of the appearance of the US is the same in the presence and in the absence of the CS, contingency will be null and no conditioning will occur. On the contrary, if the probability of the US appearing in the presence of the CS is greater than the probability of it appearing without the CS, then conditioning occurs. Therefore, classical conditioning is defined as the capacity of organisms to detect and learn about the predictive relation between signals and important events. Currently, classical conditioning is not considered to be simple and automatic learning about the establishment of CS-US relations but rather a complex process in which numerous responses intervene, not just glandular and visceral responses, as was thought at first. It also depends on the relevance of the CS and the US, the presence of other stimuli during conditioning, how surprising the US may be, and other variables.

Lastly, the growth of connectionist modeling has renewed interest in classical conditioning as a fruitful method for studying associative learning.

## 2. BASIC CONDITIONING PROCEDURE

The procedure of classical conditioning consists of the repeated presentation of two stimuli in temporal contiguity. First, a neutral stimulus (NS) is presented—that is, a stimulus that does not elicit regular responses or responses similar to the unconditioned response (UR). Immediately after that, the US is presented.

NS (a bell) → OR (orientation response)

US (presence of food in the dog's mouth) → UR (salivary response)

CS (a bell) → CR (salivary response)

**FIGURE 1** Diagram of the sequence of events in classical conditioning.

Because of this pairing, the NS will become a CS and, therefore, will be capable of provoking a conditioned response (CR) similar to the UR that, initially, only the US could elicit (Fig. 1).

On the initial trials, only the US will elicit the salivation response. However, as the conditioning trials continue, the dog will begin to salivate as soon as the CS is presented. In salivary conditioning, the CR and the UR are both salivation. However, in many other conditioning situations, the CR is very different from the UR. According to Pavlov, the animals learn the connection between stimulus and response (CS–UR). Currently, it is understood that animals learn the connection between stimuli.

### 3. BASIC PROCESSES

Some basic processes that affect all sorts of conditioning learning have been identified. They include the following:

1. Acquisition: the gradual increase in strength of CR, linking the two stimuli (CS followed by the US).
2. Extinction: the reduction and eventual disappearance of the CR at the CS onset, after repeated presentations of CS without being followed by US.
3. Generalization: when a CR is linked to a CS, the same CR is elicited by other CSs in proportion to their similarity to the original CS.
4. Discrimination: occurs when subjects have been trained to respond in the presence of one stimulus but not in the presence of another.

## 4. EXPERIMENTAL PROCEDURES

### 4.1. Appetitive Conditioning

Appetitive conditioning is a modality in which the US (and CS) will originate a positive (approaching, consummative) behavior in the organism.

### 4.2. Aversive Conditioning

In aversive conditioning, the US and CS have negative properties and will generate an escape or evitation behavior. Aversion to some taste (Garcia effect) showed the possibility of obtaining long aversive reactions to some CS taste that, after being linked to some US with strong aversive digestive qualities, would generate an evitation response after a long interstimuli interval. Learning taste aversion plays an essential role in the selection of food because it prevents the ingestion of toxic foods. It has also helped to understand and treat food aversions in people suffering from cancer and undergoing chemotherapy.

## 5. EXCITATORY CONDITIONING

Excitatory conditioning occurs because the CS allows the subject to predict when the US will occur. In excitatory conditioning, there is a positive relation between the CS and the US. The excitatory stimulus (CS+) becomes a stimulus that elicits a CR after its association with the US. Traditionally, classical conditioning research has focused on this kind of conditioning.

## 6. INHIBITORY CONDITIONING

Inhibitory conditioning occurs because the CS allows the subject to predict that no US will be presented; that is, an inhibitory conditioned stimulus (CS-) signals the absence of an unconditioned stimulus (either positive or aversive), provoking the inhibition of the CR in the organism. This kind of conditioning was studied by Pavlov, and there has been a renewed interest that continues today.

One of the clearest examples of this procedure is the one presented by M. Domjan: A red traffic light (CS+) signals the potential danger (US), which is the car. If a policeman indicates that we can cross the street (CS-) despite the red light, he is signaling that there is no danger (US). Inhibitory stimuli, such as “out of order” or “do not cross” signs, provide useful information.

## 7. WHAT INFLUENCES CLASSICAL CONDITIONING?

The capacity of stimuli to become associated is modified by many phenomena, such as the CS–US interval, spatial

contiguity between stimuli, the similarity of the CS and the US, the intertrial interval, the characteristics of the CS and the US (intensity, novelty, and duration), or previous experience of the CS (latent inhibition), the simultaneous presence of other more salient stimuli (overshadowing), or more informative stimuli (blocking).

### 7.1. CS–US Interval

The interstimulus interval is the time between the appearance of the CS and the appearance of the US. It is one of the critical factors that determine the development of conditioning. Different CS–US interstimulus intervals are usually employed.

Simultaneous conditioning describes the situation in which the CS and the US begin at the same time. Delayed conditioning occurs when the onset of the CS precedes the US onset. Trace conditioning occurs when the CS and US are separated by some interval in which neither of them is present. The US is not presented until after the CS has ended. Backward conditioning occurs when the CS is presented after the US. This procedure casts doubt on the principle of stimulus contiguity because although the CS and the US are presented in inverse order, the association should be equally efficient.

The simultaneous and backward conditioning procedures produce poorer conditioning in comparison to delayed and trace procedures. However, today, due to the use of more sophisticated measurements than those traditionally employed, all the procedures can produce efficient conditioning.

## 8. APPLICATIONS

Classical conditioning is relevant to behavior outside the laboratory in daily life and plays a key role in understanding the origins and treatment of clinical disorders. In humans, classical conditioning can account for such complex phenomena as an individual's emotional reaction to a particular song or perfume based on a past experience with which it is associated; the song or perfume is a CS that elicits a pleasant emotional response because it was associated with a friend in the past. Classical conditioning is also involved in many different types of fears or phobias, which can occur through generalization.

Pioneering work on this type of phenomena was carried out by J. B. Watson and collaborators in the 1920s. They were able to condition a generalized phobia

to animals and a white furry object in a young child, “Albert” (creating a learned neurotic reaction), and also to extinguish a phobia in another child, “Peter,” through a learning procedure (carried out by Mary C. Jones). Both cases have been considered pioneering work in the behavior modification field.

## 8.1. Everyday Human Behavior

### 8.1.1. Classical Conditioning and Emotional Response

Recent reanalyses have evidenced the relevance of individual differences and cognitive variables. People who are more anxious in general develop phobias from experiences that do not produce fear in others.

Fears can be learned by multiple means that include direct Pavlovian conditioning, but the Pavlovian paradigm, with the revisions of the neoconditioning theory, continues to be one of the most influential theories on phobic anxiety. All fears or phobias do not seem to be explainable by simple Pavlovian conditioning, although many may be. Many phobias are reported to be the result of indirect or cognitive learning, including learning by observation and by information transfer. Classical conditioning has also been useful to explain panic attacks.

### 8.1.2. Classical Conditioning and Food Aversion

Food aversions are a serious clinical problem, especially those developed by association with the nausea produced by chemotherapy for cancer, which may later interfere with eating many foods. Applying modern knowledge of the classical conditioning process, treatment usually consists of trying to prevent the aversion by minimizing the predictability relationship of the CS and the US (nausea). Giving the patients a food with a highly salient, novel taste in novel surroundings results in blocking, and the novel cues overshadow the tastes of more familiar foods. The result is a strong aversion to cues that will not be present after chemotherapy is finished. This is helpful and causes no problems because there is little development of conditioned nausea to familiar foods when eaten at home.

## 8.2. Clinical Treatment

Behavior therapy, based on the principles of classical conditioning, has been used to eliminate or replace behavior, to eliminate the emotional responses of fear

and anxiety, and as treatment for nocturnal enuresis, alcoholism, and so on.

### **8.2.1. Systematic Desensitization Therapy**

As behavioral methods developed over time, a behavior therapy technique called systematic desensitization was devised based broadly on the classical conditioning model. Undesirable responses, such as phobic fear reactions, can be counterconditioned by the systematic desensitization technique. This technique inhibits expressions of fear by encouraging clients to face the feared CS and thus allowing extinction to occur. In systematic desensitization, anxiety is associated with a positive response, usually relaxation. Systematic desensitization is a procedure in which the patient is gradually exposed to the phobic object; training in progressive relaxation is an effective and efficient treatment for phobias.

### **8.2.2. Implosive (Flooding) Therapy**

One approach to treating phobias with classical conditioning was originally called implosive therapy (flooding). It is used to extinguish the conditioned fear response by presenting the CS alone, repeatedly, and intensely. The phobic individual experiences the CS, and all the conditioned fear is elicited, but no aversive US follows, nothing bad happens to the subject in the presence of the CS, and so the conditioned fear of the CS disappears.

### **See Also the Following Articles**

Cognitivism

### **Further Reading**

- Chance, P. (2002). *Learning and behavior* (5th ed.). New York: Brooks/Cole.
- Davey, G. (Ed.) (1987). *Cognitive processes and Pavlovian conditioning in humans*. Chichester, UK: Wiley.
- Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge, UK: Cambridge University Press.
- Domjan, M. (2000). *Essentials of conditioning and learning* (2nd ed.). Belmont, CA: Wadsworth/Thomson.
- Domjan, M. (2003). *Principles of learning and behavior* (5th ed.). Belmont, CA: Wadsworth/Thomson.
- Klein, S. B. (2001). *Learning: Principles and applications* (4th ed.). New York: McGraw-Hill.
- Leahey, T. H., & Harris, R. J. (2000). *Learning and cognition* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Leslie, A. M. (2001). Learning: Association or computation? *Current Directions in Psychological Science*, 10(4), 124–127.
- Lieberman, D. A. (2000). *Learning: behavior and cognition* (3rd ed.). Pacific Grove, CA: Wadsworth.
- Mackintosh, N. J. (1983). *Conditioning and associative learning*. Oxford, UK: Clarendon.
- Mackintosh, N. J. (Ed.). (1994). *Animal learning and cognition*. New York: Academic Press.
- Mazur, J. E. (2002). *Learning and behavior* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- O'Donohue, W. (1998). *Learning and behaviour therapy*. Boston: Allyn & Bacon.
- Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.
- Schwartz, B., Wasserman, E. A., & Robbins, S. J. (2002). *Psychology of learning and behavior* (5th ed.). New York: Norton.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory*. New York: Appleton-Century-Crofts.



# Clinical Assessment

Irving B. Weiner

University of South Florida, Tampa, Florida, USA

1. Nature and Purpose of Clinical Assessment
  2. Sources of Information in Clinical Assessment
  3. Issues in Clinical Assessment
  4. Future Directions
- Further Reading

Clinical assessment is a procedure in which various forms of information drawn from interviews, behavioral observations, standardized tests, collateral reports, and historical documents are integrated to facilitate decisions about diagnosis and treatment planning.

## GLOSSARY

- deception* A conscious and deliberate attempt to appear less disturbed and more capable than is actually the case; it is also known as “faking good.”
- idiographic* A way of describing people according to their individual and unique characteristics.
- malingering* A conscious and deliberate attempt to appear more disturbed and less capable than is actually the case; it is also known as “faking bad.”
- nomothetic* A way of categorizing people according to characteristics they share in common with other people.
- performance-based measures* Assessment instruments that sample how people respond when given a task to do.
- self-report inventories* Assessment instruments that sample how people describe themselves when asked to indicate whether, or to what extent, various statements apply to them.
- standardized tests* Formal measuring instruments that consist of a specific set of stimuli and instructions and that are always presented to respondents in the same way.
- structured interviews* Interviews based primarily on a preset schedule of specific questions.
- unstructured interviews* Interviews consisting primarily of open-ended inquiries and comments.

## 1. NATURE AND PURPOSE OF CLINICAL ASSESSMENT

Clinical assessment consists of collecting various kinds of information about the nature of people to facilitate arriving at certain kinds of decisions intended to promote the welfare of persons being assessed. These decisions most commonly address interrelated issues of differential diagnosis and treatment planning, and they typically are reached in consultation with persons who have been assessed concerning what would be in their best interest.

### 1.1. Differential Diagnosis

Treatment of persons with physical or psychological problems proceeds best when the nature and origins of these problems can be identified with reasonable certainty. Adequate differential diagnosis includes both nomothetic and idiographic components. The nomothetic component consists of categorizing the type of disorder a person has and noting the kinds of characteristics the person shares in common with other people who have the same disorder. The idiographic

component consists of specifying the particular life experiences that have given rise to the disorder in this individual case and identifying current circumstances that appear to be perpetuating the disorder.

### 1.2. Treatment Planning

Differential diagnostic information derived from clinical assessments guides treatment planning in four key respects. First, the category of disorder with which a person is diagnosed calls attention to modes of treatment that have been shown to be effective in treating people with the same or a similar category of disorder. Second, the diagnosed severity of a person's disorder indicates whether treatment can be provided adequately on an outpatient basis or instead requires a hospital or other residential setting. Third, diagnostic information about past and present events in a person's life often points to topics and goals on which the treatment sessions should focus. Fourth, in addition to identifying critical treatment topics and goals, diagnostic information can alert therapists to patient characteristics that might slow progress in the treatment, at least until these potential obstacles to progress have been minimized or overcome.

## 2. SOURCES OF INFORMATION IN CLINICAL ASSESSMENT

Clinical assessment is a multifaceted process in which information is drawn from diverse sources. These sources of information include interviews, behavioral observations, standardized tests, collateral reports, and historical documents. The data obtained from these sources sometimes converge, and in such cases the confirmatory information they provide broadens the base of the assessment process and lends certainty to inferences drawn from it.

In other instances, diverse sources of information may diverge, for example, when a man being interviewed describes himself as a laid back person who is easy to live with, whereas his wife (a collateral informant) describes him as an irritable and demanding person who is nearly impossible to live with. Instances of divergent data from different sources often help clinical assessors to sharpen their decision making by identifying why these divergences have occurred and determining which sources have yielded the most reliable and useful information. Ordinarily, prior to collecting assessment data, it is difficult to anticipate which sources of information will converge (and thus be redundant) and

which will diverge (and thus deepen understanding of the person being assessed).

### 2.1. Interviews

Mental health specialists who conduct clinical assessment interviews ask patients questions and invite them to talk about their symptoms and problems, any previous treatment they have received, their family backgrounds, and their developmental, educational, occupational, and social histories. Depending on the preferences of the clinician and the capacities of the respondent, these interviews may be relatively structured or relatively unstructured. Relatively structured assessment interviews are based primarily on a preset schedule of specific questions (e.g., "How many brothers and sisters do you have?," "Do you have trouble sleeping at night?"). Relatively unstructured interviews, in contrast, consist primarily of open-ended inquiries and comments (e.g., "What was growing up like for you?," "You seem a little anxious right now"). In actual practice, most clinical assessment interviewers combine some open-ended inquiries with various specific questions intended to clarify items of information.

### 2.2. Behavioral Observations

Clinical assessors learn a great deal about people from observing how they sit, stand, walk, gesture, show their feelings, and express themselves in words. These observations are made mostly while conducting assessment interviews, but clinicians sometimes have opportunities to observe people in other contexts as well, for example, as patients in a residential treatment setting or as inmates in a correctional facility.

Regardless of their context, behavioral observations can take a variety of forms. For example, clinicians working with families may take note of how family members talk to each other during a family group interview or how they interact during a home visit, each of which is a sample of fairly natural behavior. Alternatively, clinicians may watch and listen to how family members collaborate when they are given some task to do or are asked to decide on a single set of answers to a questionnaire, each of which is an artificially contrived situation.

### 2.3. Standardized Tests

Standardized tests are formal measuring instruments that consist of a specific set of stimuli and instructions and that are always presented to respondents in the same

way. The systematized materials and procedures that characterize standardized tests make it possible to collect normative reference data concerning how certain groups of people (e.g., men, women, children, Americans, Japanese) are likely to respond to these tests on average.

The standardized tests commonly used in clinical assessments are categorized as being either self-report inventories or performance-based measures. Self-report inventories sample how people describe themselves when asked to indicate whether, or to what extent, various statements apply to them (e.g., “I worry quite a bit over possible misfortunes,” “In school, I was sometimes sent to the principal for bad behavior”). Performance-based measures sample how people respond when given a task to do (e.g., putting blocks together to make various designs, describing what a series of inkblots look like). Conjoint use of self-report inventories (what people say about themselves) and performance-based measures (how people go about doing assigned tasks) typically maximizes the scope and utility of the assessment data obtained with standardized testing.

#### **2.4. Collateral Reports**

In addition to interviewing, observing, and testing people being evaluated, clinical assessors may obtain collateral information from other individuals who know them or have observed them. Collateral informants are most commonly family members, but they may also include friends, teachers, and employers. Collateral reports usually enhance the scope and reliability of clinical assessments, especially when the people being assessed are unable or reluctant to talk freely or to cooperate fully with testing procedures.

At times, however, mistaken impressions on the part of collateral informants may detract from the accuracy of their reports, which turn out to be less dependable than what the people being assessed have to say. Informants' reports may also be of questionable reliability if personal issues they have with the people being assessed bias their perspectives. Clinical assessors may also confront instances in which people being evaluated exercise their rights to confidentiality by denying access to informants other than themselves or to historical documents.

#### **2.5. Historical Documents**

Historical documents provide real accounts of events in people's lives for which their own recollections may be vague or distorted. School, medical, military, and criminal records (when available) enhance the dependability

of clinical assessments by documenting life experiences that people may be reluctant to talk about or unable to remember clearly. In addition, when there is concern that people being assessed have suffered some decline in mental or emotional capacity from previously higher levels of functioning, historical documents can provide a valuable baseline for comparison. This is especially the case when the past records include clinical assessments.

### **3. ISSUES IN CLINICAL ASSESSMENT**

Contemporary clinical assessments must typically take account of four issues, especially with respect to collecting and using standardized test data. First, are the test results reliable and valid? Second, could the test results reflect malingering or deception on the part of the respondent? Third, are the test findings applicable in light of the respondent's cultural context? Fourth, can one make appropriate use of computer-based test interpretations?

#### **3.1. Reliability and Validity**

The reliability and validity of test results refer to whether the results are dependable, that is, whether they provide an accurate measurement of whatever they are measuring. The reliability of tests is most commonly estimated by examining the internal consistency of their items or by determining the retest stability of the scores they generate. The validity of tests refers to how much is known about what they do in fact measure, as determined by the extent to which their scores correlate with other phenomena such as a diagnosed condition and a favorable response to a particular form of treatment.

Currently available clinical assessment instruments vary in their demonstrated reliability and validity, and examiners must be cautious about relying on measures that are limited in these respects. On balance, however, research findings indicate that most of the tests commonly used in clinical assessment, when administered and interpreted properly, meet the criteria for acceptable reliability and validity.

#### **3.2. Malingering and Deception**

Malingering among persons being assessed consists of a conscious and deliberate attempt to appear more

disturbed and less capable than is actually the case. Malingering, also known as “faking bad,” commonly emerges as an effort to minimize being held responsible for some misconduct (e.g., being found not guilty of a crime) or to maximize being held deserving of receiving some type of recompense (e.g., being awarded damages for an alleged injury).

Deception, in contrast, consists of attempting to appear less disturbed and more capable than is actually the case. Deception, also known as “faking good,” usually arises as an attempt to gain some benefit (e.g., being employed in a position of responsibility) or to avoid losing some privilege (e.g., being deprived of child custody or visitation rights).

Clinical assessors can draw on various guidelines for identifying malingering and deception on the basis of interview behavior and responses to standardized tests. Based on criteria related to consistency versus inconsistency, these guidelines include specific kinds of interview questions, validity scales on standardized tests, and certain standardized tests designed specifically to reveal efforts to fake bad or fake good. Used in conjunction with data from observational, collateral, and historical sources, these interview and test measures are usually effective in detecting malingering and deception. Nevertheless, research findings indicate that resourceful and motivated respondents, especially when they have been coached with respect to what they should say and how they should conduct themselves, can succeed in misleading even experienced examiners.

### 3.3. Cross-Cultural Contexts

Cross-cultural clinical assessment generally proceeds in two phases. The first phase consists of drawing on sources of assessment information to identify the nature of an individual’s psychological characteristics. The second phase consists of determining the implications of these psychological characteristics within the person’s cultural context. With respect to collecting information, some assessment methods are relatively culture free (e.g., copying a design from memory). Other assessment methods, especially those involving language (e.g., interviews, self-report inventories), are relatively culture bound. To obtain reliable results, particularly with culture-bound methods, assessment data should be collected by an examiner or with a test form in a language with which the person being examined is thoroughly familiar.

Following collection of the data in cross-cultural assessment, the process of interpreting these data should be attuned as much as possible to available normative

standards within groups to which the person being examined belongs. On the other hand, normative cross-cultural differences in assessment findings, such as scores on some test variable, may indicate an actual difference between cultures in the psychological characteristic measured by that variable rather than a need to adjust standards for interpreting the variable. In the case of standardized tests, in fact, many variables identify psychological characteristics in the same way regardless of the person’s cultural context.

The main challenge in doing adequate cross-cultural assessment arises during the second phase of the process, that is, when clinicians must estimate the implications of a person’s identified psychological characteristics for how the person is likely to function in his or her daily life. In some instances, psychological characteristics are likely to have universal implications. For example, being out of touch with reality will ordinarily interfere with adequate adjustment in any social, cultural, ethnic, or national surroundings. In most instances, however, the implications of psychological characteristics are likely to depend on their sociocultural context. For example, whether being a passive, dependent, group-oriented, and self-sacrificing person promotes effective functioning and a sense of well-being will depend on whether the person is part of a society that values these personality characteristics or that instead is embedded in a family, neighborhood, or societal context that values assertiveness, self-reliance, and ambitious striving for personal self-fulfillment.

### 3.4. Computer-Based Test Interpretation

Computer software programs are available for most standardized and widely used clinical assessment instruments. The data entry for these programs consists of the score for each test item or response, and the output typically includes (a) a list of these scores; (b) a set of indexes, summary scores, and profiles calculated on the basis of the test scores; and (c) a narrative report concerning the interpretive significance of these indexes, summary scores, and profiles. This narrative report is the key feature of what is commonly called computer-based test interpretation (CBTI).

Self-report inventories also lend themselves to a computerized test administration in which the test items and the alternative responses to them (e.g., “agree strongly,” “agree,” “disagree,” “disagree strongly”) appear on the screen. Computerized administration of

performance-based measures has not yet been shown to be practicable, however, owing to the fact that the response data are observational, infinite in their variety, and frequently interactive (in the sense that responses may require certain kinds of inquiry to clarify how they should be scored).

Computerized testing programs usually reduce scoring and calculation errors and also save time in recording and arraying the various indexes, summary scores, and profiles that provide the basis for test interpretation. Computerization also facilitates the interpretive process itself by virtue of a computer's complete scanning and perfect memory. A good CBTI program does not overlook any encoded implications the data are believed to have for psychological characteristics of the person being assessed.

On the other hand, CBTI is limited by its unavoidable nomothetic focus. Computer-generated interpretive statements describe features of a test protocol and present inferences that are applicable to people in general who show the same or similar test features. These statements do not necessarily describe the person being assessed, nor do they identify idiographic ways in which the person differs from most people who show certain similar test scores. Consequently, because of human individuality, CBTI narrative reports nearly always include some statements that do not apply to the respondent or that are inconsistent with other statements.

Accordingly, informed and professionally responsible computerized testing requires thorough knowledge of how and why test scores generate certain interpretive statements. Sophisticated clinical application of computer-generated test interpretation also requires appreciation of what a computer-generated report implies in the context of each respondent's personal history and current life circumstances. Unfortunately, CBTI programs can be obtained and used by persons with little or no knowledge of the tests on which they are based, especially in the case of self-report software that conducts the test administration, and with little appreciation of how and why to translate the nomothetic statements in computer-generated reports into the idiographic statements that are essential for an adequate clinical report.

#### 4. FUTURE DIRECTIONS

Research findings have validated a large number and a wide variety of clinical assessment methods as providers of accurate and dependable information about the psychological characteristics of people. This information

has, in turn, been shown to be helpful to mental health professionals in making decisions about differential diagnosis and treatment planning. However, little is yet known about the long-term impact of decisions made on the basis of clinical assessment findings. Are people better off as a result of having been assessed than they would have been otherwise? Do assessment-based decisions enhance the quality of life that people are able to enjoy? These questions set an important agenda for further directions in clinical assessment research that should emphasize follow-up studies of the long-term effectiveness of assessment procedures.

#### See Also the Following Articles

Attitude Measurement ■ Educational and Child Assessment ■ Forensic Mental Health Assessment ■ Intelligence Assessment ■ Measurement and Counseling ■ Mental Measurement and Culture ■ Personality Assessment ■ Psychological Assessment, Overview ■ Psychological Assessment, Standards and Guidelines for ■ Psychometric Tests ■ Psychophysiological Assessment

#### Further Reading

- Beutler, L. E., & Groth-Marnat, G. (2003). *Integrative assessment of adult personality* (2nd ed.). New York: Guilford.
- Butcher, J. N. (Ed.). (2002). *Clinical personality assessment* (2nd ed.). New York: Oxford University Press.
- Butcher, J. N. (2003). Computerized psychological assessment. In J. R. Graham, & J. A. Naglieri (Eds.), *Assessment psychology* (pp. 141–164). Vol. 10 in I. B. Weiner (Ed. in Chief), *Handbook of psychology*. Hoboken, NJ: John Wiley.
- Dana, R. H. (Ed.). (2000). *Handbook of cross-cultural and multicultural personality assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Maruish, M. E. (Ed.). (1999). *The use of psychological testing for treatment planning and outcome assessment* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Maruish, M. E. (2003). Psychological assessment in treatment. In J. R. Graham, & J. A. Naglieri (Eds.), *Assessment psychology* (pp. 119–140). Vol. 10 in I. B. Weiner (Ed. in Chief), *Handbook of psychology*. Hoboken, NJ: John Wiley.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128–165.
- Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford.
- Weiner, I. B. (2003). The assessment process. In J. R. Graham, & J. A. Naglieri (Eds.), *Assessment psychology* (pp. 3–26). Vol. 10 in I. B. Weiner (Ed. in Chief), *Handbook of psychology*. Hoboken, NJ: John Wiley.



# Cognition and Culture

Deanna Messervey and Li-Jun Ji

Queen's University, Kingston, Ontario, Canada

Yukiko Uchida

University of Michigan, Ann Arbor, Michigan, USA

1. Self
  2. Attention and Perception
  3. Reasoning Styles
  4. Categorization
  5. Interpersonal Perception and Memory
  6. Social Inference
  7. Decision Making
  8. Language and Communication Style
  9. Intelligence
  10. Conclusions
- Further Reading

## GLOSSARY

**analytic thinking** The type of thinking that focuses on the object (instead of context) and the categories to which the object belongs. Analytic thinkers rely on rules and formal logic in understanding the world. This type of thinking is more prevalent among Westerners, particularly European North Americans.

**correspondence bias (fundamental attribution error)** The tendency to overemphasize a person's dispositions and to downplay the importance of the situation when explaining behaviors.

**entity theory of intelligence** The belief that intelligence is fixed and unchangeable.

**holistic thinking** The type of thinking that emphasizes context and relationships, and makes relatively little use of categories and formal logic. This way of thinking is prevalent among East Asians.

**incremental theory of intelligence** The belief that intelligence is mutable and can be changed.

**independent self** A type of self-construal, or self-concept, in which the self is perceived as a distinct entity that has a unique combination of traits, abilities, and other inner attributes. The independent self acts in accordance with his or her private thoughts and emotions rather than in a manner that is motivated by the thoughts and emotions of others.

**interdependent self** A type of self-construal in which the self is seen in relation to others, and the self is motivated to create and maintain interpersonal relationships with these relevant others. People with interdependent selves generally place the needs and goals of their in-group above their own.

**perspective taking** The ability to take the point of view of another person.

Culture refers to shared values, beliefs, and norms among a group of people who most often speak the same language and live in proximity. Cognition is defined as the process and products of human perception, learning, memory, and reasoning. Culture affects many aspects of cognition, and cognitive tendencies also guide the propagation of culture. This article provides an introduction to understanding how culture interacts with cognitive processes at basic and social levels from the perspective of cultural psychology. We discuss how culture affects the way that people perceive themselves, categorize information, and make social inferences. Moreover, we address how decision

making, memories, and communication and reasoning styles may be influenced by our cultural backgrounds.

### 1. SELF

In 1991, Markus and Kitayama outlined two types of self-concepts or self-construals: independent and interdependent. An independent self is perceived as a distinct entity who has a unique combination of traits, abilities, and other inner attributes. Moreover, the independent self acts in accordance with his or her private thoughts and emotions rather than in a manner that is motivated by the thoughts and emotions of others. It is worth noting that the study of the self has largely taken a “monocultural approach” that has originated in the West. Because North Americans predominantly have independent self-construals, it is not surprising that North American psychologists have adopted an approach that treats the self as an independent entity. When researchers, however, examined how the self is construed in other cultures, they discovered that independent self-construals were not as prevalent as in North America. Indeed, interdependent selves are found to be quite common among some other cultural groups, such as Asians. People with interdependent selves see themselves in relation to others and are motivated to create and maintain harmonious interpersonal relationships with these relevant others. Although people with interdependent selves hold internal attributes, they also understand the importance of context in shaping these inner attributes. Consequently, they are less likely to act in accordance with their thoughts, emotions, and other attributes, especially when such behavior affects relevant others. Thus, interdependent selves are guided by the way the self is related to others. According to Markus and Kitayama, interdependent selves are predominant in Chinese, Japanese, Hispanic, Filipino, African, and Hindu populations, and independent selves are predominant in Western populations.

The distinction between independent and interdependent selves followed Triandis’s work on individualism and collectivism, which has received tremendous attention in the cross-cultural psychology literature. Individualist selves place their own personal goals above their groups’ goals, whereas collectivists give priority to their groups’ goals over their own. It is worth noting that research suggests that individualism and collectivism are not opposite dimensions. In 2002, Oyserman and colleagues conducted a meta-analysis

examining individualism and collectivism across numerous cultures and found that people who are high in one dimension (e.g., collectivism) are not necessarily low in the other dimension (e.g., individualism). For example, European North Americans tended to be high in individualism yet were not less collectivistic than Japanese or Korean samples.

### 2. ATTENTION AND PERCEPTION

Culture affects how we perceive the world and where we focus our attention. Particularly, it shapes the extent to which we attend to the central object or to the field. In addition, differential attention may affect people’s subsequent cognitive performances. To test cross-cultural differences in perception, researchers have used the rod-and-frame test (RFT) among other tests. RFT is made of a box (also called a frame) and a black line (also called a rod) at the end of the box; the frame and the rod can be turned independently of each other. When European American and Asian participants were asked to make perceptual judgments about the position of the rod, Li-Jun Ji and colleagues found that the position of the frame affected Asians’ perceptions of the rod more than Americans’ perceptions. In other words, European Americans were better than Asians at disassociating the object from the background and hence were more field independent. Conversely, Asians tended to make perceptual judgments based on the relationship between the object and the background and hence were more field dependent. In 1976, Berry argued that cognitive styles (such as field dependence and independence) are related to one’s lifestyle. In particular, he found that sedentary agriculturalists were inclined to have a field-dependent cognitive style, and nomadic hunters and gatherers were more likely to have a field-independent cognitive style.

Similarly, Masuda and Nisbett presented European American and Japanese participants with an ocean scene on a computer screen. They found that Japanese participants paid attention to the background objects more than European Americans did (although Japanese participants did not necessarily pay less attention to the central object than did European Americans). As a result, Japanese participants were able to recall more details about the background than were European Americans. Such differential sensitivity to the background or the context has been found to facilitate performance on certain tasks but debilitate performance on other tasks. When the context remains constant, it

provides a helpful cue for Japanese participants to recognize the central object. When the context changes, however, it becomes difficult for Japanese participants to correctly identify the central object because of a possible cognitive binding between the central object and the background. On the other hand, changes in context do not have as much impact on European American participants, who generally pay little attention to the context anyway. Similarly, when a cognitive task requires disassociation between the object and the background, the task is easier for European North Americans than for East Asians. European North Americans are more likely to focus their attention on the central object and are better at ignoring the context in comparison to many Asians.

### 3. REASONING STYLES

For the past decade, Nisbett and colleagues have conducted a series of empirical studies on reasoning across cultures, which demonstrate that culture shapes people's thinking and reasoning styles in general. Specifically, many Westerners, particularly European North Americans, adopt an analytic system of thought. That is, they focus on the object (instead of context) and the categories to which the object belongs, and they rely on rules and formal logic in understanding the world. In contrast, East Asians tend to think holistically such that they emphasize context and relationships between objects. Moreover, they rely more on intuition and experience and make relatively little use of categories and formal logic compared to Westerners.

In a study conducted by Norenzayan and colleagues, Korean and European American participants were asked to rate the extent to which various arguments were convincing. Participants were presented with either typical arguments (e.g., "All birds have ulnar arteries. Therefore, all eagles have ulnar arteries.") or atypical arguments (e.g., "All birds have ulnar arteries. Therefore, all penguins have ulnar arteries."). Koreans were more inclined to rely on their experiences and to indicate that typical arguments were more convincing than atypical arguments. On the other hand, European Americans, who tend to rely on formal logic and think analytically, saw both arguments as equally convincing. Similarly, when Korean and European American participants were asked to rate the logical validity of valid and invalid arguments that lead to either plausible or implausible conclusions, cultural differences emerged. In particular, Koreans were more likely

than European Americans to report that arguments that had plausible conclusions were more logically valid than arguments that had implausible conclusions.

According to Peng and Nisbett, differences in the reasoning styles between the East and the West also affect the way people think about contradictions. Chinese are much more likely than Westerners to seek compromising solutions when presented with opposing views. For example, European Americans who are told that (i) felines eat fish or (ii) felines do not eat fish often favor one argument at the expense of the other. That is, Westerners often report that one of the options is correct and the other is incorrect. Chinese respondents, however, tend to seek a solution that reconciles the two options. In this manner, Chinese tolerate apparent contradiction in an effort to find the "Middle Way." Similar cultural differences emerge when people attempt to resolve social conflicts. American and Chinese participants were presented with two scenarios that describe everyday conflicts (i.e., a conflict between a mother and daughter and a conflict that pitted focusing on school versus having fun). Peng and Nisbett found that, in general, Americans tended to favor one side of the conflict at the expense of the other (e.g., mothers need to respect their daughters' autonomy more), and Chinese generally favored a compromising solution (e.g., both parties need to better understand each other).

### 4. CATEGORIZATION

In order to make sense of the world that is full of infinite information, we classify and categorize information in a way that can be culturally bound. For instance, Lopez, Atran, Coley, Medin, and Smith found that European Americans and Itzaj categorize mammals in ways that are similar in some situations yet different in others. The Itzaj are Mayans who live in the Peten rain forest region in Guatemala. They spend considerable time interacting with their environment, such as engaging in hunting activities or agriculture. In some regards, both European Americans and Itzaj have developed similar classification systems for mammals. For example, both cultural groups classify opossums in the same category as porcupines, even though opossums are marsupials and are in a class that is completely distinct from porcupines. In other respects, Itzaj are more likely than European Americans to classify mammals on the basis of ecological properties, such as behavior and habitat. For example, Itzaj classify

animals that dwell in trees together, whereas European Americans do not.

Deeper cultural differences emerge when people are asked to classify a variety of information. For example, European American participants tend to category things based on their category memberships, such as grouping “banana” and “orange” together because they are both fruits. Chinese children and Chinese University students tend to group things based on their functional relationships with each other, such as grouping “banana” and “monkey” together because monkeys eat bananas. According to Mishra, this latter type of grouping is also present in other cultures, such as the Birhor, Asur, and Oraon of India. Among European North Americans, however, this latter form of categorization is only found among children and is considered less sophisticated.

When categorizing information, East Asians rely more on intuitive experience, whereas European Americans rely more on formal rules. Generally, European Americans tend to assign objects to categories based on properties that fit the rule. For them, understanding, describing, and explaining objects involve classifying and seeking fundamental physical causes. Conversely, East Asians see an object as it relates to its context and the relationships among various objects. Because Asians historically did not generally perceive and classify objects as being distinct entities, they sought practical knowledge rather than developed formal rules or models.

## **5. INTERPERSONAL PERCEPTION AND MEMORY**

Given that other people are a significant part of Asians’ self-definitions, Asians also tend to pay more attention to other people’s behaviors and emotions. They are expected to remember other people’s preferences, read other people’s minds, and anticipate the needs of others. This finding is related to the fact that Asians tend to be less direct than North Americans in expressing themselves.

Accordingly, European North Americans and Asians also differ in the way they remember information. Indeed, East Asians pay more attention to and remember more behaviors of their friends than do North Americans. When asked to recall and describe situations where they provided support to a friend, Japanese spend more time discussing the situations of their

friends than their own actions. Conversely, European Americans describe their own actions more than their friends’ situations. Furthermore, Japanese accounts of helping behavior tend to correspond more closely to their friends account compared to European Americans. Interestingly, there was little or no consensus between American participants and their friends regarding the nature of support that is exchanged in their relationship.

A special case of interpersonal perception is perspective taking. Perspective taking refers to the ability to take the point of view of another person. In a conversation, East Asians (such as Japanese and Chinese) sometimes identify themselves from another person’s perspective. For example, a Japanese mother might call herself “Oka-san” (mother) in conversation with her child. The self is experienced through the other’s perspective and his or her role in that relationship (e.g., mother–child). In other words, Asians tend to experience their own actions from the points of view of others. Similar differences have also been found in memory research. Research by Qi Wang has found that East Asians’ memories tend to focus on other people or groups, whereas European North Americans’ autobiographical memories tend to be self-focused and vivid. In one study, Asian North Americans were more likely to remember past episodes from a third-person’s point of view, in which their memories included themselves. In contrast, European North Americans remembered the past episodes as they originally saw them; thus, the episodes they remembered tended not to include themselves. In addition to focusing on differential information when reporting autobiographical memories, European Americans tend to recount more emotional memories than do Chinese.

## **6. SOCIAL INFERENCE**

Cultural differences are also found in social inference. Research suggests that Asians emphasize the context more than Westerners when explaining social events. For example, Asians tend to attribute the cause of behaviors to situational factors, whereas North Americans tend to attribute the same behaviors to personal factors. A well-known finding in North American culture is that people often fail to take contextual information into account when they judge the causes of others’ behaviors. This phenomenon is called the correspondence bias, which is also known as the fundamental attribution error. In the classic paradigm,

participants are shown one of two types of essays: One essay expresses a favorable attitude toward a particular issue (e.g., capital punishment), and the other essay expresses an unfavorable attitude toward the issue. Participants are informed that the person who wrote the essay was able to choose which side of the issue he or she would like to discuss (i.e., choice condition) or was instructed to endorse a particular point of view (i.e., no choice condition). Next, the experimenter asks the participants to indicate the true attitude of the person who wrote the essay. Surprisingly, European Americans tend to judge the target person's attitude as corresponding to the view expressed in the essay, even when they know the target person did not have a choice. Consistent with an analytic thinking style, European Americans pay attention to the focal object (i.e., the essay writer) and erroneously believe that the person is independent from the situation. In contrast, Asians are more likely to recognize the situational constraints on the person's behavior when such constraints are made salient.

## 7. DECISION MAKING

Historically, there has been a dearth of research on how culture influences decision making. Recently, however, researchers have begun to give this area more attention. Much of this burgeoning research has focused on differences between East Asians and European Americans.

East Asians and European Americans tend to make decisions using different information. In one study, Incheol Choi and colleagues provided Korean and European American respondents with vignettes describing a target's prosocial and deviant behaviors. Participants were also given a list indicating possible reasons that might or might not explain the causes of the target's behaviors. Koreans considered generally more information than Westerners when asked to explain the target's behaviors. Likewise, other researchers have found that East Asians are more inclined to use historical information when making judgments about the present compared to Westerners. These cultural differences have been attributed to the different underlying theories people hold, in which East Asians have more complex and holistic theories than Westerners. In this manner, East Asians believe that many factors can interact to cause a particular outcome, whereas Westerners tend to believe that fewer factors can lead to a given outcome.

Culture can also affect the extent to which decision makers are willing to take risks. For example, Chinese are more likely to take financial risks than European Americans, although Chinese are more risk averse when they make medical or academic decisions. One possible explanation for these domain-specific patterns of risk taking is that Chinese have greater social ties (including family and friends) than European Americans, which can provide a "cushion" for any negative consequences associated with risk taking in financial domains.

Culture can also influence how confident an individual feels about his or her decisions or judgments. Typically, North Americans tend to make overly confident judgments. For example, after a general knowledge test, Americans are likely to judge that they are correct more often than they actually are correct. Yates and colleagues also found such overconfidence among Chinese, even though they did not find that Japanese made overly confident judgments. There are also different values associated with decision styles in different cultures. For example, decisiveness is not valued in Japan as much as it is in the United States or in China. These findings suggest that there are differences even between East Asian cultures, which call for further research.

## 8. LANGUAGE AND COMMUNICATION STYLE

Language is deeply interconnected with culture. On the one hand, language carries cultural meanings and values. On the other hand, language also transforms culture. Language functions as a tool that enables an individual to focus on certain aspects of the world. Research has found that Japanese pay greater attention to contextual factors during communication compared to European North Americans. For example, Japanese and Chinese do not use the first-person pronoun in some situations. When a teacher is talking to her students, she calls herself "teacher," and when she is talking to her child she refers to herself as "mother." In many English-speaking Western countries, however, the language focuses on the self, which promotes the idea that people are separate entities that are independent from their context. For example, in English people refer to themselves as "I" in almost every context. Moreover, English speakers usually do not drop personal pronouns when they are speaking.

Researchers have examined how language activates cultural identity. For example, Mike Ross and colleagues found that Chinese students in Canada reported interdependent self-perceptions when responding in the Chinese language. Interestingly, when asked to report their self-perceptions in English, they responded in a more westernized way. That is, they tended to see themselves as being more independent than interdependent. Likewise, Li-Jun Ji and colleagues found that Chinese bilinguals who learned English at a relatively late age categorized objects in a more relational (and Chinese) way when tested in Chinese than when tested in English. Thus, it appears that bilinguals' self-perceptions and cognition are closely related to language.

What is the role of language cross-culturally? According to Kim, talking is connected to thinking more in Western culture than in East Asian culture. When Asian Americans and Westerners were asked to think-aloud while solving reasoning problems, talking impaired Asian Americans' performances but enhanced the performances of European Americans. In the West, articulating one's thoughts and conveying such information to others enables the speaker to organize his or her thoughts. In contrast, verbal self-presentation is not encouraged among many Asians; thus, articulation of one's ideas does not occur naturally and actually hinders the reasoning process.

## 9. INTELLIGENCE

Jean Piaget had a considerable impact on the field of cognitive psychology and on intelligence research. He was the first psychologist to empirically measure the cognitive development of children over time. In this manner, he examined how the experiences of children are qualitatively different than adults. He advocated that the cognitive development takes place in four main stages that are universal regardless of culture (i.e., sensorimotor intelligence, preoperational, concrete operations, and formal operations). For instance, in the sensorimotor intelligence stage, children acquire mental representations, learn to imitate, and discover that objects continue to exist even when they are not in perceptual view. Much of Piaget's research has been criticized for a number of reasons. First, many sociocultural researchers do not support Piaget's universalist approach to cognition and maintain that development can take many routes rather than a single direction. Second, researchers contend that Piaget's notion of stages overlooks contextual factors that

shape an individual's development. Third, some researchers have argued that Piaget's stages are ethnocentric, in which people from cultures that do not develop in accordance with Piaget's stages are perceived as lacking in intelligence. Thus, these researchers see the theory as lacking generalizability. Although Piaget's research has played an important role in intelligence and cognitive development research, it appears that his theory is limited.

The history of defining and measuring intelligence has largely taken place in the West. When researchers began examining whether intelligence as defined by Western science could be applied to other cultures, they realized that other cultural groups often had different notions of what constituted intelligence. For example, the Trobriand Islanders may regard superior navigational skills as a key characteristic of an intelligent man. The !Kung of the Kalahari desert may define intelligence as the ability to spot sources of water. In China, intelligence may be viewed as knowing when and how to use language properly, such as knowing when (and when not) to speak. For Hindus, transcendental knowledge may be viewed as highly as practical knowledge. It appears that cultures define intelligence differently and that these differences stem from contrasting ecological and social experiences.

Not only can culture lead to diverse definitions of intelligence but also culture can influence the way individuals develop and maintain implicit theories concerning the malleability of intelligence. Dweck, Hong, Chiu, and colleagues argue that people hold either an "incremental theory" of intelligence or an "entity theory" of intelligence. Entity theorists believe that intelligence is fixed and unchangeable. For instance, European Americans tend to hold an entity theory. Consequently, they do not generally try to improve their performance on intelligence tasks when they have performed poorly because they generally believe that increased effort will have little effect on improving future performance. On the other hand, incremental theorists believe that intelligence is mutable and can be changed. East Asians tend to hold an incremental theory toward intelligence. Accordingly, they try to improve their performance on intelligence tests because they believe that increased effort leads to better performance.

Different definitions of intelligence make it difficult to develop a test of intelligence that can be used in all cultures. In addition, experience with Western cultures can sometimes have a considerable impact on test performance because most available intelligence tests

are developed in the West. For example, Australian Aboriginals who have more interactions with people from European backgrounds tend to perform better on standardized intelligence tests than Aboriginals from more isolated communities. In addition, some research suggests that performance on cognitive tasks is related to familiarity with testing materials. Children in Zambia often build intricate cars from wire, whereas children from England often draw pictures using pencils and crayons. Thus, Zambian children are more familiar with the medium of wire and English children are more familiar with paper-and-pencil mediums. Serpell found that Zambian children performed better than English children on a task that involved assembling a pattern using wire. Conversely, English children performed better than Zambian children on tasks that involved paper and pencils.

## 10. CONCLUSIONS

Research on culture and cognition has enriched our understanding of human behavior and has improved the quality of psychological theories by including people from non-Western cultures. This flourishing area of research has many great implications. Culture fosters people to think in different ways, and these differences may lead to misunderstandings and hostilities, from the long-lasting clashes between Jews and Arabs to the recent conflict between the U.S.-led coalition and Iraq. A world that appreciates and understands people from diverse backgrounds may help resolve existing and future conflicts.

Knowledge of culture and cognition may also help to smooth interactions at interpersonal levels. With increased globalization, the world becomes smaller and cross-cultural interactions become more frequent. Understanding how your colleagues from other cultures may view the world and reason differently than you do will help you to enhance your multicultural partnerships and collaborations. Examples of failing to recognize cultural differences in cognition can be found in business. For example, in a joint-venture company, making internal attributions for success by a

European American could be perceived as arrogant by a Chinese, whereas making external attributions for failure by a Chinese might be perceived as irresponsible by the European American. Ignorance of each other's thinking styles is likely to lead to negative interpersonal experiences and could even have negative consequences on the business outcome.

Understanding cultural differences in cognition can also play an important role in education. For example, teachers who deal with students from different cultural backgrounds would benefit from knowing more about how culture influences cognition. Some teachers may encourage students to verbalize their thoughts as a means of helping them develop their ideas. Although this approach to teaching may benefit Western students, students from other cultural backgrounds may actually be hindered by such an approach.

Understanding the relationship between culture and cognition enhances our understanding of the world, and we can benefit from such knowledge in different aspects of our lives.

### See Also the Following Articles

Attention ■ Cross-Cultural Psychology, Overview ■ Decision Making ■ Intelligence and Culture ■ Interpersonal Perception

### Further Reading

- Berry, J. W., Dasen, P. R., & Saraswathi, T. S. (Eds.). (1997). *Handbook of cross-cultural psychology* (2nd ed., Vol. 2). Boston: Allyn & Bacon.
- Berry, J. W., Irvine, S. H., & Hunt, E. B. (Eds.). (1988). *Indigenous cognition: Functioning in cultural context*. Dordrecht, The Netherlands: Martinus Nijhoff.
- Lonner, W. J., & Malpass, R. S. (Eds.). (1994). *Psychology and culture*. Boston: Allyn & Bacon.
- Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253.
- Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York: Free Press.
- Triandis, H. C. (1995). *Individualism & collectivism*. Boulder, CO: Westview.



# Cognitive Aging

Douglas H. Powell

Harvard Medical School, Boston, Massachusetts, USA

1. Introduction
  2. Cognitive Changes across the Life Span
  3. Age-Related Variability
  4. Evidence for Optimal Cognitive Aging
  5. Correlates of Optimal Cognitive Aging
  6. Looking Ahead
- Further Reading

## GLOSSARY

**cognition** (or “*intellectual functioning*”) Global intelligence or a mental capacity that consists of several specific abilities; an IQ (for intelligence quotient) test such as versions of the Wechsler Adult Intelligence Scale (WAIS) are examples.

**cross-sectional studies** A type of research that studies age-associated changes in different individuals who are grouped by age; for example, the decline in cognitive functions might be measured over the life span by comparing the scores of individuals in age groups of 20, 40, 60, and 80 years.

**crystallized abilities** (or “*verbal*” or “*left hemisphere*” aptitudes) Aptitudes that are largely dependent on acquired knowledge and formal education and that develop rapidly during the first 20 years of life; vocabulary is an example.

**dedifferentiation** The relative decline in specific aptitudes from decade to decade that is nested within a greater decline in global intellectual functioning in normal older adults.

**fluid abilities** (or “*performance*” or “*right hemisphere*” aptitudes) Aptitudes that are largely independent of acquired knowledge and education and that are influenced

by biological factors and the integrity of the central nervous system; block design is an example.

**immune system** The body’s complex and multidimensional defense against invading bacteria, viruses, other pathogenic microorganisms, and tumors; it also assists in the repair of damaged tissue after injury and creates specific antibodies to combat diseases.

**longitudinal studies** A type of research that studies age-related changes in the same participants who are reexamined several times over their lifetimes.

**MicroCog** A computerized test of intellectual functioning that consists of 13 subtests measuring both crystallized and fluid abilities; it was administered to 1002 physicians (ages 25–92 years) and 561 normal respondents.

**neuroimaging** Imaging technology that enables investigators to observe how the brain functions when it carries out specific intellectual tasks; single photon emission computed tomography (SPECT), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI) are examples.

**optimal cognitive agers (OCAs)** Those older individuals whose cognitive test scores are statistically similar to those in midlife.

**psychoneuroimmunology** The study of how the central nervous system is influenced by psychological and social factors.

**Public Law 99-592** The law that eliminated age-based mandatory retirement for workers in the United States; it became effective January 1, 1994, and excludes only a small number of occupations (e.g., Federal Bureau of Investigation agents, airline pilots).

**variability** (or “*variance*” or “*standard deviation*”) How widely dispersed, from top to bottom, the individual scores within an age group may be.

Interest in cognitive aging has grown substantially during the past 25 years. For instance, the divisions of the American Psychological Association (APA) expanding most rapidly during this period have been those with a strong interest in aging such as aging and neuropsychology, a field that examines cognitive impairments associated with aging (e.g., Alzheimer's disease). Responding to the growing interest in Alzheimer's disease and other conditions afflicting numbers of older people, the U.S. Congress designated the 1990s as the "Decade of the Brain." It might be said that this interest applies especially to "older" brains. Funds are being invested at record high levels to increase the understanding and amelioration of the effects of age on intellectual vigor as well as on physical health. These efforts may be more important today than ever before. The reasons for this are demographic, legislative, physical, and economic.

## 1. INTRODUCTION

### 1.1. Demographic Trends

America is graying. At the start of the new millennium, one of every seven U.S. citizens was over 65 years of age. By 2025, this proportion will be one of every five. The group of those age 85 years or over, currently numbering in excess of 3 million, is the fastest-growing segment of the population. This increase in the numbers of elders is not a uniquely American phenomenon. By 2025, those age 65 years or over in Japan will double. The People's Republic of China, Korea, and Malaysia anticipate a tripling of this age group. There is little benefit in a longer lifetime if the quality of health and mental acuity are greatly diminished during these added years. The budget of the National Institute on Aging has tripled during the past decade. Substantial portions of that agency's funds are earmarked for research to reduce the incidence of Alzheimer's disease and other forms of cognitive impairment and to enhance intellectual vitality during the later years.

### 1.2. Legislative Changes

Compelling workers to retire due to age was banned in the United States by Public Law 99-592 as of January 1, 1994. With the exception of a small number of occupations, such as Federal Bureau of Investigation agents and airline pilots, this law eliminated age-based mandatory retirement for nearly everyone. A primary reason for

passage of this law was that many studies had found that a large number of older individuals can match the physical and mental vigor of their younger colleagues. Because variability increases with age, it is necessary to differentiate between functional age and chronological age among seniors. Currently, organizations and professions are struggling with the implementation of objective and fair methods for retaining their older physically and cognitively vigorous employees while creating scenarios for encouraging the retirement of those whose abilities have declined to the point that their job performance is compromised.

### 1.3. Increased Physical and Cognitive Vigor

Despite the growing trend toward obesity in the United States, the physical and cognitive vigor of a large proportion of 60-, 70-, and even 80-year-olds today is superior to that of their parents and grandparents of the same age generations ago. Research confirms these observations. A 1997 report by the National Academy of Sciences showed a dramatic increase in the proportion of Americans over 65 years of age who are able to care for themselves. Reasons offered to account for these gains include reductions in drinking and smoking, better diets, weight loss, control of blood pressure, and use of aspirin to reduce heart attacks and strokes. Also, modern medicine has been chipping away at physical conditions that greatly affect cognition, including medications for diabetes, hypertension, and heart problems as well as advances in cardiac surgery and neurosurgery. Better physical health is strongly correlated with higher levels of intellectual functioning among older individuals.

### 1.4. Economic Needs

One unanticipated benefit of Public Law 99-592 is that the American economy will need its aging "baby-boomers"—those born during the 1946–1964 era—to work well past what used to be the normal retirement age into the first quarter of the 21st century. The reason is that the labor pool is shrinking due to the "baby-bust" generation—those born during the 1965–1977 era—now entering the workforce. Because their yearly birth pools were much smaller than those during the previous 20 years of the baby boom, the labor force will experience a shortfall in those age groups that organizations have traditionally relied on to form the backbone of their workforces, that is, those 30 to 44 years of age.

By 2009, there will be 10.6% fewer men and women in that age group than there were a decade earlier. Census projections estimate that there will not be as many people in that age group as there were in 2000 until 2025. A source of workers that could fill this projected labor shortage are men and women in their 60s and early 70s. Fully 70% of the baby boomers, now at midlife, say that they want to stay on their jobs after reaching 65 years of age. A large proportion of these individuals are mentally and physically fit to continue working if opportunities are made available.

## 2. COGNITIVE CHANGES ACROSS THE LIFE SPAN

What happens to cognitive abilities over the life span? Which aptitudes are impaired most dramatically, and which are most likely to be spared? How much variability is there among people of a given age as they grow older? This section examines global changes in cognition through the life cycle and trace the age-associated pattern of decline among particular aptitudes, paying close attention to the degree to which some are impaired or spared. It also discusses evidence in support of increasing differentiation of cognitive functions with advancing age.

### 2.1. Overall Cognitive Changes through the Life Span

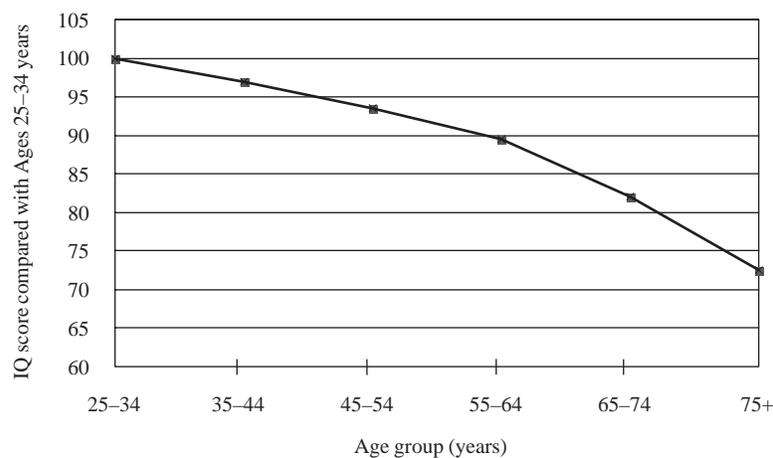
Figure 1 shows the decline in Full Scale IQ from 30 to 75-plus years of age on an earlier version of the Wechsler

Adult Intelligence Scale (WAIS). The curve is derived by comparing the average of the scaled scores for each older age group decade with the norms for 25- to 34-year-olds. Overall, cognitive scores slope gently downward until 60 years of age. At that point, the level of intellectual functioning of the average 60-year-old is only 10% less than that of the average 30-year-old. But then the cognitive decline accelerates. During the next decade, the overall IQ drops 9%. The scores of those age 75 years or over are 11% lower than those of 70-year-olds.

A limitation of these data is that the WAIS respondents were not matched for characteristics (other than age) that affect cognitive changes over the life span. These other characteristics are educational level, social class background, and access to health care. Those who have more education, are wealthier, and are able to avail themselves of adequate medical services usually score higher on cognitive tests than do their less fortunate age mates. Studies that have rigorously matched age groups on these characteristics have found that the shape of the curves of overall cognitive decline is similar but that the test scores are higher at every age for those with more advantages.

### 2.2. Decline of Specific Aptitudes through the Life Span

Although overall intellectual functioning declines predictably from 30 years of age onward, the downward progression of individual aptitudes varies greatly. For instance, the various versions of the WAIS all have found that the Verbal IQ scores slope downward far more gradually than do the Performance IQ scores. For



**FIGURE 1** Average WAIS Full Scale IQ scores for older age groups compared with 25- to 34-year-olds. Adapted from Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.

instance, scores on the Verbal IQ decline 10% from 30 to 70 years of age, whereas scores on the Performance IQ drop 25% during this same period.

The types of abilities in the WAIS Verbal and Performance sections explain why there is so much difference in the rate of decline. The Verbal IQ score is based largely on crystallized abilities. In contrast, the Performance IQ subtests are composed of largely fluid aptitudes. Research by Salthouse and colleagues has demonstrated that fluid aptitudes, such as processing speed and working memory, are greatly impaired by aging. Tests involving processing speed typically require solving complex problems under time pressure. Being required to learn a number–symbol code and then to enter the proper number next to a list of symbols is an example. Working memory involves storing an important piece of information while working on another task. Backward number recall is an example.

Age differences in processing speed and working memory apply much more to complex tasks than to simpler ones. For instance, little difference occurs between younger and older adults on analogies of low and medium difficulty under time pressure. Where the performance of older test takers falls off, relative to that of young adults, is on the hardest of the analogies (e.g., “Fission is to splitting as fusion is to \_\_\_\_\_”) under stressful time limits. The differences are shown in Fig. 2.

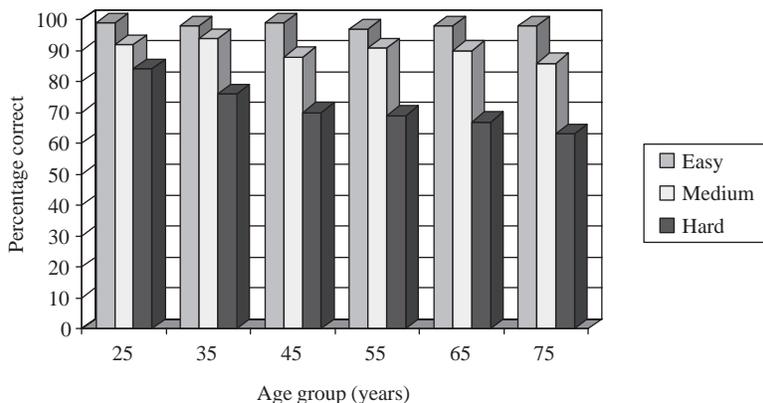
Research reports summarized by Heaton and colleagues in 1981 show other aptitudes whose scores decline earlier and more steeply with increasing age. Timed spatial recall, reasoning, and verbal memory after delay are examples. In contrast, attention, word knowledge, and calculation skills remain relatively stable and are largely spared by the aging process.

All age groups have more difficulty with tasks requiring dual-task or divided attention. The increased incidence of automobile accidents among drivers using cell phones is an example. As people age, their dual-task performance declines rapidly. Over the past three decades, scientific studies have consistently reported that younger adults outperform older adults on many different tests of dual-task attention, for example, classifying a list of words appearing on a computer screen as verbs or nouns while simultaneously listening to a string of numbers and pushing a button when hearing two odd numbers in a row.

Although there is clear evidence of substantial disparity in the rate of decline of specific aptitudes, a number of studies have reported data showing dedifferentiation of cognitive functions with advancing age. That is, in normal older adults, the decline in specific aptitudes from decade to decade is embedded within a greater decline in global intellectual functioning. The evidence is stronger in cross-sectional data and among test scores based on accuracy alone. Longitudinal studies and test results combining speed with accuracy show weaker but similar patterns. Explanations for this phenomenon vary. One theory is that dedifferentiation is simply a symptom of cognitive decline. Another view, based on imaging studies, is that older adults compensate for diminished skills by drawing resources from other areas of the brain that are not normally activated during mental tasks.

### 2.3. Alzheimer’s Disease

If the 1990s was indeed the Decade of the Brain, it could also be said that the 1980s was the decade when Alzheimer’s disease (AD) and other forms of dementia began to attract public attention as potential



**FIGURE 2** Effect of time pressure on analogy scores by difficulty level. Adapted from Salthouse, T. A. (1992). Why adult age differences increase with task complexity. *Developmental Psychology*, 28, 905–919.

threats to those entering the decades beyond middle age. The Alzheimer's Disease and Related Disorders Association was founded in 1980. In that same year, according to Lexis/Nexis (an electronic database for media publications), major U.S. newspapers mentioned AD or dementia twice. In the year 2000, there were 250 separate pieces about AD or dementia.

AD was first described by a German psychiatrist, Alois Alzheimer, in 1907. His first case was a 50-year-old woman whose memory and other cognitive functions deteriorated rapidly over a 5-year period before she died. An autopsy found abnormal brain structures that are referred to as neurofibrillary tangles, and neuritic plaques. As this case demonstrates, the relatively young onset differentiates AD from senile dementia. Today, health care specialists distinguish between presenile AD (prior to 65 years of age) and senile AD (age 65 years or over).

The criteria for AD include a gradual decline in at least two intellectual functions (e.g., attention, memory, language, reasoning, spatial ability) that are sufficient to significantly impair social relationships and/or performance at work. This decline cannot be explained by head trauma, neurological conditions, ill health, severe psychiatric symptoms, or drug/alcohol abuse. The cognitive deterioration also may be associated with depression, emotional outbursts, and/or apathy. It is not unusual to find AD occurring in individuals who otherwise are in good health.

At the start of the new millennium, there were an estimated 5 million people in the United States with AD or other forms of dementia. Prevalence estimates vary, however, and depend on the criteria used for making the diagnosis. For instance, a research team in England rated 100 elderly, community-dwelling volunteers using seven different sets of diagnostic criteria for AD. They found that between 3 and 63% of this population would be judged as "having" AD, depending on which standards were applied. A survey of the incidence of AD among older people in seven countries found proportions ranging from 1.9 to 52.7% of the population. One explanation for the disparate findings on the frequency of AD is the differences in the way in which the condition was assessed. Diagnoses that used tests that included measures of both crystallized and fluid abilities produced more consistent estimates. Applying these test-based standards, AD appears to afflict approximately 3% of 65- to 69-year-olds, 6% of 70- to 74-year-olds, and 11% of 75- to 79-year-olds. Beyond 80 years of age, the proportion of AD rises sharply.

### 3. AGE-RELATED VARIABILITY

Studies of age-related changes in cognitive functions have concentrated on what are called "measures of central tendency," which are most often the mean or average scores. Figures 1 and 2 are illustrations of this approach. Typically, the mean scores of 30-year-olds are compared with the averages of groups of those in older decades. The downward slope of these age group means is then reported. What has been overlooked by focusing on averages alone is what happens to the variability of the individual scores with each advancing decade. The increasing variability with age found in the cross-sectional study of the cognitive aging of 1002 physicians using a computerized test, MicroCog, is an example.

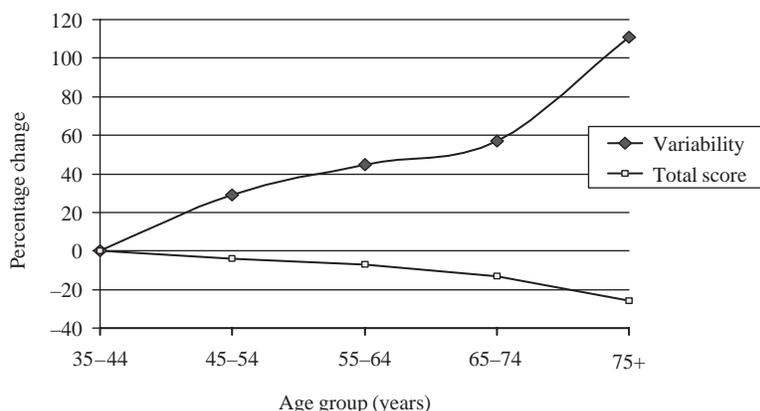
Figure 3 shows the percentage of decline in MicroCog total score in each age group after 40 years. This is compared with the percentage growth of the variability for each decade during this same period. It can be seen that the standard deviations of the age groups rise much more rapidly than the mean scores decline. At 70 years of age, the average physician's MicroCog total score was 13% lower than that of the typical 40-year-old. But the variability of the scores for these 70-year-olds was 57% greater than that for their younger colleagues.

Growing variability within advancing age groups is not specific to these data. Summaries of other studies of age-associated changes in cognition have reported that four of every five studies found that variance within each age group increased with advancing age. These findings occurred equally in both cross-sectional and longitudinal research.

The practical value of being aware of this growing age group variability is that it indicates that a number of older individuals continue to function as well as those in their prime. These older individuals might be called optimal cognitive agers (OCAs).

### 4. EVIDENCE FOR OPTIMAL COGNITIVE AGING

Newspapers and magazines often feature stories of elders who have been productive and successful until very late in life. Novelists James Michener and Barbara Cartland were still writing into their 90s. Heart transplant pioneer Michael DeBakey, South African political leader Nelson Mandela, musician Lionel Hampton, and humanitarian Mother Teresa worked productively in their 80s. Grandma Moses did not begin painting until



**FIGURE 3** Percentage changes in MicroCog total score and variability from 40 years of age: 1002 Physicians. Reprinted by permission of the publishers from *Profiles in Cognitive Aging* by Douglas H. Powell in collaboration with Dean Whitla, Cambridge, MA: Harvard University Press, Copyright © 1994 by the President and Fellows of Harvard College.

she was 78 years old and continued well beyond her 100th birthday.

Social scientists are sometimes skeptical of such examples, wondering whether these people are merely statistical anomalies rather than representative of the vast majority of their age mates. However, empirical evidence now confirms the presence of large numbers of elders who continue to function cognitively at a level similar to those in midlife. This group of individuals, the OCAs, can be distinguished from their contemporaries by applying statistical standards.

Consider the following example. The MicroCog study described earlier obtained cognitive test scores on a large number of physicians from 25 to 92 years of age. Suppose that we wanted to know what proportion of older physicians were OCAs. We could use those doctors in their prime (say, 45–54 years of age) as a comparison group and then set a statistical threshold for optimal cognitive aging. For example, we might choose a cutting score of the 10th percentile, approximately  $-1.3$  standard deviation, of those 45 to 54 years of age. In other words, for doctors to qualify as OCAs by this standard, their scores on MicroCog must be higher than the bottom 10% of those in midlife. The percentages of physicians age 65 years or over who meet this statistical standard are shown in Table I.

These data confirm what informal observation has told us: A large number of people continue to function at a high level cognitively well beyond 65 years of age. Nearly three-quarters of the 65- to 69-year-old physicians, and half of the 70- to 74-year-old physicians, scored as well on this cognitive test as did those in their prime. One-quarter (25%) of 75- to 79-year-old physicians, and

**TABLE I**  
Proportions of Older Physicians with MicoCog Scores above the 10th Percentile of 45- to 54-Year-Old Physicians

Age group	Percentage above 10th percentile
65–69 years	73.2
70–74 years	50.0
75–79 years	25.0
80+ years	10.4

Source. Adapted from Powell, D. H., in collaboration with Whitla, D. K. (1994). *Profiles in cognitive aging*. Cambridge, MA: Harvard University Press.

slightly more than 10% of those beyond 80 years of age, performed as well as the reference group. These data are not specific to physicians. Similar proportions were found among nonphysician respondents in this study.

Although there are advantages to using respondents matched for education, social class, and access to health care for these calculations, nearly any cognitive test data set that reports individual test scores by age group through the life span can be analyzed to yield the percentage of older participants who are OCAs.

### 5. CORRELATES OF OPTIMAL COGNITIVE AGING

What distinguishes these OCAs from their contemporaries whose cognitive decline follows the average downward trend for their age group? Are there behaviors

within individuals' control that are correlated with optimal cognitive aging? If so, identifying these factors may contribute to behavioral changes that can improve the quality of these later years for many people.

But first, another question must be asked: What are the relative contributions of heredity and environment to the quality of cognitive aging? Research finds that heredity does play a strong role in the acquisition, maintenance, and decline of most human qualities. It is also true that environmental forces exert a strong effect. Swedish scientists tested 80-year-old identical twins who had been separated at birth and raised in different home environments. The researchers estimated that approximately 60% of the variance (i.e., factors contributing to the similarity of cognitive ability) was genetic and that the rest was related to environmental influences. Reports from dozens of other twin studies have estimated that heredity accounts for 50 to 60% of what goes into mental ability. So, it seems that somewhere between 40 and 50% of the factors influencing the quality of cognitive aging is related to environmental influences, that is, factors that are partly within the control of each individual.

This section summarizes those behaviors that are correlated with optimal cognitive aging and that are within individuals' control to varying degrees. These include activities that have been demonstrated to be directly related to higher levels of cognition in older adults. Also, this section describes a number of activities that indirectly benefit cognition because they are associated with better physical health. Finally, the section discusses selective optimization with compensation as a strategy for minimizing the inevitable negative effects of cognitive aging.

### **5.1. Activities That Directly Benefit Cognition**

Activities that directly benefit cognition include regular exercise, mental enrichment, active social networks, and cognitive training. Of all the activities that could be recommended to maintain cognitive vitality, regular exercise has the greatest positive effect. This is because exercise strengthens the cardiovascular system, which maintains necessary blood flow to the brain. Cardiovascular disease is a major cause of cognitive impairment among individuals of any age. Meta-analyses of research over the past 40 years have found that younger and older individuals working out for 30 minutes or more every other day for at least 3 months perform at a higher

level on tests of attention, memory, spatial ability, processing speed, and reasoning than do their sedentary neighbors. Of interest is that those individuals whose regimen combines strength/flexibility and aerobic training obtain slightly higher cognitive scores than do those who engage in aerobic or strength/flexibility exercise alone.

Physical and mental vigor can be improved until quite late in life. An illustration of the power of these programs was strength training for 90-year-olds at the Hebrew Rehabilitation Center for the Aged in Boston. After just 8 weeks of regular workouts, these nonagenarians increased their muscle strength by 174% and increased their walking speed by 48%. Although pre- and postcognitive tests were not given, it is likely that improved intellectual vitality accompanied the physical gains. Baylor Medical School researchers found that physically active women and men in their late 60s had greater cerebral blood flow that resulted in higher cognitive test scores than those of their inactive contemporaries.

Mental enrichment stimulates cognitive functions. K. Warner Schaie, the director of the Seattle Longitudinal Study, and his coworkers identified factors that they believe are correlated with maintaining high levels of mental vigor well into later life. These include being open to new and stimulating experiences and having close relationships with individuals who have strong intellectual abilities. Scientific support for this thinking has come from the laboratory of Marian Diamond at the University of California, Berkeley. She was among the first to discover that laboratory animals that were raised in a more complex and challenging environment, along with other animals, developed a thicker frontal cortex than did those that lived in normal but unstimulating cages. She discovered that this occurred in both young and middle-aged animals. The increased thickness was because the existing brain cells in the cortex produced more connections (called dendrite branching) in response to outside stimulation.

A research team at the University of California, Los Angeles, confirmed these findings with humans. The researchers looked at the brains of 20 adults at autopsy. Then, they correlated the size of the nerve cells in the brains' cortexes with the activities reported by these individuals prior to death. They found that the cortexes of those older people who were more mentally and socially active were thicker than those of individuals who were not.

Social support has long been known to be correlated with physical health, but until recently little attention had been paid to its effect on cognition. Scientists from the MacArthur Studies of Successful

Aging studied the impact of social networks on a battery of aptitude tests given to 70- to 79-year-olds. They found that those who had more emotional support obtained higher test scores. At follow-up testing nearly 8 years later, higher retest aptitude scores were moderately correlated with having had greater social support. Other research has found that having larger and more diverse social networks and having more frequent contact with friends and family are related to less cognitive decline.

The question of whether age-related cognitive decline can be reversed through training has raised the interests of specialists in gerontology for the past quarter-century or so. Overall, it appears that most normally functioning individuals in the “young-old” years (approximately 60–75 years of age) and many of those who are older have the capacity to enhance their mental capabilities through practice. One example is the improvement made by participants in the Seattle Longitudinal Study who received 5 hours of training when they were 67, 74, and 81 years old. On average, the two younger groups improved their reasoning and spatial abilities to the levels of their test scores of 14 years earlier. Octogenarians were able to improve their performance to the level of 7 years earlier.

Follow-up studies of cognitive training have found that only those mental skills selected for remediation improved. Those not targeted for tutorial work remained the same as when the instruction began. Thus, if the goal of the program was to better remember a list of words, the participants did not wind up also exhibiting stronger calculation skills at the end of training.

Surveys of the published reports on cognitive training have found that many different methods produce positive results. Older participants at the Max Planck Institute in Berlin who spent 5 hours studying self-help books on improving fluid (spatial) abilities performed as well on tests as did those who had 5 hours of training. These gains were maintained on tests 6 months later. Similar results have been achieved with a wide range of other interventions, with some of them being quite indirect. For instance, members of groups that practiced transcendental meditation regularly improved their scores on cognitive tests far more than did those who did not.

So far, the effects of the training are largely untested in real life. After completing a course on memory improvement, is a woman better able to recall a list of errands she must run? After cognitive training, can a man more easily remember where he left his car in a supermarket parking lot?

## 5.2. Activities That Indirectly Benefit Cognition

Intellectual functions are indirectly influenced by a number of activities that enhance physical health. In an otherwise reasonably healthy aging population, these include behaviors that benefit cardiovascular functions and strengthen the immune system.

In addition to regular exercise, two crucial contributors to cardiovascular health are weight control and blood pressure management. Obesity is a growing problem in the United States. Depending on which criteria are applied, the proportion of overweight adults in the United States currently varies from 37 to 67%. By any standard, the percentage of obese adults has nearly doubled during the past 40 years. Excess weight is associated with greater risk for diabetes, hypertension, and cardiovascular disease—three conditions that adversely affect cognition. Obesity increases with age, partly because fewer calories are required to support normal activity with advancing years and partly because many older adults are physically inactive. In the year 2000, 35% of American older adults reported that they had engaged in no physical activity during the previous month.

Controlling blood pressure reduces the risk of cardiovascular disease and stroke. Participants in the Framingham Longitudinal Study with untreated hypertension scored lower on cognitive tests than did those with normal blood pressure. Increased diastolic blood pressure had particularly negative consequences for intelligence. The intellectual consequences of high blood pressure are far more serious for younger people. Hypertensives under 40 years of age had far lower scores on cognitive tests of attention, memory, mental flexibility, and reasoning than did age-matched normal individuals. Beginning to control high blood pressure early has substantial cognitive and physical advantages.

Physical vitality and cognitive vitality are greatly affected by the efficiency of the immune system. The immune system declines with age. Just as with age-related cognition, however, considerable variability occurs in immune system functions among older people. Until the past decade, it was assumed that the immune system functioned independently of psychosocial influences. Over the past 10 years, however, evidence has accumulated that the immune system is significantly influenced by mental states. This has led to increasing scientific interest in what is called psychoneuroimmunology. Of particular interest is the influence of stress on the suppression of the immune

system given that this, in turn, can adversely affect health and cognition.

How much stress individuals confront, how well this stress is managed, and what other resources individuals can draw on to ameliorate the impact of negative events and bring comfort into their lives all are found to be correlated with health status. For example, more than 400 adult volunteers in Pittsburgh, Pennsylvania, were given nose drops containing a common cold virus. Not surprisingly, a large number of the volunteers developed severe colds soon afterward. But not everyone developed colds, and the severity levels of the colds varied widely. These scientists also measured the total amount of stress that these individuals had to deal with during the past year. Then, they compared these data with the susceptibility to, and the severity of, the colds. They found that individuals with more stress in their lives developed colds far more often, and with more severity, than did those whose stress scores were below average.

Elevated stress-related emotions of anxiety, depression, and anger correlate with lower scores on mental tests. This may be because high levels of stress diminish intellectual resources. An illustration is the finding that neuronal loss in the hippocampus area of the brain was moderately correlated with the number of months in combat in the war in Vietnam. The negative effect of stress emotions appears to be greatest on fluid abilities. For example, higher levels of anxiety have been demonstrated to be associated with lower scores on aptitudes such as analogies, picture memory, and block design.

### 5.3. Selective Optimization with Compensation

One of the most promising strategies for minimizing the impact of aging on cognitive performance is selective optimization with compensation. Developed by Baltes and coworkers, the essential ingredients in this approach are broken down into three steps. The first step is selection, that is, giving up some activities to concentrate the available mental energies on those remaining activities. The activities that continue to have priority are usually selected due to skills, motivations, and opportunities. An aging physician who wishes to remain active might be an example. In midlife, the physician has a busy practice and also teaches medical students. He simultaneously does clinical research, serves on professional committees and boards, and does consulting work. By his early 60s, the physician may begin to focus his mental energies on those patients and types of

activities that he likes best, cutting back on those professional responsibilities in which his interest is waning. So, the physician keeps working at his practice but takes no new patients, and he continues his teaching and research but gives up committee work and consulting. At 70 years of age, the physician may limit his professional activities to teaching and research.

The next ingredient is optimization. This means doing everything possible to maximize one's performance at those things that one chooses to do by anticipating the potential negative effects of cognitive aging. For example, the physician may begin to limit his practice to those patients who he believes he can treat most effectively, referring the others to colleagues. Keeping up with the recent developments in medicine and taking short courses to maintain and upgrade his medical skills become priority items. The physician may try to work "smarter," that is, using computerized expert analyses and digital libraries more readily as opposed to trying to keep all of this information in his head. Optimization requires more time for preparation. For an older physician, this might entail carefully reviewing the chart of a patient in advance of an appointment rather than counting on his ability to get a picture of the medical history by a quick scan while seeing the patient. The aging physician also begins to dictate notes right after seeing the patient instead of waiting until the end of the day or later in the week to do so. In his 60s, the physician must accept the reality that far more time needs to be set aside to prepare lectures than was the case a decade earlier.

Compensation is doing everything that can be done to bolster fading abilities. Eyeglasses and hearing aids, to see and listen as well as possible, are examples from the physical standpoint. From the cognitive standpoint, compensation can include doing one task at a time and recognizing the effect of circadian rhythm (day-night cycle) and glucose/caffeine on cognition. The still working older physician compensates by avoiding dual-task activities such as making notes in a patient's chart while talking about another patient on the telephone. Cognitive compensation might include putting to use an understanding of his personal circadian cycle. The average older person's mental functions are stronger in the morning than later in the day, so the physician may schedule his most difficult patients in the morning and schedule less challenging professional activities in the afternoon. Finally, the physician may compensate for diminished mental acuity in the late afternoon with a cup of coffee and a cookie. For most healthy older individuals, moderate amounts of caffeine and glucose enhance attention and memory for a short time.

## 6. LOOKING AHEAD

Judging from the remarkable progress and growth of interest in all aspects of aging during the past quarter-century, there is little doubt that this will continue to be an area of expanding scientific and academic interest, as well as an area of considerable opportunity, during the decades ahead.

Several relatively recent developments make cognitive aging an especially interesting subspecialty within the field of aging. For instance, developments in neuroimaging technology over the past decade or so have enabled investigators to see in real time how the brain functions when carrying out specific intellectual tasks, to visualize the differences between the brain structures of younger individuals and those of older individuals, and to observe how the brain is compromised by conditions such as AD. Neuroimaging techniques have made possible a quantum leap in understanding the relationship between the brain and behavior. A new discipline, called cognitive neuroscience, draws deeply from this technology. At the time of this writing, the field is attracting the interest of some of the most influential thinkers in the social sciences and medicine.

Another important development has been the availability of findings from longitudinal studies of human development through much of the life span. Examples of these longitudinal studies include the Oakland/Berkeley Growth and Guidance Studies, the Terman Study of Children With High Ability, the Grant Study of Adult Development, the Seattle Longitudinal Study, and the Duke Longitudinal Studies. Some of these were begun during the 1920s and 1930s, whereas others started during the third quarter of the 20th century. Although many of these studies have limited generalizability because their participant pools are not representative of the U.S. population in terms of gender, education, social class, and race, they still provide an increased understanding of the aging human body and mind. As the participants from these investigations are moving into their later years, this research has begun to bear fruit. The numerous books and articles published about these studies have contributed much to the understanding of cognitive aging.

But these publications do not nearly exhaust the opportunities to learn much more about cognitive aging. The raw data from many of these longitudinal studies—the actual responses from each participant every time he or she was examined—are currently archived and are available for analysis. For instance, the Henry A. Murray Research Center for the Study of Lives

at Harvard University contains nearly 300 complete data sets from these longitudinal studies as well as other cross-sectional research that is better balanced for demographic variables. The rapid growth in data storage technology and computer memory and speed makes it possible today to access the raw data from these aging studies and to examine them using a computer in the privacy of a distant laboratory, professorial office, or student's room. Today's scholars can examine this vast storehouse of information, posing new questions, formulating new hypotheses, and looking for new answers in the responses of participants studied generations earlier.

The verity of any research finding is strengthened when it is demonstrated across studies, across generations, across geographical regions, and across investigators. Sophisticated statistical packages and advances in meta-analyses make it possible to compare findings from several studies at once. For instance, a student working on his or her thesis may want to know whether the findings from cross-sectional studies about the significant correlation between more social support and less cognitive decline also could be confirmed by looking at these same relationships in longitudinal studies of earlier epochs. The archived raw data can be easily loaded into the student's computer for analysis of this new question.

Finally, because substantial scientific interest in cognitive aging is relatively new, much interesting work begs attention. Following are two examples. First, what are the characteristics of people who benefit most from cognitive training? What are the mediating effects of variables such as education, motivation, temperament, and degree of impairment? Do people with more education and higher levels of motivation, who are more extraverted and less intellectually impaired, benefit more from cognitive training? Or, is it just the opposite—do those people with less education and lower levels of motivation, and who are less extraverted and more intellectually impaired, make greater gains following training?

The second example is Public Law 99-592. Its passage solved one problem but exposed another. What standards will replace the arbitrary age limits of 55, 60, or 65 years to determine when someone is no longer able to do the job? Because of the currently shrinking labor pool, valid and fair procedures are urgently needed to determine whether an older person continues to have the skills necessary to perform competently in an occupation. So far, no techniques have been developed that have the necessary sensitivity to identify only those individuals with true cognitive

impairment without falsely identifying others who, in fact, do not suffer from cognitive impairment.

Knowledge about cognitive aging has accrued rapidly during the past 25 years or so. More is on the way. As these new findings accumulate, they raise the probability that all individuals, through their behaviors in small increments, will be able to positively influence a significant portion of the cognitive quality of their later lives.

### See Also the Following Articles

Aging and Competency ■ Aging, Cognition, and Medication Adherence ■ Anxiety Disorders in Late Life ■ Cognitive and Behavioral Interventions for Persons with Dementia ■ Dementia in Older Adults ■ Depression in Late Life ■ Elder Caregiving ■ End of Life Issues ■ Personality and Emotion in Late Life ■ Psychoneuroimmunology ■ Psychotherapy in Older Adults

### Further Reading

- Baltes, P. B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as a foundation of developmental theory. *American Psychologist*, 52, 366–380.
- Birren, J. E., & Schaie, K. W. (Eds.). (2001). *Handbook of the psychology of aging* (5th ed.). San Diego: Academic Press.
- Cohen, S. (1996). Psychological stress, immunity, and upper respiratory infections. *Current Directions in Psychological Science*, 5, 86–90.
- Craik, F. I. M., & Salthouse, T. A. (Eds.). *The handbook of aging and cognition*. Hillsdale, NJ: Lawrence Erlbaum.
- Damasio, A. R. (1994). *Descartes' error: Reason, emotion, and the human brain*. New York: Avon Books.
- Diamond, M. C. (1988). *Enriching heredity: The impact of the environment on the anatomy of the brain*. New York: Free Press.
- Glaser, R., & Kiecolt-Glaser, J. (Eds.). (1994). *Handbook of human stress and immunity*. San Diego: Academic Press.
- Heaton, R. K., Grant, I., & Damasio, A. R. (1981). Normative observations on neuropsychological test performances in old age. In I. Grant, & K. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 100–120). New York: Oxford University Press.
- Kosslyn, S. M., & Koenig, O. (1992). *Wet mind: The new cognitive neuroscience*. New York: Free Press.
- Lexak, M. D. (1983). *Neuropsychological assessment*. New York: Oxford University Press.
- McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Pettrill, S. A., & Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins eighty or more years old. *Science*, 276, 1560–1563.
- Poon, L. W., Rubin, D. C., & Wilson, B. A. (Eds.). (1989). *Everyday cognition in adulthood and later life*. New York: Cambridge University Press.
- Rowe, J. W., & Kahn, R. L. (1987). Human aging: Usual and successful. *Science*, 237, 143–149.
- Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Lawrence Erlbaum.
- Salthouse, T. A. (1992). *Mechanisms of age–cognition relations in adulthood*. Hillsdale, NJ: Lawrence Erlbaum.



# Cognitive and Behavioral Interventions for Persons with Dementia

Jiska Cohen-Mansfield

Research Institute on Aging, Hebrew Home of Greater Washington, Rockville, Maryland, USA

1. Introduction
  2. Purpose of Interventions
  3. Theoretical Frameworks
  4. Types of Interventions
  5. Efficacy, Effectiveness, and Utility
- Further Reading

## GLOSSARY

**affect** A person's mood or state of mind such as depression or happiness.

**agitation** Inappropriate verbal, vocal, or motor activity that is not judged by an outside observer to result directly from the needs or confusion of the individual; these behaviors have been labeled problem behaviors, disruptive behaviors, challenging behaviors, disturbing behaviors, behavioral problems, and agitation, all terms that are used interchangeably.

**functional status—ADL** The ability of a person to complete activities of daily living (ADL), that is, activities required for personal self-maintenance; some of these activities include bathing, dressing, grooming, feeding, and toileting.

**self-affirmation** The enhancement of the sense of selfhood/personhood; it can be achieved through things such as meaningful activities, social interactions, and successful accomplishment of objectives.

Cognitive and behavioral interventions for persons with dementia include a large range of tools used to enhance cognitive, behavioral, and affective functioning as well as self-maintenance. Such interventions should follow a detailed analysis of the problems, goals, and abilities of these persons. The interventions are described under the following categories: cognitive, behavioral, self-affirming, sensory, social contact, structured activities, environmental, medical/nursing care, and staff/caregiver training and management. Although these interventions have shown promise in prior research, the field of investigation is still in the early stages, and larger controlled intervention studies are needed. Furthermore, research is needed to clarify the process of intervention and which intervention ingredients are crucial for success. These studies would lead the way to effectiveness and cost analyses.

## 1. INTRODUCTION

The scope of cognitive and behavioral interventions for persons with dementia can be conceptualized from a narrow or a wide perspective. The narrow perspective includes only interventions traditionally referred to as

cognitive behavioral, including reinforcement techniques and cognitive restructuring. A wider perspective includes a range of techniques based on different theoretical frameworks and disciplines, many of which are used either to prevent or to substitute for use of psychoactive medication. This article uses the latter approach, with the goal of clarifying the wide range of techniques available to clinicians in enhancing the quality of life of persons with dementia.

The various intervention techniques that are used with persons with dementia can be organized along several dimensions: the function of the intervention, the theoretical framework serving as a basis for the intervention, the type of activity undertaken during the intervention, and the population subgroups for whom a technique is appropriate. Interventions should be ranked by their effectiveness within a subgroup for certain goals with specific outcome criteria. Some of these dimensions are reviewed briefly in what follows.

## 2. PURPOSE OF INTERVENTIONS

Cognitive and behavioral interventions are used for improving functions in many domains, including the following:

- Cognitive goals include enhancement of memory as measured by memory tests, enhancement of memory for daily tasks, enhancement of ability to communicate, an increase in task engagement, an increase in orientation to time and place, and a decrease in overall cognitive decline.
- Functional goals usually involve an increase in independence or in participation in the performance of activities of daily living (ADL) (e.g., dressing, eating, toileting) and a decrease in risk of falling.
- Affective goals include a decrease in negative affect (e.g., depression, anxiety) and an increase in positive affect (e.g., pleasure, contentment).
- Behavior change involves a decrease in problem behaviors (agitation) such as aggressive behaviors, repetitive nonaggressive and restless behaviors, and repetitive verbal/vocal behaviors. Such behaviors are frequently indicative of discontent in persons with dementia and may also be disruptive to caregivers.
- In terms of self-affirmation, a loss of sense of self and of identity is a common characteristic of dementia. Even when the inner sense of self may remain intact, the socially defined self is often diminished by the decline in social, work, and leisure roles. Therefore,

some interventions target enhancement of the sense of selfhood/personhood in dementia.

## 3. THEORETICAL FRAMEWORKS

Many theoretical frameworks underlie cognitive and behavioral interventions. Some of the basic concepts used in these frameworks are described in what follows.

### 3.1. Cognitive

Research on cognitive processes differentiates between explicit and implicit memory. Explicit memory requires the ability to produce the memorized material and usually involves conscious knowledge of the material. In contrast, implicit memory affects behavior but does not involve awareness of the memory. Explicit memory is usually lost earlier in the dementia process than is implicit memory. Procedural memory involving everyday activities frequently involves implicit memory. Therefore, some simple procedures can be learned during the early and middle stages of dementia even when there is no awareness of the learning process.

### 3.2. Behavioral Theory

Behavioral theory asserts that behavior manifestation is affected by its antecedents and consequences. Environmental stimuli can become associated with a behavior and trigger it, whereas consequences such as reinforcement can affect the likelihood of the behavior's manifestation. To change a behavior, the contingencies between a behavior and its consequences need to be changed, for example, withdrawal of reinforcement or introduction of reinforcement for different behaviors under different circumstances. Alternatively, an intervention can alter the environmental stimulus that triggers the behavior.

### 3.3. Cognitive and Behavioral Theories of Depressed Affect

Cognitive and behavioral theories of depressed affect include (a) cognitive theory that ascertains that depression is a consequence of negative and distorted cognitions about oneself (based on work by Beck and Ellis), (b) theory that claims that depression involves learned helplessness and a sense of loss of control (based on Seligman's work), and (c) theory that depression relates

to an insufficient level of reinforcements or pleasurable experiences (based on Lewinsohn's work).

### 3.4. Unmet Needs Theory

The unmet needs theory asserts that people with dementia manifest inappropriate behavior due to unmet needs. The theory claims that people with dementia have the same needs as do other people. These needs, which are generally summarized under Maslow's hierarchy of needs, include physiological, safety, love and belonging, esteem, and self-actualization needs, with the latter sometimes being manifested as a need for cognitive stimulation or for meaningful activity. According to this theory, these needs frequently go unrecognized by caregivers, and even when they are recognized, these needs are often not satisfied because caregivers do not understand their significance or do not know how to address them. The unmet needs result in inappropriate behavior in one of several mechanisms. First, the inappropriate behavior may be a call for help due to the unmet need. Second, the inappropriate behavior may be an attempt by the person with dementia to respond to his or her need (e.g., self-stimulation as a response to sensory deprivation). Finally, the inappropriate behavior may be a direct result of the unmet need (e.g., screaming due to pain).

### 3.5. Humanistic Theory

Humanistic theory maintains that a person's subjective understanding of his or her experiences is more important than objective reality.

### 3.6. Stages of Human Development

Erikson described eight stages of human development culminating in the eighth stage (age 65 years or over), described as "Integrity versus Despair," in which a person looks back at accomplishments and ahead to the unknowns of death and feels either fulfilled or saddened. Feil added a ninth stage to Erikson's developmental framework that she termed "resolution versus vegetation." It is viewed as an opportunity to resolve conflict and achieve a sense of integrity before death.

## 4. TYPES OF INTERVENTIONS

This section includes a descriptive list of interventions used with people with dementia. In understanding the

interventions on this list, several underlying principles need to be clarified.

### 4.1. General Issues in Considering Types of Interventions

- An intervention should follow a thorough assessment such as a functional analysis that examines the nature of the behavior as well as possible causes, antecedents, and consequences. The assessment should also include information about past stresses, identity, habits, and preferences as well as current medical and psychosocial status. This assessment would guide the understanding of the etiology of presented symptoms as well as realistic goals and options for treatment. Once an intervention is implemented, it should be evaluated. If it does not meet the goals, a different treatment alternative should be taken. Such alternatives might be a different intervention or might involve a different aspect of the intervention (e.g., different timing, differing amounts [dosage], different presentation style, different therapist).
- The list of interventions is not exhaustive but provides examples of the currently available interventions. Interventions are in the process of being developed and tested, so the list is an evolving body of knowledge.
- Interventions often have multiple purposes, and there are multiple options for interventions in any functional domain (see Table I). For example, cognitive interventions can be used for improving cognition, function, behavior, and/or affect.
- The descriptions in what follows provide only a general concept of the various interventions. However, each type of intervention can be considered as a whole range of possible interventions and, therefore, can be delivered in many different ways. It is the specific operational details that may make any treatment effective.
- The manner in which an intervention is presented can be crucial for its success.
- An important issue in understanding the utility of interventions is discerning the active ingredient of the interventions. Two interventions might seem quite different, yet the potent factor in both may be the opportunity to interact with a supportive person. The understanding of the underlying differences among interventions should be a focus for future research.
- Matching interventions to the person and the purpose is important. Cognitive and behavioral interventions cannot be applied uniformly to all persons with dementia. Interventions are appropriate for specific

**TABLE I**  
**Nonpharmacologic Interventions: Examples of Type of Intervention by Goal**

<i>Goal/type</i>	<i>Treatment modalities</i>	<i>Environmental</i>
Enhance cognitive performance	Cognitive remediation/memory training Spaced retrieval Reality orientation Cognitive tasks, such as reading group activities or sorting activities	Memory books (Bourgeois et al., 2001; Brooker & Duce, 2000) Use of signs and pictures to enhance orientation
Enhance functional performance	Independence promoting strategies (Beck, 1998; Namazi, Gwinnup, & Zadorozny, 1994) Prompted (Schnelle, 1990) and scheduled voiding (Jarvis, 1981)	Increased cue visibility. Change lighting levels (Beck & Heacock, 1988; Koss & Gilmore, 1998) Increase toilet visibility, raise toilet seat
Enhance self-perception	Reminiscence therapy (Gagnon, 1997; Hoerster et al., 2001) Validation therapy (Morton & Bleathman, 1991; Toseland et al., 1997) Self-identity based interventions	Provide objects (e.g., furniture, photographs) that were meaningful in the past
Enhance pleasure	Interventions based on past or present pleasant activities Sensory interventions Structured activities	Provide objects (e.g., pictures, piano) that were liked in the past
Decrease behavior problems	Sensory interventions Behavioral interventions Social contact interventions Structured activities Staff training Environmental interventions Medical/Nursing care interventions Combination therapies (Cohen-Mansfield, 2001)	Allow access to a secure outdoor garden

Table © Cohen-Mansfield (2004).

goals and need to be matched to individuals based on their cognitive and sensory abilities as well as on their identities, habits, and preferences. In addition to tailoring the type of intervention to the person and their specific needs, the content of the intervention, such as the type of music that is played, needs to be matched to the person's preferences. This need to individualize treatments makes both research and practice more complex.

## **4.2. Cognitive Interventions**

### **4.2.1. Memory Training**

Positive results have been described for memory training with older persons. However, these are usually conducted with persons with either no memory impairment or mild cognitive impairment. Interventions reported for

persons with dementia usually involve more support and less reliance on the person than traditional memory training methods. Interventions such as mnemonics are therefore usually inappropriate for this population. However, positive results have been demonstrated in small studies that involved daily cognitive stimulation by caregivers and with a computerized cognitive remediation system created by Butti and colleagues.

### **4.2.2. Cognitive Restructuring**

Cognitive interventions based on Beck's and Ellis's work have been used to treat depressed persons with mild levels of dementia. These interventions involve cognitive restructuring, that is, challenging or changing distorted thought processes and thoughts that cause depressed affect.

#### **4.2.3. Task Simplification/Cognitive Prosthesis**

The most common and straightforward cognitive methods involve task breakdown, enhanced instruction, modeling, rehearsal, cueing, and gradual approximations of a task. These principles apply across different domains. The notion of simplifying tasks can also be applied across daily activities such as by providing consistency in caregiver assignment (this helps to orient older persons across different tasks). Similarly, many environmental interventions have been used to simplify orientation such as placing items that are frequently needed in a permanent location, using labels and signs, wearing a large wristwatch, and placing a card with important information in a pocket.

#### **4.2.4. Memory Books**

Memory books are booklets with autobiographical, daily schedule, and problem resolution information. When nursing assistants were trained to use the memory books, nursing home residents showed improvement in conversation and affect.

#### **4.2.5. Spaced Retrieval for Procedural Memory**

Spaced retrieval is a memory training technique that involves repeated rehearsal, with testing occurring over gradually increasing time intervals. For example, the resident may be trained to check whether his or her glasses are being worn. The resident would first be asked to check immediately after instruction, then after 20 seconds, after 40 seconds, after 1 minute, after 2 minutes, and so on. In this way, people with dementia are able to learn procedures so as to consistently check their calendars. This procedural learning is essential for independent functioning and is based on the use of procedural implicit memory, which is retained until the later stages of dementia.

#### **4.2.6. Reality Orientation**

Reality orientation is an intervention in which staff members present orienting information, including information about time, place, and person, to persons with dementia. Such information may be conveyed in special group sessions, which may include discussion of current events or may be ongoing in interactions between (usually formal) caregivers and persons with dementia.

#### **4.2.7. Cognitive Tasks**

Cognitive tasks include both group and individual activities that involve use of cognitive and memory skills. Group examples include “question-asking readings” in which a group reads a relatively low-reading level script that is typed in large font and high contrast and that is accompanied by questions typed on cards that allow participants to discuss related topics. Another group memory task is “memory bingo,” in which participants have bingo-type cards with endings of popular sayings. For each card, the group facilitator reads the beginning of the saying, and participants use long-term memory to complete the sentence and use recognition skills to find it on the card in front of them. The sayings can also be used to stimulate discussion. Individual cognitive tasks include activities such as card or object sorting by category.

### **4.3. Behavioral Interventions**

#### **4.3.1. Differential Reinforcement**

Differential reinforcement may involve positive reinforcement contingent on nonagitated behavior or “time out” (e.g., moving the person to a quiet area when agitated) or “restriction” (e.g., denying the person goods [e.g., candy], activities, or access to a location or another person when agitated).

#### **4.3.2. Stimulus Control**

Stimulus control is based on behavior being emitted under specific antecedent conditions and not others. Stimulus control involves changing the stimuli that tend to trigger a behavior, so that the behavior will be less likely to be triggered, or changing the association between the antecedent stimulus and the behavior by changing contingencies depending on the presence of various antecedent stimuli. Camouflaging exit doors and placing stop signs on them, along with training that the stop sign means stop and walk away, are examples of the use of stimulus control.

#### **4.3.3. Maximizing Control**

Helplessness has been shown to relate to depressed affect, whereas control is associated with well-being. Providing residents with opportunities to exercise control can involve having a door that they can open to an outdoor area, having a plant that they can care for, or allowing them to make decisions about meals, clothes, and the like.

#### **4.3.4. Increasing Levels of Pleasant Activities or Noncontingent Reinforcements**

Clarifying which activities or experiences are reinforcing to individuals with dementia and providing these activities or experiences is a method that has been used to treat depressed affect in persons with dementia.

### **4.4. Self-Affirming Interventions**

#### **4.4.1. Reminiscence Therapy**

Reminiscence therapy encourages persons with dementia to recall their pasts. Sessions may use audio-visual aids such as old family photos and other objects. Reminiscence can enhance patients' sense of identity and sense of worth and/or general well-being, and may also stimulate memory processes.

#### **4.4.2. Validation Therapy**

Validation therapy, developed by Feil, involves communication with a therapist who accepts the disorientation of the person with dementia and validates his or her feelings. The assumption is that the person returns to unfinished conflicts in the past, around which a meaningful conversation can take place addressing the emotions that are important for the elderly person rather than trying to correct the disorientation. Thus, it fits within the perspective of humanistic theory. Validation therapy can take place in one-on-one treatment or in a group.

### **4.5. Sensory Interventions**

#### **4.5.1. Massage/Touch**

Massage or therapeutic touch has been used to decrease behavior problems in people with dementia. Touch is sometimes considered a form of communication when verbal communication is no longer available. In many of the articles reporting on massage therapy, the procedures took approximately 5 minutes and were performed once or twice a day.

#### **4.5.2. Music and White Noise**

Music has been used for decreasing behavior problems and for improving cognition or function. Music has been used to relax patients during meals or bathing or as a means of providing sensory stimulation to people who are understimulated. Music interventions take many forms, including listening to a music tape, playing

musical games, dancing, moving to music, and singing. Prior to using music therapy, hearing would need to be checked, and an amplifier, headphone, or hearing aid might need to be used. Music is more effective when it is individualized to match the person's preferences. White noise has also been used to induce relaxation, thereby improving sleep and decreasing restlessness.

#### **4.5.3. Sensory Stimulation**

Sensory stimulation involves the presentation of stimuli that affect different sensory modalities, including hearing, touching, seeing, and smelling. Therefore, it can include elements of the massage and music interventions described previously as well as aromatherapy, moving lights, pictures, and the like. The "Snoezelen" sensory stimulation program, which was developed in The Netherlands and includes a variety of relaxing stimuli, is also a type of sensory stimulation.

### **4.6. Social Contact Interventions**

Loneliness is highly correlated with depressed affect, and being alone has been shown to relate to behavioral problems. The best intervention for loneliness is positive interaction with a person who is meaningful to the elderly person. However, this is often not feasible. Therefore, a variety of alternative social interventions have been developed. In addition to the following social contact interventions, most group activities, including cognitive ones, can be used as vehicles to promote social contacts.

#### **4.6.1. One-on-One Interaction**

One-on-one interaction is a potent intervention for loneliness and behavior problems. It can be conducted with relatives, paid caregivers, or volunteers.

#### **4.6.2. Pets/Dolls**

Pet therapy often involves visits with a dog that last from a half-hour to an hour. However, other pets (e.g., a cat, fish) can be used as well. Even plush stuffed animals have been used successfully, as have robotic pets. Dolls have also been used to simulate companions, since they can be viewed as real babies. Pet therapy not only involves interaction with the pet but also serves as a topic for interaction with other people.

#### **4.6.3. Stimulated Interaction/Family Videos/Interaction Videos**

Simulated presence therapy uses an audiotape to simulate phone interactions. The tape contains a relative's portion of a telephone conversation and leaves pauses that allow the older person to respond to the relative's questions. Family videotapes also simulate interaction by having a loved one talking to the person with dementia. These videotapes have been found to produce engagement and decrease agitation in persons with dementia. Interaction videos are those in which the persons on the videos, usually professionals, interact with the persons with dementia, often recalling information from the past or inviting the persons with dementia to sing along, using long-term memory for well-known songs.

### **4.7. Structured Activities**

#### **4.7.1. Indoor Activities**

Structured activities are used to improve affect and decrease behavior problems. They can take many forms and include activities that Buettner termed "simple pleasures," including both group and individual activities. The activities may involve manipulation (e.g., ball throwing), nurturing (e.g., watering a plant), sorting, cooking, sewing, or engaging in sensory intervention as described previously (e.g., music, tactile stimulation with a fabric book). The content of activities may be based on information regarding which activities were or are reinforcing to the individual.

#### **4.7.2. Outdoor Activities**

Outdoor walks and physical activities are forms of structured activities that have the potential to improve function and affect and to decrease problem behaviors. Outdoor walks may take place in the company of a caregiver and so may also involve a social component, or they may occur in a secure outdoor area involving an environmental intervention. Outdoor walks are often limited by weather conditions.

### **4.8. Adapted/Enhanced Environmental Interventions**

Environmental interventions can be used for multiple goals such as to decrease or accommodate behavior

problems, to enhance functional status, and to increase orientation.

Environments for increasing functional ability have included different levels of environmental adaptation, from decreasing clutter to providing grab bars or handrails. The following are specific types of adaptation or enhancements.

#### **4.8.1. Visibility of Cues Needed for Activities of Daily Living**

The importance of visibility has been demonstrated in several research studies. For example, use of toilets increased when toilets were more visible, and eating behavior improved with better light and increased contrast between plates and the table.

#### **4.8.2. Seating in Motion**

Rocking chairs and gliding swings have been used for relaxation, improved affect, stimulation, and reduced physical agitation.

#### **4.8.3. Homelike Ambience**

Homelike furnishings in the institutional setting and enhanced ambience can affect many aspects of well-being for both patients and their caregivers.

#### **4.8.4. Reduced Falls**

Decreasing bed height, placing mattresses on the floor, using hip protectors, improving light on the way to the bathroom, and improving call systems all can help to decrease the risk of injury due to falling.

#### **4.8.5. Privacy and Intrusion Deterrence**

To prevent trespassing into other people's rooms or through emergency exit doors, doors and doorknobs can be camouflaged with cloth panels or murals, thereby disguising the doors. In addition, providing alternative doors, which can be controlled by the patient and permit movement into another secured area, can be useful in reducing trespassing.

#### **4.8.6. Clear Labels and Signs**

Clearly labeled signs with large font and high contrast can be used to increase orientation. Nursing homes can also use boxes with pictures of residents taken in

the past, or other memorabilia that is meaningful to them, to help orient residents in finding their rooms. Clear labels on public toilets are useful in decreasing incontinence.

#### **4.8.7. Wandering Areas**

Places that allow people to walk in a safe environment can be created either indoors or outdoors. These usually involve some type of walking loop that allows for walking but does not have an exit door that might allow for egress into an insecure area. In contrast, exit doors that can be opened by residents and allow them control of their own exit to another secure area can be therapeutic.

#### **4.8.8. Peace and Quiet**

Reduced stimulation environments, such as quiet rooms with soft colors, an absence of paging systems, and only a few objects, have been reported to be helpful. Reduced stimulation interventions may use neutral colors on pictures and walls and a consistent daily routine, with no televisions, radios, or telephones (except one phone for emergencies).

#### **4.8.9. Nature Environments**

Natural environments, including pictures of fish and/or sounds of bird songs or bubbling brooks, have been used to decrease agitation in the shower. Visual, auditory, and olfactory stimuli have been used to make nursing home corridors feel like home or like a natural outdoor environment, thereby improving resident and staff well-being and decreasing trespassing and exit-seeking behaviors.

### **4.9. Medical/Nursing Care Interventions**

Although medical and nursing care interventions would not usually be considered cognitive or behavioral treatments, they are included here because it is the behavioral assessment that is used to reveal the underlying etiology of negative affect or behavior problems that allows a medical or nursing intervention to relieve the condition. Therefore, interventions such as light therapy to improve sleep, pain management, reduction of discomfort by improved seating or positioning, and removal of physical restraints all have

been associated with improvement in behavior. Similarly, the provision of eyeglasses or hearing aids can be an important intervention for treating sensory deprivation and loneliness as well as the ensuing depressed affect and problem behaviors.

A behavioral approach may also guide nursing interventions directly in helping elderly persons to achieve maximal functional levels. One example is that of cognitive interventions that include cueing and task breakdown. Another example involves toileting management protocols. Two types of toileting protocols have been described. In scheduled or timed voiding toileting, patients are taken to the toilet either at fixed times (usually every 2 hours) or on a schedule that is based on their voiding pattern. Prompted voiding involves asking residents on a regular basis whether they need assistance with toileting. Patients are helped when they indicate such a need. Both types of protocols can be effective in reducing incontinence.

### **4.10. Staff/Caregiver Training and Management**

Staff training programs can take many forms and focus on any of the interventions outlined previously. Many focus on improved understanding of older persons and the impact of the disease and on improving verbal and nonverbal communications with persons suffering from dementia. Given that good communication skills are crucial for proper assessment of a problem area and for delivery of any intervention, training for such communication is the basis of good caregiving. Examples of methods for simplifying and clarifying communication using both verbal and nonverbal communication channels were provided by Beck and Heacock in 1988. Changing caregiver behavior through training is a complex and difficult challenge and often requires ongoing instruction, modeling, monitoring, feedback, and support of the caregiver. Therefore, in institutional settings, staff training is closely tied to management.

#### **4.11. Combination Therapies**

Good care involves using combinations of the available interventions and tailoring them to the needs, abilities, and preferences of the older person. In addition, many intervention programs combine elements from various intervention modalities.

## 5. EFFICACY, EFFECTIVENESS, AND UTILITY

There are two basic questions concerning interventions. First, are interventions effective? Second, what is the cost of interventions? There is insufficient research to answer either question, although partial answers have been suggested. One study reported the benefits of validation therapy for reducing aggressive behavior, and another reported on the benefits of reality orientation on orientation and affect, in comparison with control groups. However, in a review of reality orientation, validation therapy, and reminiscence therapy, Gagnon concluded that reality orientation and validity therapy do not produce sufficient change to justify their costs. Cohen-Mansfield reviewed 83 studies of nonpharmacological interventions for inappropriate behaviors in dementia and described the majority as reporting a positive, but not always significant, impact. Many of the studies included small samples and other methodological limitations, most often resulting from limited funding for this type of research.

There are few studies that have examined the question of cost. One study by Rovner and colleagues implemented an intervention that included daily group-structured activities, reduction of psychotropic medications, and weekly educational rounds with staff members. They calculated the cost of the 6-month intervention program at \$8.94 per patient per day. Another study by Schnelle *et al.* examined ways in which to improve nursing practices concerning exercise and incontinence care and concluded that current staffing levels are inadequate to provide good care.

Given the early stage of research in this field, there is an urgent need to increase our understanding about the following basic questions. Which interventions are efficacious for which individuals? Which aspects of an intervention are necessary for it to be efficacious? What are the active ingredients, or principles at work, in various interventions? Which personal characteristics (e.g., gender, culture, prior stress) need to be considered when matching an intervention with a person? Process variables (e.g., the person who is delivering the intervention, the timing of the intervention) and their significance need to be elucidated. Only after we have answers to these basic questions can the questions of effectiveness and costs be addressed properly.

## See Also the Following Articles

Aging and Competency ■ Cognitive Aging ■ Cognitive Skills: Training, Maintenance, and Daily Usage ■ Dementia in Older Adults ■ Elder Caregiving ■ End of Life Issues ■ Personality and Emotion in Late Life ■ Psychotherapy in Older Adults

## Further Reading

- Beck, A. (1997). The past and future of cognitive therapy. *Journal of Psychotherapy Practice & Research*, 6, 276–284.
- Beck, C., & Heacock, P. (1988). Nursing interventions for patients with Alzheimer's disease. *Nursing Clinics of North America*, 23, 95–124.
- Camp, C. J., Foss, J. W., O'Hanlon, A. M., & Stevens, A. B. (1996). Memory interventions for persons with dementia. *Applied Cognitive Psychology*, 10, 193–210.
- Cohen-Mansfield, J. (2000). Theoretical frameworks for behavioral problems in dementia. *Alzheimer's Care Quarterly*, 1(4), 8–21.
- Cohen-Mansfield, J. (2001). Nonpharmacologic interventions for inappropriate behaviors in dementia: A review, summary, and critique. *American Journal of Geriatric Psychiatry*, 9, 361–381.
- Day, K., & Calkins, M. P. (2002). Design and dementia. In R. B. Betchel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 374–393). New York: John Wiley.
- Ellis, A. (1962). *Reason and emotion in psychotherapy*. New York: Lyle Stuart.
- Feil, N. (1982). *Validation: The Feil method*. Cleveland, OH: Feil Productions.
- Gibbons, A. C. (1988). A review of literature for music development/education and music therapy with the elderly. *Music Therapy Perspectives*, 5, 33–40.
- Lewinson, P. M., & Youngren, M. A. (1976). The symptoms of depression. *Comprehensive Therapy*, 2(8), 62–69.
- Opie, J., Rosewarne, R., & O'Connor, D. (1999). The efficacy of psychosocial approaches to behaviour disorders in dementia: A systematic literature review. *Australian and New Zealand Journal of Psychiatry*, 33, 789–799.
- Schnelle, J. (1990). Treatment of urinary incontinence in nursing home patients by prompted voiding. *Journal of the American Geriatrics Society*, 38, 356–360.
- Seligman, M. E. P. (1974). Depression and learned helplessness. In R. J. Friedman, & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research*. Washington, DC: Winston-Wiley.
- Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco: Freeman.



# Cognitive Skills: Training, Maintenance, and Daily Usage

Karlene K. Ball, Virginia G. Wadley, David E. Vance, and Jerri D. Edwards

University of Alabama, Birmingham, Alabama, USA

1. Cognitive Skills and Their Importance for Daily Functioning
2. How Cognitive Skills May Be Maintained
3. How Cognitive Skills Can Be Trained and the Impact of Training
4. Conclusion  
Further Reading

Cognitive skills are basic and higher order mental abilities, such as attention, information processing, memory, executive functioning, and reasoning, that interact with sensory and perceptual systems to determine success in performing daily activities. Older adults are at risk for cognitive decline but may benefit from strategies to maintain cognitive skills and from training programs to enhance certain skills.

## GLOSSARY

*divided attention* The ability to split a person's attentional focus or to maintain vigilance with respect to two or more objects or activities.

*executive function* A set of higher order cognitive abilities primarily associated with frontal and prefrontal structures of the brain; it involves planning, organizes information, inhibits responses, and orchestrates mental resources.

*hippocampus* A subcortical structure of the brain that is key for memory function and is specifically involved in encoding new information, a vital first step for subsequent recall.

*mild cognitive impairment (MCI)* A term used to describe individuals with focal memory impairments but no evidence of frank impairment in other cognitive domains and no evidence of impaired daily functioning; MCI is often thought of as a preclinical stage of dementia.

*transfer of training* The capacity for training-induced improvements in cognitive skills to translate into improved performance of cognitively demanding activities.

## 1. COGNITIVE SKILLS AND THEIR IMPORTANCE FOR DAILY FUNCTIONING

The ability to think, remember, and reason allows people to negotiate the world in which they live, whether it be planning the steps in cooking a meal, remembering to take their medications on time, deciding how to spend their money, or reacting quickly to avoid hitting a dog in the road. People's cognitive abilities provide them with the flexibility to engage in a number of unique and demanding situations. Declines in any of these abilities can subsequently diminish the ability to function in day-to-day life. Cognitive abilities that tend to decline with age, with potential impact on the ability to perform everyday tasks, include attention, speed of processing, memory, executive functioning, and reasoning. These abilities interact with sensory and

perceptual systems to determine performance in daily life.

### **1.1. Sensation and Perception**

Information processing must begin with acquiring the information in the first place. Thus, age-related changes in sensory–perceptual abilities (e.g., poorer vision, hearing) can influence older adults' ability to perform measures of cognitive ability as well as everyday activities. For the most part, older adults can compensate for age-related sensory decline through corrective lenses and /or hearing aids. Older adults also compensate for degraded sensory information through cognitive strategies in which they use the context of a situation (e.g., conversation, other words in a sentence, situational cues) to understand what is being presented. However, impairments in sensory processing can lead not only to poorer performance on cognitive tests, and perhaps a false impression of cognitive impairment, but also to restrictions in everyday activities such as driving. Such restrictions may, in turn, lead to increased depression and/or the need for formal care.

### **1.2. Attention**

Attention encompasses a rather broad range of abilities, from simply orienting (e.g., turning attention toward a sudden noise), to sustained attention or vigilance (e.g., maintaining focus over an extended period of time). Divided attention is quite commonly used in everyday contexts and occurs when people try to do two things at the same time. This ability may or may not decline with age, depending on the individual and the difficulty of the two tasks being performed at the same time. The use of cell phones while driving is an example of a divided attention task that can be fairly difficult. There is mounting evidence that crash risk is elevated for drivers of all ages while using cell phones, and this risk may increase with age. In general, research indicates that older drivers are susceptible to decrements in driving performance when increased demands are placed on attention, whether it be cell phones or multiple events occurring at a congested intersection. Selective attention, or the ability to pay attention to relevant information while ignoring irrelevant information, is also germane in this context. It has generally been found that the detrimental effect of irrelevant information, or distraction, is higher for older adults.

### **1.3. Speed of Processing**

Speed of processing, or the ability to perceive and process information quickly, is one of the first cognitive abilities to decline with age. Over the past decade or so, many studies have demonstrated relationships between cognitive processing speed and everyday function in older adults. Various indexes of mobility have been linked to processing speed impairments in older adults with and without dementia. For example, reductions in life space, or the extent of a person's travel in his or her environment inside and outside the home, have been associated with impairments in processing speed. Furthermore, multiple studies have demonstrated that processing speed impairment is an excellent predictor of increased crash involvement in older drivers and is an even stronger predictor of injurious crashes. Reduced processing speed also is related to slower performance of instrumental activities of daily living (IADLs) such as looking up a telephone number, finding items on a crowded shelf, and reading medication labels.

### **1.4. Memory**

Memory is one of people's most obvious cognitive abilities. Of all the mental abilities, people report complaints about this ability more often than about any other ability, probably because it is fundamental to most everyday tasks. Also, memory decline can be obvious when a person is attempting to recall necessary information such as telling a doctor what medications he or she is taking or trying to remember a phone number. Although activation of both cortical and subcortical areas of the brain is necessary for good memory functioning, the hippocampus, a subcortical structure, has the primary responsibility of encoding new information, a vital first step for subsequent recall of the information. As seen with Alzheimer's disease, when the hippocampus is severely damaged, profound memory loss is experienced. Although some normal age-related decline in memory does occur with age, such changes generally do not severely affect daily functioning to the same degree as found in a dementia such as Alzheimer's disease. Regardless, age-related memory declines have been found to influence activities of daily living (ADLs), whether they be forgetting to adhere to medical appointments or not remembering to pay bills. Research has linked memory declines to self-reported impairment with daily shopping, preparing a hot meal, housecleaning, managing finances, and completing forms.

### 1.5. Executive Function

One of the most complex cognitive abilities is executive function, which includes people's ability to plan, organize information, inhibit responses, think abstractly, and reallocate mental resources. This ability is primarily associated with the frontal lobes of the brain and is known to decline with age. Executive functioning is necessary to complete tasks that require complex behavior or have multiple steps. For instance, executive functioning is useful in paying bills because this task requires planning (e.g., determining how much money to transfer from a savings account to a checking account), calculating how much money remains in the checking account after paying the bills, and (in many cases) deciding what is the least amount that can be paid to prioritize money to other uses. Thus, financial capacity can be impaired in individuals with executive dysfunction, as is seen in adults with Alzheimer's disease. However, dysfunction in executive abilities can also affect less cognitively demanding activities. For example, scientists have found that lower levels of executive functioning are related to impairment in the ability to perform other IADLs such as medication management, even among noninstitutionalized retirees.

### 1.6. Reasoning

Reasoning is another complex cognitive ability affected by aging. This ability is similar to executive functioning; in fact, damage to the frontal lobes also impairs the ability to reason. Unlike executive functioning, reasoning focuses on using logical constructs, knowledge, and principles to extrapolate a solution to a problem. Reasoning is a sophisticated problem-solving skill that requires both memory and executive functioning. Reasoning is used in a variety of real-world tasks, ranging from medical decisions to driving behavior. Thus, declines in reasoning ability can impair decision making in a variety of real-world scenarios. As with other cognitive skills, reasoning ability has been found to predict IADL performance. Scientists have found that inductive reasoning is important for figuring out problems associated with everyday tasks such as shopping and managing money. In fact, researchers have found that as inductive reasoning is enhanced, everyday problem solving also improves.

### 1.7. Cognitive Impairment

Whereas age-related declines in cognition and function may be relatively subtle and selective in normal aging, the declines experienced by individuals with mild

cognitive impairment (MCI) or dementia are more pronounced and become pervasive over time.

MCI is a term used to describe individuals with focal memory impairments but no evidence of frank impairment in other cognitive domains and no evidence of impaired daily functioning. MCI is often thought of as a preclinical stage of dementia. Approximately 5 to 15% of adults age 65 years or over may be affected by MCI. Individuals with MCI, by virtue of their memory impairments and their risk of progression to dementia, are at high risk for declines in the abilities needed to function effectively and independently in society. These individuals are likely to experience a gradual loss of functional skills of a magnitude that lies somewhere between the subtle decrements associated with normal aging and the much more obvious deficits associated with dementia. Little has been done to directly evaluate the impact of cognitive changes in MCI on the performance of everyday tasks. A critical question is how changes in memory and other cognitive abilities influence the performance of everyday tasks over the period of time that, for many individuals, will mark the transition from MCI to dementia. To date, no one has mapped cognitively demanding everyday activities to component cognitive abilities in MCI. Thus, it is difficult to determine whether some tasks of daily life are dependent on specific abilities or how levels of ability influence functional performance. A clearer understanding of this relationship within the MCI population is needed.

Dementia includes a category of diseases, such as Alzheimer's disease, characterized by more severe cognitive impairments that hamper a person's ability to perform everyday tasks. The incidence of dementia rises with age. For example, only approximately 5% of adults over 65 years of age have Alzheimer's disease, whereas 10 to 20% of adults over 80 years of age are affected, and estimates have indicated that as many as 40% of adults age 85 years or over suffer from the disease. Research in the field of cognitive aging has demonstrated that the deterioration of cognitive function that occurs with dementia contributes to a decline in the performance of everyday activities. Research has also shown a clear link between declining everyday function (e.g., managing finances, grocery shopping, driving) and subsequent hospitalizations, need for care, and death.

## 2. HOW COGNITIVE SKILLS MAY BE MAINTAINED

Cognitive decline in numerous areas can result in difficulty in performing tasks important for everyday life and

for maintaining independence throughout old age. Thus, an important question to consider is, "How can individuals maintain cognitive abilities and avoid cognitive decline?" Several factors have been related to sustained cognitive function with age, including education level, physical activity level, pulmonary health, and feelings of self-efficacy. By and large, healthy and active lifestyles that include proper diet and exercise, prevention of disease, and the avoidance of trauma all are important for sustaining cognitive capacity with advancing age.

Higher education levels are associated with better cognitive performance during older adulthood. Education level affects cognitive function both directly, through enhancing brain function, and indirectly, through individuals with more education being more likely to participate in intellectually stimulating activities that, in turn, preserve cognitive abilities.

Physical activity has various benefits for older adults and positively affects cognitive function in a variety of ways. For example, physical activity has been shown to decrease depression (which can cause cognitive deficits, particularly for older adults) and to increase overall physical health and improve quality of life. Strenuous exercise directly benefits brain and central nervous system functioning, which are vital to maintaining cognitive abilities. Accordingly, aerobically active individuals perform better on measures of working memory, reasoning, and speed of processing than do those who exhibit more sedentary behavior.

In addition to exercise, many facets of physical health have been related to cognitive function. For example, a specific physiological measure, pulmonary peak expiratory flow rate, has been found to be predictive of sustained cognitive ability with age. This physiological index is related to cardiovascular health and physical activity level. Studies have noted in particular that preserved vascular and cardiovascular health is vital for cognitive function. For example, decreased cerebral blood flow is associated with poor cognitive performance. Furthermore, poor cardiovascular health is associated with increased incidence of stroke, a leading cause of cognitive impairment. In the Seattle Longitudinal Study (SLS), absence of cardiovascular disease was linked to better reasoning ability and speed of processing performance for older adults. A number of other diseases, such as diabetes, emphysema, and Parkinson's disease, can also result in cognitive deficits. In general, the fewer chronic diseases a person has, the better his or her chances of maintaining cognitive abilities with advancing age.

Overall, a healthy lifestyle is important not only for physical health but also for mental health. Obesity has been related to higher incidence of many chronic

diseases, which in turn are related to cognitive decline. Cigarette smoking and excessive alcohol consumption both have been linked to higher incidence of stroke, a leading cause of cognitive impairment, as well as to many other poor health outcomes, thereby negatively affecting cognition. Furthermore, alcoholism can be directly detrimental to cognitive function. Most obviously, traumatic head injury results in cognitive impairment and is associated with increased incidence of dementia later in life. Therefore, simple preventive behavior, such as using seat belts combined with air bags while in automobiles and using helmets while riding on motorcycles, bicycles, or scooters, is advised.

In addition to physical health and healthy lifestyles, psychological well-being is important for preserving cognitive function. An individual's attitude can affect his or her cognitive performance with age. For example, self-efficacy, a positive evaluation of one's own cognitive capabilities, is related to better cognitive performance. Individuals in the SLS who rated their personalities as "flexible" during mid-life were found to be at reduced risk for cognitive decline in later life. Furthermore, individuals who indicated satisfaction with their life accomplishments during middle age performed better on cognitive measures in later life.

Sustained social activity also is associated with better cognitive performance. For example, researchers have found that those who engage in social, domestic, and leisure pursuits are less likely to have cognitive impairment. Similarly, the SLS indicated that older adults who had high socioeconomic status and were socially active experienced the least cognitive decline over a 7- to 14-year period. Risk of cognitive decline was lessened for individuals who read extensively, traveled, attended cultural events, were involved in clubs and professional associations, and pursued continuing education opportunities. Extensive social networks and support systems, including intact families, have also been related to better cognitive functioning with age. Conversely, depression has been identified as a psychological/biological cause of decreased cognitive function. Particularly for older adults, depression can manifest itself as cognitive impairment.

### **3. HOW COGNITIVE SKILLS CAN BE TRAINED AND THE IMPACT OF TRAINING**

#### **3.1. Cognitive Training**

Given that cognitive abilities can and do decline with age, often resulting in difficulty in performing everyday

tasks, the possibility of cognitive training to prevent, slow, or reverse age-related cognitive decline has been investigated. A growing number of studies now support the protective effects of intellectual stimulation on cognitive abilities for older adults without dementia. Early studies in the area of cognitive training were conducted within the Adult Development and Enrichment Project (ADEPT) and the SLS. Both of these studies provided 5 hours of strategy training, preceded and followed by cognitive assessment. Significant cognitive training gains were observed in both studies for the specific abilities that were trained.

A large randomized clinical trial, Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE), recently evaluated the impact of three promising cognitive interventions—speed of processing training, memory training, and reasoning training—on the maintenance of both cognitive and day-to-day abilities in community-living older adults. The study showed that for all three interventions, there were significant and specific improvements in cognitive ability as well as an increased benefit of additional booster training. The amount of training gain for cognitive abilities was equal to or greater than the amount of decline that would be expected in older adults without dementia over 4 to 14 years of aging in the absence of any training.

### 3.2. Durability of Training

In the ACTIVE study, among those individuals who experienced immediate cognitive improvements, training gains were found to persist for at least a 2-year follow-up period for all three interventions. Prior cognitive training studies have also pointed to the durability of training effects, with maintenance of reasoning training demonstrated at even 7 years after original training. Those older adults in the SLS who received reasoning booster training were found to substantially outperform their initial baseline levels 7 years later. Although studies of memory training have consistently shown that memory improvement can be maintained 1 week to 6 months, results on the durability of training beyond 6 months are less consistent. Some investigators have reported a significant decrease in memory performance and a substantial decrease in mnemonic strategy use over a 3-year interval in a memory training group. Multifactorial memory training programs, such as those incorporating psychomotor training in addition to memory training, have resulted in maintenance of memory improvement extending to 3½ years.

Maintenance of speed of processing training effects has not been evaluated extensively beyond the ACTIVE study, which demonstrated maintenance for at least 2 years. In an early training study, speed training effects were found to persist over a 6-month period. Furthermore, participants in the older age group (>60 years) performed at the same level following training as the middle age group (40–59 years) had performed at baseline. In a subsequent study among older adults, training effects were found to persist over 18 months and remained significantly better than baseline. Thus, in general, cognitive training studies have supported the durability of training effects.

### 3.3. Transfer of Training

Given that cognitive training can result in improvements of targeted cognitive skills, can these cognitive improvements transfer to improvements in performance of everyday activities? Older adults are constantly faced with learning new things such as skills for using computers, home medical devices, automated teller machines, videocassette recorders, and job-related skills. In addition, older adults would benefit from preservation of functional abilities that were most likely acquired at a younger age (e.g., driving, other IADLs). Although the onset of decline occurs for many cognitive abilities during a person's 60s, the onset of decline for everyday abilities typically occurs during the 70s or later. Obviously, functional competence is multidimensional, relying on multiple cognitive abilities. Consequently, assessment of functional competence/everyday abilities has included multiple methods, including global measures, performance-based measures, and self-report measures.

An example of a cognitively demanding task of daily living is the Everyday Problems Test (EPT), which assesses a person's ability to interact with 14 everyday situations (e.g., doubling a recipe, interpreting transportation schedules, understanding a Medicare benefit chart). Similarly, the Observed Tasks of Daily Living (OTDL) assessment presents tasks in three domains: medication use, financial management, and telephone use. Participants are asked to perform actions required to solve everyday tasks (e.g., searching medication label for side effects, making change). Performance on these measures has been related to multiple basic cognitive abilities; therefore, it may improve with cognitive training.

Other measures of everyday ability emphasize the speed of responding to real-world stimuli. For

example, slower processing abilities can present a risk with respect to safe driving. The speed of processing research has examined age differences in the Useful Field of View (a collection of speeded tasks evaluated in the ACTIVE clinical trial) as well as the benefits of speed of processing training to both cognitive abilities and everyday tasks. Prior training research found that improved speed of processing resulted in improved braking times in a driving simulator in addition to significantly reducing the number of dangerous maneuvers made by drivers during an open-road driving evaluation and maintaining the extent of people's driving over an 18-month period. Self-reported difficulty in driving, driving avoidance, and reducing the extent of people's driving all have been found to be associated with slower processing speed. Despite the fact that individuals with impairments appear to self-regulate their driving performance by reducing their exposure to challenging driving situations, studies have shown that these modifications have not been sufficient to eliminate their increased risk of crash involvement. Another measure, the Timed IADL (TIADL) test, provides an index of the time needed to successfully complete five tasks (e.g., finding a number in a telephone book, finding specific items of food on a grocery shelf). The time needed to complete the TIADL tasks is similarly reduced following speed of processing training among older individuals with impaired processing speed. Thus, speed of processing training has been found to generalize to important everyday abilities in addition to boosting cognitive ability.

#### 4. CONCLUSION

Clearly, there are many aspects of cognition that can potentially decline as a part of the aging process, and some of these declines can affect performance of

everyday tasks. There is a growing body of literature demonstrating that these declines can be prevented or slowed through changes in lifestyle. Progress also has been made in the area of cognitive training, indicating that cognitive decline is not necessarily irreversible during older age and that interventions may allow longer maintenance of cognitive function into older age. Furthermore, the fact that many cognitive declines are amenable to moderating effects implies that the independence and mobility of individuals can be extended into later life. In short, the outlook for preventing, delaying, or reversing the onset of age-related cognitive decline is bright.

#### See Also the Following Articles

Cognition and Culture ■ Cognitive Aging ■ Cognitive and Behavioral Interventions for Persons with Dementia ■ Dementia in Older Adults

#### Further Reading

- Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Morris, J. N., Rebok, G. W., Smith, D. M., Tennstedt, S. L., Unverzagt, F. W., & Willis, S. L., for the ACTIVE Study Group. (2002). Effect of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association*, 288, 2271–2281.
- Birren, J. E., & Schaie, K. W. (Eds.). (1995). *Handbook of the psychology of aging* (4th ed.). San Diego: Academic Press.
- Fisk, A. D., & Rogers, W. A. (Eds.). (1996). *Handbook of human factors and the older adult*. San Diego: Academic Press.
- Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study*. New York: Cambridge University Press.
- Willis, S. L. (1987). Cognitive training and everyday competence. In K. W. Schaie (Ed.), *Annual review of gerontology and geriatrics*, (Vol. 7, pp. 159–188). New York: Springer.



# Cognitivism

Pío Tudela

University of Granada, Granada, Spain

1. Cognitivism During the Behaviorist Era
  2. The Information-Processing Approach
  3. Cognitive Science
  4. Tenets of Cognitivism
  5. Cognitivist Influence
- Further Reading

## GLOSSARY

**adaptive control of thought (ACT)** An influential unified theory of cognition based on a symbolic architecture of the mind, developed by John Anderson. This theory had several different versions; one of the most influential is known as ACT\*.

**automatic and controlled processes** A cognitive process is said to be automatic when its functioning may be triggered by the mere presence of a stimulus, without need for conscious attention. It is called a controlled process when its functioning depends upon conscious attention.

**bottom-up and top-down processing** Cognitive processing takes place in a bottom-up manner when the direction of the information processing flow goes from simple (or low level) to complex (or higher level) processes. Top-down processing goes in the opposite direction and refers to the influence of high level on lower level processes.

**cognitive neuroscience** An interdisciplinary enterprise in which cognitive psychology, neuroscience, and computer science cooperate in order to study and understand the relationship between mind and brain; considered a subdivision of cognitive science.

**cognitive science** An interdisciplinary enterprise in which different branches of scientific knowledge dedicated to

the study of cognition merge. Six disciplines are involved: psychology, computer science, linguistics, philosophy, anthropology, and neuroscience.

**connectionism** An approach within cognitive science contending that information processing takes place through the interactions of a large number of simple processing elements called units, each sending excitatory and inhibitory signals to other units. Parallel processing and distributed representations are the main characteristics of connectionist (also called PDP) models.

**information processing** An approach to the study of cognition that considers the human mind as a symbol-manipulating system similar to a computer. Cognition can be divided into a series of stages; at each stage, certain particular operations are performed on incoming information. The eventual performance of the system is thought to be the outcome of this series of stages and operations.

**levels of explanation** Different viewpoints from where the behavior of a complex information processing system may be considered, described, and eventually explained.

**local and distributed representations** A representation is said to be local if there is a one-to-one correspondence between entities in the represented world and elements in the representation. In a distributed representation, each entity is represented by a pattern of activity distributed over many elements in the representation, and each element is involved in representing many different entities of the represented world.

**neural networks** Sets of elementary, neuron-like information-processing units related to each other by excitatory or inhibitory connections. Neural networks form the basic structure of connectionist architectures.

**physical symbol system** A system consisting of physical symbols. A physical symbol is a physical entity able to

designate or refer to another entity that is the *meaning* of the symbol. Physical symbols form the basic structure of symbolic architectures.

**serial and parallel processing** Information processing is organized in a serial way when processes are carried out one after another, each taking as input the output of the previous one and beginning operation when the preceding process has finished. When processes are organized in parallel, they operate simultaneously and their input is independent of the output of other processes.

**situated cognition** A nonrepresentational approach within cognitive science contending that mind, body, and environment are coupled systems, dynamically interacting with each other. Hence, they should not be studied as isolated systems; rather, their mutual interactions should be the main goal of cognitive science.

**Soar** A general cognitive symbolic architecture for developing systems that exhibit intelligent behavior. It was created by John Laird, Paul Rosenbloom, and Allen Newell.

In the broad sense of the term, cognitivism encompasses all those theoretical approaches that consider the study of the human mind and its constituent processes as the main objective of psychological science. In this sense, psychology, as an experimental enterprise, was cognitive from its very beginning. For example, W. Wundt considered the task of psychology to be the analysis of conscious processes into its constituent elements, and many chapters in William James' book *The Principles of Psychology* were dedicated to analyze processes such as attention, memory, and consciousness, the core cognitive processes. In a more strict sense, cognitivism refers to a particular theoretical approach to studying cognitive processes that originated as a reaction against behaviorism in the midst of the 20th century and is known as the information-processing approach. This article primarily concentrates on the origin and subsequent development of this approach, though attention is also paid to other cognitive views influencing psychological research.

## 1. COGNITIVISM DURING THE BEHAVIORIST ERA

From a cognitive point of view, the contribution of behaviorism to our knowledge of psychological processes was almost restricted to the realms of learning, motivation, and emotion. Little if any concern was shown about cognitive processes such as perception, attention, memory, and consciousness in the writings

of Hull or Skinner. Even when they made a foray into the territories of thought (C. Hull) or language (B. Skinner), their nonrepresentational approach turned out to be too restricted and had little influence on subsequent research. Among behaviorists, E. Tolman was one of the few who did not show disdain for using terms and concepts with a cognitive flavor. Even in his case, however, those terms did not refer to processes to be studied for their own interest; rather, they were conceptualized as intervening variables useful to predict behavior.

It should be mentioned, however, that behaviorism did provide a lasting contribution to cognitive and psychological research in general. Emphasis on the measurement of variables, the careful design of experiments, and the wide-ranging use of objective scientific methods is part and parcel of the invaluable legacy of behaviorism. Thanks to this methodological heritage, modern cognitivism could establish itself on a solid scientific foothold.

Despite the general sway of behaviorism during the second quarter of the 20th century in the United States, interest in cognitive research remained active within restricted domains of psychological inquiry. This was particularly so in Europe, where the influence of behaviorism took longer to be felt and was never as strong as it was in the United States. The following streams of investigation deserve to be pointed out:

1. The contribution of Gestalt psychologists to the development of perceptual theory.
2. Research on concept formation, problem solving, and thinking by researchers such as Edouard Claparède, Wolfgang Köhler, and Frederic Bartlett.
3. The sociohistorical approach to cognitive development carried out in Russia by Lev Semenovitch Vygotsky.
4. Jean Piaget's research program on cognitive development and genetic epistemology in Switzerland. Piaget was one of the first thinkers to realize the interdisciplinary nature of cognitive research. In 1955 he founded the International Center for Genetic Epistemology in Geneva, where scientists from different disciplines tried to set up the foundations of a true cognitive science. Piaget influence on cognitive development and educational psychology is still noticeable today; research along his line of thinking is an integral part of contemporary cognitive research.
5. It is important to recognize that, in some areas of psychology, behaviorism never played the dominant role it did within experimental psychology. References

to mind and mentalistic concepts never disappeared from social and clinical psychology. Authors such as Kurt Lewin, Leon Festinger, Erich Fromm, Carl Rogers, and Abraham Maslow are only a few among the many who did not adhere to the dominant behavioristic approach.

All these streams of investigation paved the way for what has come to be known as the cognitive revolution.

## 2. THE INFORMATION-PROCESSING APPROACH

Cognitivism, as it was articulated within the information-processing approach, was from its beginning the result of multiple lines of thought coming from different scientific disciplines. Within psychology, there was a profound dissatisfaction with behaviorism as a method of tackling complex human behavior. World War II compelled psychologists to face practical problems, such as the manipulation of aircraft, radar, and other highly technical devices involving multiple stimuli and responses and posing heavy demands on human operators. In these situations, concepts from the animal learning labs had little application, so psychologists turned to communication engineering and information theory for help. The human operator was conceptualized as an information transmitter and decision maker, and research on selective attention followed as a matter of course.

In 1959, the linguist Noam Chomsky published a devastating review of Skinner's book *Verbal Behavior*; his review was also an attack on other behavioristic accounts of language. As an alternative to studying linguistic performance, Chomsky emphasized linguistic competence, knowledge about the rules of grammar that the native speaker of a language owns. Under Chomsky's influence, psychologists became interested in the cognitive processes responsible for the structural aspects of human language. Subsequently, the significance of linguistic structures for our understanding of memory and thought was increasingly acknowledged.

The most determining influence, however, came from the realm of computer science. Computers are general purpose machines that can be programmed to perform any well-defined task. As information-processing systems, they are able to receive, store, transform, and retrieve information. A computer performance may be analyzed at the physical level of circuits used to implement the

system (hardware), but there is a different, more abstract program level (software) that lends itself as an adequate analogy for the human mind. Programs are sets of instructions ready to act upon data; they may take diverse sequences of actions depending upon different conditions. As both instructions and data are stored in the same symbolic form, programs can also act upon themselves in a recursive manner, changing their own instructions if the task demands it.

When computers first appeared, they were largely considered to be mathematical devices, number-manipulating machines that would make calculations easier. But it soon became apparent that they were able to deal not only with numbers but also, more generally, with symbols. Thus, computers came to be conceptualized as symbol-manipulating systems. Computers can take symbolic input, recode it, store it, retrieve it, make decisions based on the recoded information, and produce symbolic output. These operations closely resemble mental processes and offer a precise and mechanistic way of thinking about them. Hence the computer became the best metaphor to lead the scientific research of the human mind.

There is general agreement that the critical time period giving a boost to cognitivism took place between 1954 and 1960. In 1956, a summer seminar took place at Dartmouth, where Allen Newell, Herbert Simon, and other major workers on artificial intelligence gathered to establish a research program for the new discipline interested in building programs able to generate intelligent behavior. In 1958, another summer seminar was organized by Newell and Simon at the RAND Corporation in Santa Monica in order to acquaint psychologists with computer simulation techniques and their application to the study of cognitive processes. In 1960, the Center for Cognitive Studies was created at Harvard, where Jerome Bruner and George Miller served as a point of reference for the new cognitive psychology.

## 3. COGNITIVE SCIENCE

The time from 1960 to approximately 1976 was a consolidation period for cognitive psychology. Processes such as attention, memory, language, and thought moved to the forefront of psychological research. In 1967, Ulric Neisser published a textbook, titled *Cognitive Psychology*, that served as a reference for the psychological research that was being done within the new cognitive framework. The heuristic value of the computer analogy was widely accepted, though it was

not understood the same way by all researchers. For some, the computer was just a metaphor that helped to conceptualize human mind as an information-processing system. For others, however, the computer was much more than a metaphor; they assumed that both the computer and the mind were examples of a type of system with specific properties. In 1976, Newell and Simon proposed the term physical symbol system as a name for this type of system and as a fundamental concept for a general and unifying approach to cognition that came to be known as cognitive science. That same year, the Sloan Foundation became interested in this approach and created an interdisciplinary research program to explore its possibilities. In 1977, the journal *Cognitive Science* was founded, and the Cognitive Science Society followed 2 years later. Nowadays cognitive science has been established as an academic discipline in several universities in different countries.

Cognitive science was an interdisciplinary enterprise in which different branches of scientific knowledge dedicated to the study of cognition merged. Originally, there were six disciplines involved, namely, psychology, computer science, linguistics, philosophy, anthropology, and neuroscience. Each had developed a particular way of looking at cognition, and some already had productive bilateral relations that had given rise to new fields of cognitive research, as was the case for psycholinguistics or computational linguistics. Nevertheless, there was the shared belief that progress within each discipline largely depended upon progress in the rest of them. Today, links between the six disciplines exist, though some are stronger than others, and new fields of research, as is the case for cognitive neuroscience, have emerged.

## 4. TENETS OF COGNITIVISM

### 4.1. Levels of Explanation

Cognitivism is not reductionist; it maintains that complex information-processing systems, such as the computer or the human mind, cannot be understood as a simple extrapolation of the properties of its elementary components. To understand a complex system, we have to take into consideration the fact that its behavior can be described at different levels of abstraction, each related to particular questions that may be asked about the system. If an explanation is conceptualized as an answer to a specific question, then we must be open to expect different explanations depending upon

the description level associated to the question. The classical view of cognitive science distinguishes at least three different levels of analysis, named according to David Marr terminology.

#### 4.1.1. The Computational Level

The computational level is the level of abstract problem analysis and refers to what the system does and why. At this level, explaining why people or computers do something requires pointing to their goals and objectives and to the strategies and means to carry them out. Take, as a simple example, the case of a calculator able to perform number addition. An explanation of what the machine does and why should be found in arithmetic theory, because the task to be performed by the system has to comply with the principles of this theory.

#### 4.1.2. The Algorithmic Level

The algorithmic level is the program level in which a representation for coding the system's input and output should be specified, together with an algorithm providing the appropriate output for a given input. This level refers to how the system brings about its task. It is also called the symbol level, because the goals established at the computational level are encoded by symbolic expressions. In the case of the calculator from the example above, explaining how the system resolves an addition has to do with the representation used to encode numbers (Arabic, Roman, or binary, for instance) and the specific algorithm employed, for example, adding the units first and carrying to the tens if the sum exceeds 9.

#### 4.1.3. The Level of Physical Implementation

The level of physical implementation is concerned with the particular technology used to physically realize the system. In our example, an explanation at this level would point to the working of electronic circuits, but electronic circuits are not the only way to implement an addition machine. Old cash registers, for example, were not electronic but mechanical; however, they could use the same algorithm as an electronic calculator to perform addition.

These three levels are frequently related and influence each other; for example, the choice of electronic circuits to build an addition machine may favor a binary over a decimal representation of numbers. However, it is convenient to note that, because the

three levels are not necessarily bound, some phenomena may find an explanation at only one or two of them. Above all, it should be noted that it is pointless to search for only one valid explanation of the behavior of complex information-processing systems. An explanation only makes sense in relation to a question referred to a particular level of analysis. Therefore, we need explanations at different levels in order to understand complex information-processing systems.

## 4.2. Systems Approach

Cognitivism views human cognition as an activity emerging from the interaction of a system of components. Though the search for components is considered a valuable aspect of the scientific activity, cognitivism emphasizes the significance of structure and functional architecture for our understanding of the mind.

From the beginning, cognitive psychology made wide use of flowchart models as a means of specifying the component processes involved in a particular situation. D. Broadbent's model was one of the first to specify the attention and memory components needed to explain human behavior in situations in which a person has to select one out of several different streams of information, as exemplified by a cocktail party, where a person has to follow one among many surrounding conversations. At first, models tended to be bound to specific tasks or situations, but soon interest for building general models became dominant. The model of human memory advanced by R. Atkinson and R. Shiffrin in 1968 is a good example of this trend. These authors were able to develop a memory model consisting of three different components, namely, sensory registers, short-term memory, and long-term memory. Extensive previous research had been carried out on each of these components, but Atkinson and Shiffrin managed to put them together within a general framework that was influential for years. Memory research also offers a good illustration of the analytic procedure that cognitive research employed to determine the components of a cognitive system. Five years after the model of Atkinson and Shiffrin was published, E. Tulving argued in favor of the partition of long-term memory into an episodic component and a semantic component. Later, in the 1980s, it was proposed that both episodic and semantic memory were part of a declarative or explicit component of memory having to do with conscious knowledge, but research also focused on an implicit, largely unconscious component of memory dealing with knowledge

related to cognitive, motor, and perceptual skills. Though there is no general agreement about the psychological reality of all these components and some authors consider the increasing number of them as a plain violation of the scientific principle of parsimony, a large group of psychologists views this tendency as a result of many years of analytic research and as a valuable way to make contact with brain research.

With the advent of cognitive science, the term cognitive architecture came to be used as a way to indicate that the structure of the cognitive system had some sort of primitive and permanent character. Search for this character divided the field of cognitive science into two conflicting views. On the one hand, the classical view considered the concept of the physical symbol as the building block of cognitive architecture and defined a cognitive system as a physical symbol system. On the other hand, a new connectionist view considered the concept of the neural network to be the basic unit of analysis for understanding the human cognitive system.

A physical symbol is a physical entity able to designate or refer to another entity that is the meaning of that symbol. For centuries this capacity to refer to something else was thought to belong only to the realm of ideas, and an idea was considered to be nonmaterial. As mentioned previously, the computer became a model for the human mind when scientists became aware that a computer was a symbol-manipulating machine rather than a machine dealing only with numbers. The physical symbol concept, by attributing to physical entities the capacity of having meaning, became a key concept in bridging the gap between brain and mind. In turn, a physical symbol system was defined as a system constituted by physical symbols. The adaptive control of thought (ACT) and the Soar systems, developed by John Anderson and Allen Newell, respectively, are two of the most influential current symbolic architectures.

The connectionist view represented an attempt to associate cognitive science with brain theory. Though the origin of the concept neural network goes back to Donald Hebb, and some interesting work on neural networks had already been done during the first years of cognitivism, the impulse that brought connectionism to the fore took place in the mid-1980s. Connectionists tried to build information-processing systems that were neurally inspired and called computation on such a system brain-style computation because they thought that the primitive concepts of a cognitive model should resemble the primitive units of the brain, neurons.

A neural network is composed of a set of elementary, neuron-like processing units connected to each other

in a specific way. The different strengths of the connections among units defines a pattern of connectivity that may change through experience according to a particular learning rule. An activation rule determines the way different inputs to a unit combine to determine the unit's state of activation, and an output function is in charge of mapping the unit state of activation into output. The only other component of a neural network is an environment within which the system can operate. Systems of this kind have been successfully used to explain performance on cognitive tasks as well as to model cognitive processes. For some time there were major disputes between defenders of symbolic and connectionist architectures. Nowadays, the arguments have abated, and many cognitive scientists make free use of both symbolic and neural network architectures depending upon the problem at hand. Neural networks seem best suited to model knowledge dependent on long, repetitive practice, as is usually the case in implicit knowledge. On the other hand, symbolic architectures appear preferable for modeling high-level cognitive processes such as reasoning or thought.

### 4.3. Representation

One of the main characteristics defining a particular system as cognitive is its capacity to reinstate a prior experience in the absence of a current external stimulus. We say that such a system owns the capacity to represent its environment.

Interest in mental representation as such was not a relevant feature of the first cognitive models. As discussed previously, they mainly dealt with mental structure and its components. Reference to mental representation was implicit in the use of terms such as encoding and recoding, but interest in the mental code itself had to wait until the early 1970s.

To understand what representation is we should differentiate the representing world from the represented world. The representing world has to somehow imitate the represented world, though not every possible aspect of the latter must be reflected in the former. For any representation, the following three features must be specified:

1. What aspects of the represented world are chosen to be encoded;
2. What elements of the representing world are doing the encoding;
3. What the correspondence is between the two worlds.

Research on mental images and memory brought to the forefront the very problem of mental representation. Was the representation responsible for mental images different from abstract semantic or syntactic representations? Were there analogical representations? These questions divided the field of cognitive psychology in two halves, one for and the other against analogical representations, a debate that still continues, although recent research in cognitive neuroscience seems to favor their existence. In the field of semantic memory, a plethora of representational formats for abstract knowledge were advanced within the framework of symbolic architectures. Prototypes, semantic features, semantic networks, schemata, and production systems were some of the main formats employed, but none attained general acceptance. Some investigators tried to build semantic representations based on a handful of primitive structures, but most models using symbolic architectures turned to local representations. In a local representation, there is a one-to-one correspondence between entities in the represented world and elements in the representing world.

Connectionist systems use distributed representations. In a distributed representation, each entity is represented by a pattern of activity distributed over many elements in the representing world, and each element represents many different entities. Distributed representations have proved to be very effective in dealing with aspects of cognition that were difficult to implement on symbolic architectures. Content-addressable and reconstructive memories, ability to generalize automatically to novel situations, and capacity to be modified according to a changing environment are genuine achievements of distributed architectures.

Cognitive science, symbolic cognitive science in particular, has been criticized for the way it has elaborated the notion of representation. Criticisms range from denying a particular characteristic of the representation, like that of being symbolic or abstract, to calling the entire notion into question. A common trait for these views is their concern about treating mind, body, and environment as coupled systems, dynamically interacting with each other; hence the term situated cognition is often used for this approach. It emphasizes that representations undergo continuous changes in tune with changes in the environment and encourages research on the changing patterns of behavior over time, using nonlinear differential equations to do the modeling. From this point of view, the best way to construct cognitive models is to build real agents. So far, the main contributions of this viewpoint

are to be found in the fields of perception and motor processes. To what extent this approach may be useful in understanding higher cognitive processes remains to be seen.

#### 4.4. Process

Processes are operations acting on a representation either to generate it or to transform it in some way. They may be involved in translating external input into a particular representation, modifying the representation itself, or generating output. A process may be considered a function mapping an input to an output.

Interest in elementary mental operations goes back to the Dutch physiologist F.C. Donders in 1868. He assumed that the time spent by an agent performing a task could be divided into a series of stages, each of which corresponded to an elementary component process of the task. By using simple tasks, supposedly consisting of few elementary components, and subtracting the time spent in performing one task from the time corresponding to the next more complex one (with an extra elementary component), Donders thought to have isolated the time for simple cognitive processes such as detection or discrimination. Two assumptions relevant for cognitive psychology underlay Donders' experimental program: the existence of elementary mental operations and their serial organization as component processes of a complex task.

The information-processing approach has tended to rely on the assumption that a small number of elementary mental operations are sufficient to specify the complexity of human cognitive performance. However, there is little agreement about what the elementary operations are and about the ultimate level of description appropriate for an elementary operation. Connectionist models offer an alternative because they use only activation and inhibition as elementary processes acting on neuron-like elementary units, but many researchers prefer descriptions in more intuitive terms referring to information-processing stages.

A central issue for information-processing theories is the way in which mental processes are assembled in order to perform a task. Early cognitive models assumed a serial organization of processes; they were carried out one after another, each process taking as input the output of the previous one, and starting operation when the preceding process had finished. It soon became apparent that many cognitive processes in perception, memory, and other cognitive areas could not be understood on the basis of serial

organization only; rather, they seemed to operate simultaneously, that is, in parallel. Parallel functioning of processes is typically associated with division of labor and modularity. A module, much like a subroutine in a computer program, is made up of a set of operations carrying out a particular assignment in the service of a larger task. It has been much debated whether the entire human mind or only some cognitive processes are modularly arranged. Nevertheless, there seems to be general agreement that modularity is a characteristic of low- rather than high-level cognitive processes. Connectionism has made parallel processing a distinctive feature of its approach. The term parallel distributed processing, often used as an alternative name for connectionism, highlights the two defining characteristics of this view, namely, distributed representation and parallel processing. Even so, many connectionist models group their simple, neuron-like units into layers, thus allowing for some serial organization of information processing stages.

An additional dimension introducing differences within cognitive theory has to do with the direction of the processing flow. Some theories and models emphasize bottom-up processing, in which the entire job is carried out by algorithms acting on input data to build increasingly complex representations. David Marr's influential theory of vision is a good example of a cognitive theory built from a bottom-up point of view. On the other hand, top-down processing emphasizes the feedback influence that complex cognitive component processes, such as goals, beliefs, and knowledge, exert on the functioning of simpler ones. A process is said to be data driven when its working is independent of feedback influences from goals, intentions, or other higher level processing stages. A process depending upon these influences is said to be conceptually driven. The most powerful cognitive models are usually interactive, meaning that they include both bottom-up and top-down processing.

Another frequently used classification of cognitive processes is the division between automatic and controlled processes. An automatic process usually refers to a sequence of component subprocesses that is triggered by a particular input. The sequence may include both bottom-up and top-down processing. An automatic process may be built-in as part of the physical implementation of a system, or its components may be assembled as a result of practice. Thus, to the extent that automaticity depends on practice, it should be considered a gradual rather than an all-or-none property for processes. On the other hand, when a new sequence of component

processes is required, top-down attentional influences are needed to do the assembling. In this case, the sequence is called a controlled process, meaning that it is under attentional control. Automaticity and dependence upon attentional control are complementary characteristics of mental processes; the more automatic a process is, the less attentional control it needs to act. Distinguishing between automatic and controlled processes has become very useful in psychological research. This distinction partially overlaps with that between explicit and implicit or between conscious and unconscious processes. Further research on this area will likely reveal fundamental characteristics of the human cognitive system.

#### 4.5. Methods

Cognitivism brought to psychology new research methods and new ways of approaching more traditional ones. Reaction time came to be one of the main dependent variables used to understand cognitive processes. As mentioned before, Donders' subtractive method and more sophisticated later developments, like S. Sternberg's additive factors method, served as a guide in that reaction time was not studied by itself, but as an index for making inferences about nonobservable processes. Accuracy measures were also developed with the same purpose. In 1954, signal-detection theory was applied to perception by W. Tanner and J. Swets, providing indexes of perceptual functioning uncontaminated by the observer's decision biases. Subsequently, signal detection was also applied to memory and became a valuable tool for studying diagnostic systems.

Modeling cognitive processes was a research method characteristic of the cognitive approach. Mathematical as well as flow-chart models were widely used in the early days, but computer models, either symbolic or connectionist, were soon considered the high road to understanding cognitive processes. Computer modeling helped psychologists become aware of cognitive processing intricacy and forced them to be specific about the mechanisms involved in a particular task.

Recently, methods from the realm of neuroscience have been incorporated into cognitive research. These days, electrophysiological and brain-imaging techniques allow online recording of brain activity while a person is performing specific cognitive tasks. These technological advances provide a unique opportunity to relate brain activity to cognitive performance and to establish connections between cognitive constructs and brain

structures and processes. An entirely new and promising field of research, known as cognitive neuroscience, has appeared, bringing new horizons for cognitive research.

### 5. COGNITIVIST INFLUENCE

The influence of cognitivism pervades every field of psychological research and spreads to the entire realm of social sciences. The term cognitive, considered vague and nonscientific during the behaviorist era, has now become respectable and scientific. Different approaches within cognitive science provide new ways of thinking about mental processes and intelligence. The study of human intelligence, once restricted to the statistical analysis of psychological tests, has now expanded to the working of mental machinery. The intelligence factors isolated by means of factor analysis are being studied in relation to cognitive constructs, thus bridging a long-lasting gap between differential and experimental psychology. As almost every article in this encyclopedia shows, cognitive factors are frequently mentioned as relevant causes in the explanation of different practical situations. We learn about cognitive ergonomics, cognitive behavior modification, cognitive neuropsychology, and other cognitive enterprises as truly influential approaches within applied psychology. Each one deserves to be treated separately.

#### See Also the Following Articles

Classical Conditioning ■ Cognition and Culture ■ Cognitive Skills: Training, Maintenance, and Daily Usage ■ Connectionism

#### Further Reading

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes, In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 89–195). New York: Academic Press.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
- Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience: The biology of the mind*. New York: W. W. Norton & Company.
- Gruber, H. E., & Vonèche, J. J. (1977). *The essential Piaget*. New York: Basic Books.
- Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Lawrence Erlbaum.
- McClelland, J., Rumelhart, D., & Hinton, G. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations*. Cambridge, MA: MIT Press.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Posner, M. I. (Ed.) (1989). *Foundations of cognitive science*. Cambridge, MA: MIT Press.
- Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. New York: Scientific American Library.
- Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.



# Collective Action

Bert Klandermans

Free University of Amsterdam, Amsterdam, The Netherlands

1. What Is Collective Political Action?
2. Why Do People Participate in Collective Political Action?  
Further Reading

## GLOSSARY

**collective action frames** Sets of collective beliefs that serve to create a frame of mind in which participation in collective action appears meaningful.

**collective identity** A place shared with other people.

**collective incentives** Incentives that relate to achievement of the movement's goals and the extent to which participation in a specific activity contributes to goal achievement.

**erosion of support** The process whereby people lose their sympathy for the movement in the course of a mobilization campaign.

**free rider** A person who refrains from participation, assuming that he or she will reap the benefits of the collective action anyway.

**grievances** Experiences of illegitimate inequality, feelings of relative deprivation, feelings of injustice, and/or moral indignation about some state of affairs.

**identity** Participation as an expression of identification with a group.

**ideology** Participation as a search for meaning and an expression of one's views.

**instrumentality** Participation as an attempt to influence the social and political environments.

**mobilization** The transformation of potential participants (sympathizers) into participants.

**nonconversion of support** The failure to convert sympathizers into participants because organizers failed to target them or were unable to motivate them.

**personal identity** The set of places that a person occupies in society.

**selective incentives** Incentives that relate to costs and benefits that differentiate between participants and nonparticipants.

On February 15, 2003, millions of people took to the streets in hundreds of cities in 60 countries throughout the world. In what was probably the largest collective action ever staged in human history, they protested against the imminent war in Iraq. Men and women of all ages expressed their aversion to the planned invasion. They were mobilized through the media, through organizations in which they were members, and by people who knew them. They came alone or with family and/or friends. They detested their government if it supported the war or applauded their government if it opposed the war. Obviously, these were not random samples of the population that walked the streets. The vast majority of the participants placed themselves at the left end of the political spectrum. They were predominantly white-collar workers or students who were highly educated. They were people who were interested in politics, and many had taken

part in demonstrations previously. The antiwar demonstrations were examples of an increasingly common form of political behavior, namely, collective action. What is collective action, why do people participate in such action, and how are these people mobilized? This article addresses these questions. Although the concept of collective action has a broader use, the article restricts itself to collective political action.

## 1. WHAT IS COLLECTIVE POLITICAL ACTION?

Collective political actions are challenges by people with common purposes and solidarity in interaction with elites and authorities. This definition includes two key elements that deserve some elaboration. First, collective political actions are challenges. They concern disruptive direct action against elites, authorities, other groups, or cultural codes. There is an obvious reason why this is the case. Collective political action is typically staged by people who lack access to politics. If they had access, there would be no need for the action. However, disruption forces authorities to pay attention to the claims brought forward. Second, collective political actions concern people with a common purpose and solidarity. Participants rally behind common claims. They want authorities to do something—to change a state of affairs or to undo changes. Such common claims are rooted in feelings of collective identity and solidarity. If these challenges gain sustainability, they constitute a social movement.

Collective political action has become much more frequent over recent decades. For example, in 1975, 22% of people in Britain signed at least one petition; in 1990, the figure was 75%. In 1974, 9% of people in Britain participated in at least one demonstration; in 1990, the figure was 25%. Between 1979 and 1993, the number of protest events in the French city of Marseille more than doubled, from 183 to 395 events per annum. During that same period, considerable increases in protest events in Germany were reported, a result that was confirmed in 1995 by Kriesi and colleagues' now classic study on new social movements in Europe. At the same time, diffusion to broader sectors of the population was witnessed. Across age groups and gender lines, from the left to the right, among workers and students, and in Western and non-Western societies alike, collective political action has become a common phenomenon.

## 2. WHY DO PEOPLE PARTICIPATE IN COLLECTIVE POLITICAL ACTION?

The social psychology of protest suggests three fundamental motives for participation in collective political action in that people may want to (a) change their circumstances, (b) act as members of their group, and/or (c) give meaning to their world and express their views and feelings. This article suggests that these three motives together account for most of the motivation to participate in collective political action in a society refers to these motives as instrumentality, identity, and ideology. Instrumentality refers to participation as an attempt to influence the social and political environment. Identity refers to participation as an expression of identification with a group. Ideology refers to participation as a search for meaning and an expression of one's views. This article does not suggest that these are mutually exclusive motives or are competing views on collective action participation, although some parties in the debates in the literature seem to have taken that position. However, this article concurs with those who argue that approaches that neglect any of these three motives are fundamentally flawed. To be sure, individual participants may participate due to a single motive, but all three motives are needed to understand why people take part in collective political action. Of course, that raises the question of how the three motives interact in fostering participation. The article elaborates on this in the following discussion of the motivation to participate.

### 2.1. The Motivation to Participate in Collective Action

#### 2.1.1. Instrumentality

Instrumentality begins with dissatisfaction, be it the experience of illegitimate inequality, feelings of relative deprivation, feelings of injustice, moral indignation about some state of affairs, or a so-called suddenly imposed grievance, that is, an event imposed on people (e.g., highway construction in the neighborhood). Social psychological grievance theories, such as relative deprivation theory and social justice theory, have tried to specify how and why grievances develop. Despite the fact that grievances are at the roots of collective political action, they have not featured prominently in social movement literature since the early 1970s. Resource mobilization theory and political process theory, the

two approaches that have dominated the field since those days, have always taken as their point of departure that grievances are ubiquitous and that the key question in action participation research is not so much why people are aggrieved as it is why aggrieved people participate. The instrumentality paradigm holds that participants are people who believe that they can change their social and political environments at affordable costs.

The perceived costs and benefits of collective action can be distinguished in selective and collective incentives. Selective incentives relate to costs and benefits that differentiate between participants and nonparticipants. People are spending time and/or money only if they participate, people run the risk of being beaten up by the police only if they participate, people's friends will blame them for not participating only if they stay home, and so on. Sometimes, movement organizations try to make participation more attractive by providing selective benefits (e.g., a popular music group, a train ticket to the city where the demonstration is held, a T-shirt). Authorities or opponents, for their part, can try to make participation less attractive by imposing costs on participants. Collective incentives are related to achievement of the movement's goals and the extent to which participation in a specific activity contributes to goal achievement. Obviously, it is not enough for a goal to be important to a person. What is also needed is some likelihood of success. The problem with collective action is that it is difficult to know to what extent an activity will have any influence on authorities. In any event, chances are low that an activity will have any impact if only a few people participate. Therefore, the likelihood of success depends on the expected behavior of others. If too few people participate, it is unlikely that the activity will make any difference. As a consequence, expectations about the behavior of others play an important role in the decision to participate. If someone expects that only a few people will participate, that person's motivation to participate will be low. In a way, the expectation about other people's behavior functions as a self-fulfilling prophecy. If people believe that few people will participate, they will not be motivated to participate and, thus, will help make their own expectation come true.

Instrumentality presupposes an effective movement that is able to enforce some wanted changes or at least to mobilize substantial support. Making an objective assessment of a movement's impact is not an easy task, but movement organizations will try to convey the image of an effective political force. They can do so

by pointing to the impact they have had in the past or to the powerful allies they have. Of course, they may lack all of these, but they might be able to show other signs of movement strength. A movement may command a large constituency as witnessed by turnouts at demonstrations, by membership figures, or by large donations. It may be composed of strong organizations with strong charismatic leaders who have gained respect. Instrumentality also implies the provision of selective incentives. Movements may vary considerably in the selective incentives for participation they provide. Obviously, this is also a matter of the resources that a movement commands. Surprisingly little comparative information is available on the resources that movements have at their disposal. In a similar vein, systematic documentation is lacking on the ways in which the larger political system and the alliances and opponents of movement organizations influence movement participation. In 1978, Tilly coined the now classic terms "repression" and "facilitation" to distinguish between political systems that increase and decrease the costs of participation. Indeed, repressive political environments may increase the costs of participation considerably. People may lose friends, they may risk their jobs or otherwise jeopardize their sources of income, they may be jailed, and they may even lose their lives.

Viewed from an instrumental perspective, a solution to the dilemma of collective action must be found. It was through the work of Olson that students of movement participation became aware of the problem. In 1968, Olson published his *The Logic of Collective Action*. The core of the book was the argument that rational actors will not contribute to the production of a collective good unless selective incentives persuade them to do so. Collective goods are characterized by "jointness of supply." That is, if they are made available to one person, they become available to everybody irrespective of whether people have contributed to the production of the collective good (e.g., a law against discrimination, measures against pollution). Therefore, according to Olson, a rational actor will choose to take a free ride unless selective incentives (i.e., costs or benefits that are made contingent on participation in the production of the collective good) prevent him or her from doing so. Olson's argument was soon applied to social movement participation. It helped to explain why so often people do not participate in social movements despite their interest in the achievement of the movements' goals. Movement scholars argued that movement goals typically are collective goods. If a goal

is achieved, people will enjoy the benefits irrespective of whether they have participated in the effort. In view of a goal for which achievement is uncertain but for which benefits—if they materialize—can be reaped anyway, rational actors will take a free ride.

However, social movement scholars quickly discovered that reality is more complex than Olson's simple model suggests. The problem with Olson's logic of collective action is that it provides an explanation for why people do not participate but fares poorly in explaining why people do participate. Moreover, Olson's solution that people participate for selective incentives is fundamentally flawed because it does not give a satisfactory answer to the question of where the resources needed to provide selective incentives come from. If these resources must be collected from individual citizens, the same collective action dilemma arises again. This is not to say that selective incentives are irrelevant; however, in the final instance, they cannot solve the collective action dilemma. In other words, if collective and selective incentives do not provide a sufficient explanation of movement participation, what else might make the difference? A recurring criticism was that Olson's model assumes that individuals make their decisions in isolation, that is, as if there are no other people with whom they consult, with whom they feel solidarity, and by whom they are kept to their promises. This pointed to the significance of collective identity as a factor in movement participation.

### 2.1.2. Identity

It soon became clear that instrumentality, or movement participation to achieve some external goal (social or political change), was not the only reason to participate. After all, many external goals are reached only in the long run if at all. Similarly, when it comes to material benefits, costs frequently outweigh benefits. Apparently, there is more involved in being a movement participant than perceived costs and benefits. Indeed, one of those motives relates to belonging to a valued group.

Identity can be described succinctly as a place in society. People occupy many different places in society. They are students, housewives, soccer players, politicians, farmers, and so on. Some of those places are exclusive, that is, occupied by only a small number of people (e.g., the members of a soccer team). Others are inclusive and encompass large numbers of people (e.g., Europeans). Some places are mutually exclusive (e.g., male/female, employed/unemployed), some are nested (e.g., French/Dutch/German vs European), and some

are crosscutting (e.g., female and student). All of these different roles and positions that a person occupies form his or her personal identity. At the same time, every place a person occupies is shared with other people. The author of this article is not the only professor of social psychology, nor is he the only Dutch or the only European. He shares these identities with other people, a fact that turns them into collective identities. Thus, a collective identity is a place that is shared with other people. This implies that personal identity is always collective identity at the same time. Personal identity is general, referring to a variety of places in society, whereas collective identity is always specific, referring to a specific place.

Most of the time, collective identities remain latent. Self-categorization theory hypothesizes that an individual may act as a unique person (i.e., display his or her personal identity) or as a member of a specific group (i.e., display one of the many collective identities he or she has depending on contextual circumstances). Contextual factors may bring personal or collective identity to the fore. Obviously, this is often not a matter of free choice. Circumstances may force a collective identity into awareness whether people like it or not. The Yugoslavian and South African histories have illustrated this dramatically, and there are other equally or even more dramatic examples throughout human history. But collective identities can also become significant in less extreme circumstances. For example, consider the possible effect of an announcement that a waste incinerator is planned next to a certain neighborhood. Chances are that within very little time, the collective identity of the people living in that neighborhood will become salient.

Self-categorization theory proposes that the more people identify with a category, the more they are prepared to employ a social category in their self-definition. Identification with a group makes people more prepared to act as a member of that group. This assertion refers, of course, to identity strength. Social identity literature tends to neglect that real-world identities vary in strength. But identifying more or less strongly with a group may make a real difference, especially in political contexts. Moreover, strong identities may be less affected by context. Following this reasoning, we may expect that strong identities make it more likely that people act on behalf of their group. The basic hypothesis regarding collective identity and movement participation is fairly straightforward. A strong identification with a group increases the likelihood of participation in collective political action on

behalf of that group. The available empirical evidence supports this assumption overwhelmingly.

Movements offer the opportunity to act on behalf of one's group. This is most attractive if people identify strongly with their group. The more farmers identify with other farmers, the more appealing it is to take part in farmers' protests. The more women identify with other women, the more inviting it is to participate in the women's movement. The more gay people identify with other gay people, the more they are attracted by the possibility of taking part in the gay movement. In addition to the opportunity to act on behalf of the group, collective political action participation offers the opportunity to identify with the movement's cause, the people in the movement, the movement organization, the group in which one is participating, and/or the leader of the movement. Not all of these sources of identification are always equally appealing. Movement leaders can be more or less charismatic, and the people in the movement or in someone's group can be more or less attractive. Moreover, movements and movement organizations may be, and in fact often are, controversial. Hence, becoming a participant in a movement organization does not mean taking a respected identity on oneself—quite the contrary. Within the movement's framework, of course, this is completely different. There, the militant does have the status that society is denying him or her. And for the militant, of course, in-group/out-group dynamics may turn the movement organization or group into a far more attractive group than any other group "out there" that is opposing the movement. Indeed, it is not uncommon for militants to refer to their movement organization as a second family, that is, as a substitute for the social and associative life that society is no longer offering them. Movement organizations not only supply sources of identification but also offer all kinds of opportunities to enjoy and celebrate the collective identity—marches, rituals, songs, meetings, signs, symbols, common codes, and so on.

### **2.1.3. Ideology**

The third motive—wanting to express one's views—refers at the same time to a long-standing theme in the social movement literature and to a recent development. In classic studies of social movements, the distinction was made between instrumental and expressive movements or protests. During those days, instrumental movements were seen as movements that aimed at some external goal (e.g., implementation of citizenship rights). Expressive movements, on the other hand, were

movements that had no external goals and for which participation was a goal in itself (e.g., expression of anger in response to experienced injustice). Movement scholars felt increasingly uncomfortable with the distinction because it was thought that most movements had both instrumental and expressive aspects and that the emphasis on the two could change over time. Therefore, the distinction lost its use. Recently, however, the idea that people might participate in movements to express their views has received renewed attention. Attention arose this time from movement scholars who were unhappy with the overly structural approach of resource mobilization and political process theory. These scholars began to put an emphasis on the creative, cultural, and emotional aspects of social movements such as music, symbols, rituals, narratives, and moral indignation. People are angry, develop feelings of moral indignation about some state of affairs or some government decision, and wish to make these feelings known. They participate in social movements not only to enforce political change but also to gain dignity in their lives through struggle and moral expression.

Ideology has a significant affective component. Acting on one's ideology is deemed to be one of the fundamental motives of action participation and is necessarily charged with emotion. Appraisal and action are socially constructed. That is, they are formed in interpersonal interaction, especially in the case of politically relevant emotions. Cultural and historical factors play an important role in the interpretation of the state of affairs by which politically relevant emotions are generated. Obviously, appraisal can be manipulated. Activists work hard to create moral outrage and anger and to provide a target against which these can be vented. They must weave together a moral, cognitive, and emotional package of attitudes. But emotions also play an important role in the ongoing activities of movements. The literature distinguishes two kinds of collective emotions—reciprocal emotions and shared emotions—that reinforce each other. Each measure of shared outrage against an injustice reinforces the reciprocal emotion of fondness for others precisely because they feel the same way. Conversely, mutual affection is a context in which newly shared emotions are easily created. Anger and indignation are emotions that are related to a specific appraisal of the situation. At the same time, people might be puzzled by some aspects of reality and may try to understand what is going on. They may look for others with similar experiences, and a social movement may provide an environment in which to exchange experiences, tell their stories, and express their ideologies.

Social movements play a significant role in the diffusion of ideas and values. In 1998, Rochon made the important distinction between “critical communities” (where new ideas and values are developed) and “social movements” (which are interested in winning social and political acceptance for those ideas and values). “In the hands of movement leaders, the ideas of critical communities become ideological frames,” stated Rochon, who continues to argue that social movements are not simply extensions of critical communities. After all, not all ideas developed in critical communities are equally suited to motivate collective action. Social movement organizations, then, are carriers of meaning. Through processes such as consensus mobilization and framing, they seek to propagate their definitions of the situations to the public at large. A study of flyers produced by the various groups and organizations involved in the protests against the International Monetary Fund and the World Bank in Berlin, Germany, is an excellent example in this respect. The study showed how links were constructed between the ideological frame of the organizers of the demonstration and those of the participating organizations so as to create a shared definition of the situation. In the social movement literature, such definitions of situations have been labeled “collective action frames.”

Social movements do not invent ideas from scratch; rather, they borrow from the history of ideas. They build on an ideological heritage as they relate their claims to broader themes and values in society. In so doing, they relate to societal debates that have histories of their own, and those histories are usually much longer than those of the movements themselves. In 1992, Gamson referred to the “themes” and “counterthemes” that, in his view, exist in every society. One such pair of a theme and counter-theme that Gamson mentioned is that of “self-reliance” versus “mutuality,” that is, the belief that individuals must take care of themselves versus the belief that society is responsible for its less fortunate members. Klandermans has demonstrated how, in The Netherlands, these two beliefs became the icons that galvanized debate and spurred protest over disability payments. Self-reliance became the theme of those who favored restrictions in disability payment, whereas mutuality was the theme of those who defended the existing system.

## 2.2. The Transformation of Potentiality into Action

People who are potentially prepared to take part in collective action do not automatically participate.

Potential must be transformed into action. This is in fact what most mobilization campaigns are about—targeting potential participants and turning them into actual participants. Obviously, a campaign will never be 100% successful in that regard. In the course of a campaign, people may lose their sympathy for the movement. Klandermans has labeled this “erosion,” a phenomenon that is more likely to occur if counter-movements and/or authorities are mounting counter-campaigns. On the other hand, organizers might fail to convert sympathizers into participants because they fail to target or motivate them. Klandermans has called this “nonconversion.” The rest of this section focuses on targeting.

Targeting sympathizers involves finding answers to two strategic questions: “Who are the sympathizers?” and “How can they be reached?” Every attempt to persuade must find answers to these two questions. Indeed, sophisticated movement organizations such as Greenpeace have learned to employ many of the strategies that have been developed in advertising. Social networks are of crucial importance in this regard. Movement organizations have two options here in that they can (a) try to co-opt existing networks or (b) build new networks. Both strategies are mobilization efforts in themselves. Co-optation is the easier strategy of the two because it builds on existing commitments to organizations and networks that are part of the movement organization’s alliance system. However, there are risks. The co-opted organizations may use the campaigns for their own ends, or their leadership may—for whatever reason—decide not to collaborate. The latter makes it more difficult for the rank-and-file to cooperate. Yet co-opting existing networks, such as churches, unions, political parties, and youth organizations, is frequently applied, if only because doing so implies an answer to both strategic questions. On the one hand, it works from the assumption that most members of the movement organization sympathize with the movement. On the other hand, it assumes that these sympathizers can be reached through the organization’s networks.

Building new networks implies recruiting people who are willing to spend considerable amounts of time for a prolonged period as movement activists. Therefore, it will require more effort on the part of the organization than will co-opting existing networks. But once the new networks are established, they will be more reliable. The recruitment of movement activists is a process that is determined, on the one hand, by factors that influence who is being asked and, on the

other, by factors that influence who agrees to serve as activists when asked. As for the first type of factors, a crucial determinant is someone's position in networks linked to the movement organization or, more specifically, to the movement organizer who is undertaking the recruitment effort. Movement organizers tend to recruit first among the people they know, and this often suffices. After all, they do not need many activists. They need a few to maintain the networks, and once they have those few, the return on having additional activists diminishes rapidly. Indeed, it has been argued that long-term activism is one form of activism that must cope with free-rider behavior. The latter point is important when deciding whether to serve as an activist or not. Klandermans has maintained that the people who are asked to serve as activists understand perfectly well that they are giving a free ride to most of the sympathizers to the movement, but they are prepared to do so because they care. Only people who really care a lot are prepared to sacrifice for the others. They make the effort because they believe that "If [they] don't do it, nobody else will." Klandermans has asserted that only people who strongly support the movement's ideology, and who feel responsible for maintaining and proselytising the movement and its ideology, are prepared to make the effort of serving as movement activists.

### See Also the Following Articles

Ideological Orientation and Values ■ Organizational Participation

### Further Reading

- Gamson, W. A. (1992). *Talking politics*. Cambridge, UK: University of Cambridge Press.
- Klandermans, B. (1997). *The social psychology of protest*. Oxford, UK: Blackwell.
- Klandermans, B., & Staggenborg, S. (Eds.). (2002). *Methods of social movement research*. Minneapolis: University of Minnesota Press.
- Kriesi, H., Koopmans, R., Duyvendak, J. W., & Guigni, M. (1995). *New social movements in Western Europe: A comparative analysis*. Minneapolis: University of Minnesota Press.
- McAdam, D., Tarrow, S., & Tilly, C. (2001). *Dynamics of contention*. Cambridge, UK: Cambridge University Press.
- Olson, M. (1968). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Rochon, T. R. (1998). *Culture moves: Ideas, activism, and changing values*. Princeton, NJ: Princeton University Press.
- Tarrow, S. (1998). *Power in movement: Social movements, collective action, and mass politics in the modern state* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Tilly, C. (1978). *From mobilization to revolution*. Reading, MA: Addison-Wesley.



# Color Blindness

Julio A. Lillo and Humberto Moreira

Complutense University, Madrid, Spain

1. Introduction
2. Using the Anomaloscope to Diagnose: Photoreceptors and Types of Color Blindness
3. Causes of Color Blindness
4. Color Blindness and Psychological Intervention  
Further Reading

## GLOSSARY

**anomalous trichromatism** Group of color blindness pathologies caused by disturbances in the functionality of a certain type of cone; there are three types of anomalous trichromatism: protanomaly (when the problem is related to the cones that are more sensitive to long wavelengths), deuteranomaly (when the alteration is related to the cones that are more sensitive to medium wavelengths), and tritanomaly (when the alteration is related to the cones that are more sensitive to short wavelengths).

**color blindness** Term used to name a set of color perception anomalies; the term is not totally accurate because most color-blind people do not have “black and white” vision but rather experience a more or less reduced gamut of chromatic sensations.

**daltonisms (red–green pathologies)** Term used to name the group of color blindness associated with the X chromosome; it includes two varieties of protan alterations (protanopia and protanomaly) and two varieties of deutan alterations (deuteranopia and deuteranomaly).

**dichromatisms** Group of color blindness pathologies caused by the absence of functionality in a certain cone type

(usually, the cones do not have the proper photopigment); protanopia is the dichromatism related to problems in the cones that are more sensitive to long wavelengths, deuteranopia is the dichromatism related to the cones that are more sensitive to medium wavelengths, and tritanopia is the dichromatism related to the cones that are more sensitive to short wavelengths.

**metamers** Stimuli that, being physically different, produce an identical chromatic experience.

**monochromatisms** Group of color blindness pathologies; monochromats have functionality in only one specific type of photoreceptor, and they lack color perception.

**opsin** Protein present in the receptor photopigments.

**scotopic vision** The customary vision in normal-sighted people at night; it is mediated by the retina rods and does not allow color perception (and, consequently, is a monochromatic type of vision).

The scientific definition of color blindness is explained, emphasizing that this denomination is not only applied to people without color experiences (monochromats) but also to all with no-common color experiences (dichromats and anomalous trichromats). Clinical nomenclature in color-blind diagnosis is described, specifying the causal origin of each color-blind type. The different action types available to applied psychologists to help color-blind people are described (diagnosis, perceptual problems detection, use of specific information, creation-selection of accurate materials, etc.).

## 1. INTRODUCTION

The existence of metamer stimuli (i.e., physically different but perceptually identical) has psychological implications that are essential to understand the functional logic of color perception in both normal-sighted and color-blind people. Figure 1 shows the spectracolorimetric measurement of two metamers. The upper portion (Fig. 1A) is a physical description of the light reflected by a white shirt. The

middle portion (Fig. 1B) corresponds to the light from this object as reproduced on a conventional television screen. It is easy to observe considerable differences between these two physical stimuli that, nonetheless, produce the same perceptual experience (“white”).

The existence of metamers casts serious doubts on what could be called the “commonsense theory of color perception and color blindness.” This theory can be summarized in the following three ideas:

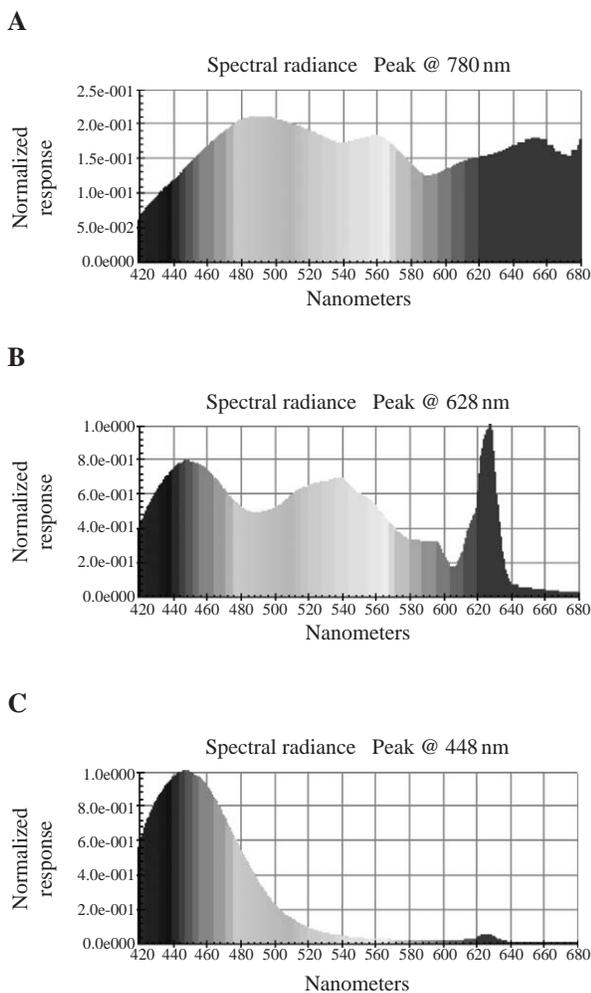
1. Color is a physical property of light (e.g., the light reflected by blood is red, the light reflected by a lawn is green).

2. The function of our visual system consists exclusively of detecting such properties from the stimulation that reaches the retina.

3. Color blindness would derive from the visual system’s limitations to detect the presence of some color (e.g., difficulty in detecting the presence of red).

From the viewpoint of the commonsense theory, the existence of metamers implies that all humans are color-blind because they experience the same color in response to physically different stimuli. It must be emphasized that such stimuli generate different color experiences when presented to other animal species (e.g., the stimuli from Figs. 1A and 1B are not metamers for bees and pigeons). Consequently, it can be concluded that the essential difference between normal-sighted and color-blind people is quantitative and not qualitative: All of them have metamers, but color-blind people have more or them. Although this conclusion is correct, a different conclusion can be inferred regarding the main premise of the commonsense theory because color is not a physical property of light. Instead, it is a property of the perceptual response generated in reaction to light, as was explicitly stated by Isaac Newton in his book *Optics*: “The rays, to speak properly, are not coloured. In them there is nothing else than a certain power and disposition to stir up a sensation of this or that colour.”

Following Newton’s ideas, the next section reveals the nature of the “certain power and disposition” present in physical stimuli and, even more important, how our visual system produces “a sensation of this or that colour” in response to them. Subsequently, it will be easy to conclude that the most important difference between normal-sighted and color-blind people is that the latter have a reduced capacity to detect energetic imbalances in light and, consequently, to produce color experiences.



**FIGURE 1** Spectracolorimetric measurements of three stimuli. Horizontal axis indicates the wavelength in nanometers (1 nm =  $10^{-9}$  meters). Vertical axis indicates the relative quantity of energy. (A) White shirt. (B) White shirt on a television screen. (C) Primary blue screen.

## 2. USING THE ANOMALOSCOPE TO DIAGNOSE: PHOTORECEPTORS AND TYPES OF COLOR BLINDNESS

Color perception is the result of a complex visual processing sequence, initiated in the retina photoreceptors, that involves a large number of brain centers. Consequently, any anomaly in this sequence can generate some type of color blindness. However, most of the problems related to color perception are caused by some functional anomaly in a photoreceptor type. Consequently, this analysis centers on these anomalies.

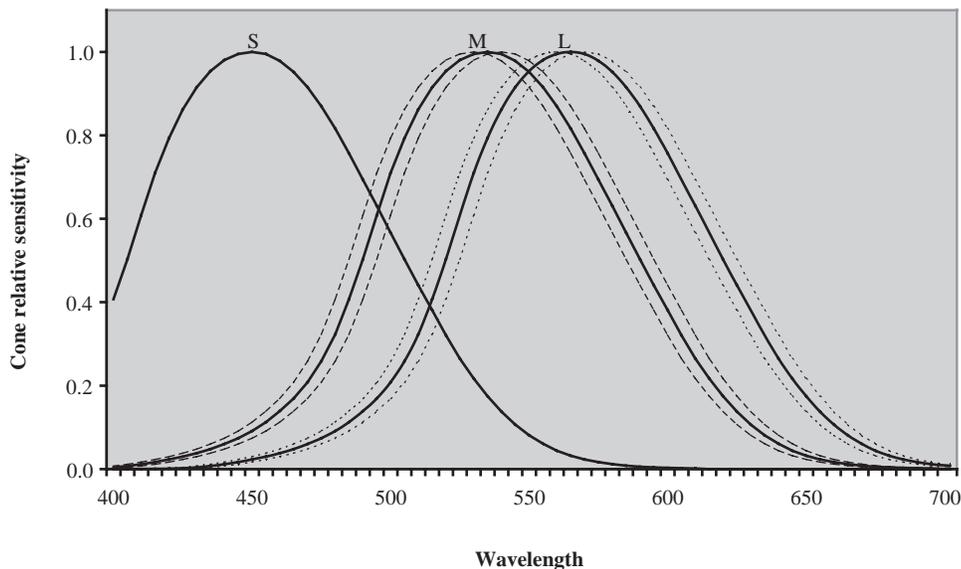
There are two types of photoreceptors in the human retina: cones and rods. Cones are especially relevant to understanding color perception and can be divided into three types (continuous curves in Fig. 2), depending on the relative position of their spectral sensitivity curves. In the standardized scientific nomenclature, these types are called “L” (long), “M” (medium), and “S” (short), depending on the portion of the spectrum that produces larger responses for each type of cone. Taking into account the relative curve position and counting from right to left, the clinical nomenclature calls the three cone types “proto,” “deutera,” and “trita,” respectively (corresponding to the Greek words for “first,” “second,” and “third”). Recent genetic evidence indicates that, at least for the L and M cones, the spectral response curves

present minor interindividual differences among normal-sighted people (discontinuous curves in Fig. 2).

The analysis of cone responses is useful to understand (a) the origin of metamers, (b) their more widespread presence among color-blind people and the various types of color blindness, and (c) how they could be diagnosed using an anomaloscope. Each of these aspects is considered in turn.

### 2.1. The Origin of Metamers

Despite their physical differences, the two metamers in Figs. 1A and 1B share an essential characteristic: They both present important energy accumulations in the three spectrum portions (long, medium, and short). That is why both stimuli generate high responses in the three cone types. Cone response similarity is the cause of the achromatic sensation generated (“white”). On the other hand, the stimulus represented in Fig. 1C accumulates energy in only one spectrum portion (the short one) and, consequently, generates an imbalance in the relative responses of the three cone types (the S cones have a larger response). This imbalance causes a chromatic sensation (“blue”). To summarize, first, the energetically unbalanced stimuli (more energy in any spectrum portion) tend to generate imbalances in the relative responses of the three cone types and, consequently, the generation of hue



**FIGURE 2** Cone spectral sensitivity curves. Horizontal axis indicates the wavelength in nanometers. Vertical axis indicates the relative sensitivities of the cones. S = cones most sensitive to short wavelengths; M = cones most sensitive to medium wavelengths; L = cones most sensitive to long wavelengths.

experiences (e.g., blue, green, purple). Second, the specific hue experienced depends on the type of imbalance detected by the cones. Third, all physical stimuli that generate the same pattern of response in the cones are metamers. For example, all stimuli that produce the same predominance of the S cones over the other cones will generate the same blue perception, and all stimuli that produce high and similar responses in the three cone types will generate the same white perception.

People with only one photoreceptor type lack the ability to detect energetic imbalances and are called “monochromats.” Their vision is very similar to that experienced by normal-sighted people when watching a black and white film. The psychophysical characteristic of these individuals is that, by performing the necessary intensity adjustments, they can experience any pair of wavelengths as metamers. Equation (1) synthesizes the cause of this fact. As can be observed, the response ( $R$ ) in the type of cone available depends on the relative wavelength visual efficacy ( $EF_\lambda$ ) and physical intensity ( $I_\lambda$ ) of the stimulus:

$$R = EF_\lambda \cdot I_\lambda. \quad (1)$$

## 2.2. Metamers and Types of Color-Blind People

The term “scotopic vision” refers to the kind of vision mediated by rods. When the illumination level is reduced and the cones become inoperative, vision depends on the rods and all humans become monochromats because they see with only one type of photoreceptor. Consequently, all of the variations have a quantitative characteristic and can be described using Eq. (1). Monochromatic vision is the only type of vision available to people who lack cones (rod monochromats). In addition, these individuals have severe acuity problems (similar to normal-sighted people at night, they cannot detect subtle details) and must use sunglasses permanently to reduce the amount of light reaching their eyes. Also considered monochromats are individuals with only one cone type (the S cone), although they do not have photophobia or acuity problems.

Strictly speaking, only monochromats should be called “color-blind” people because they are the only ones whose visual world is similar to that of normal-sighted people when watching a black and white film. However, monochromats comprise only a small portion of the people who are usually called color-blind.

People lacking a certain type of cone are dichromats. In the clinical taxonomy, this condition is indicated by

using the suffix “anope” after the cone denomination. Thus, people lacking L cones (protocones) are protanopes, people lacking M cones are deuteranopes, and people lacking S cones are tritanopes. To reach a better understanding of dichromats’ color limitations, a brief description of the most prestigious diagnostic test, the Nagel analytic anomaloscope, is presented.

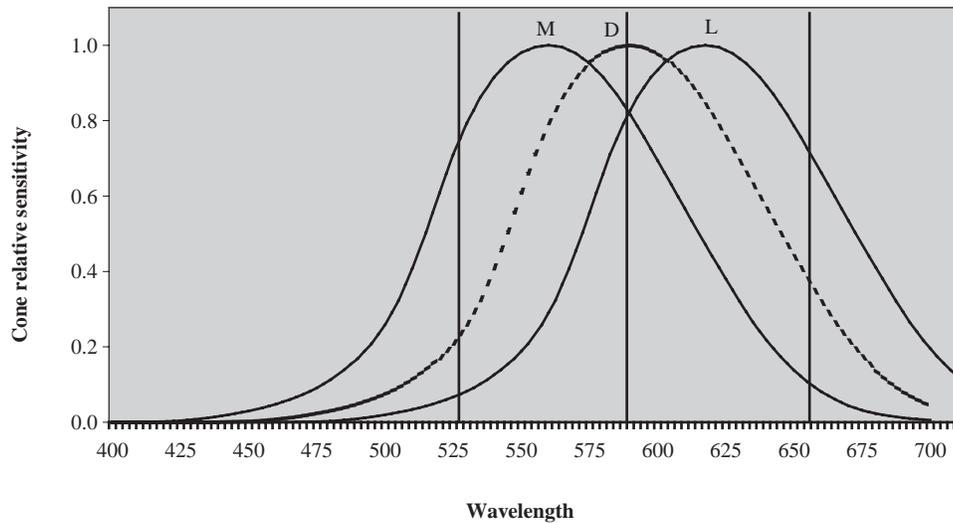
## 2.3. Diagnosing Dichromatism and Anomalous Trichromatism with an Anomaloscope

The Nagel anomaloscope uses three monochromatic stimuli of 546, 589, and 670 nm (vertical lines in Fig. 3). These values are out of the operative range of the S cones (not represented in Fig. 3). Consequently, the next analysis can focus on the other two cone types: the M cones (thick left curve in Fig. 3) and the L cone (thick right curve in Fig. 3).

The anomaloscope shows a circle divided into two halves. One half is used to present the reference stimulus, a monochromatic stimulus with a wavelength of 589 nm (represented by the central vertical line in Fig. 3). To make the explanation simpler, assume that presentation of the reference stimulus generates similar activation levels in both cone types (L and M) and that, because of this, a yellow experience is generated. The second half of the circle presents a mixture of 546- and 670-nm light (represented, respectively, by the left and right vertical lines in Fig. 3). For a normal-sighted person, a 546-nm monochromatic stimulus generates more response in the M cones than in the L cones and, consequently, the perception of green. On the other hand, the 670-nm monochromatic light is experienced as reddish because it generates a larger response in the L cones than in the M cones.

The use of a Nagel anomaloscope requires adjusting the quantities of both mixture components (the greenish and the reddish light) to match the reference stimulus. For a normal-sighted person, this means that the mixture should similarly activate the M and L cones and, therefore, produce the same yellow experience as does the reference stimulus. These adjustments are performed very precisely because every variation in the relative quantities of the mixture components will produce a greenish or reddish hue and, consequently, an easily noticeable difference in relation to the reference color.

When carrying out Nagel anomaloscope adjustments, the more common dichromat types (protanopes and



**FIGURE 3** Anomaloscope stimuli and cone spectral curves. Horizontal axis indicates the wavelength in nanometers. Vertical axis indicates the relative sensitivities of the cones. M = cones most sensitive to medium wavelengths; L = cones most sensitive to long wavelengths; D = curve that substitutes one of the two previous ones in anomalous trichromats.

deuteranopes) are not as consistent as normal-sighted people. This is because the three stimuli used act on the same cone type, that is, the only one they have (remember that S cones are irrelevant to this analysis). Consequently, the situation is very similar to the one described by Eq. (1); the response ( $R$ ) generated by the reference stimulus can be metamized by any of the two mixture stimuli presented individually, or by any combination of both mixture stimuli (given the accurate quantities).

Anomalous trichromats are the next group of color-blind people to be considered. This group consists of all people who, like normal-sighted people, have three cone types in their retinas (this is why they are trichromats) but whose spectral curve responses are somewhat different from those represented in Fig. 2 (this is why they are anomalous). The most important consequence of this is that anomalous trichromats perceive as metamers certain stimuli perceived as different by normal-sighted people.

Nowadays, it is commonly assumed that the anomalous trichromat's limitations in differentiating colors, and the peculiar way in which these are perceived, derive from an exaggerated overlapping of the spectral response curves of two cone types. To understand why, the D curve in Fig. 3 must be used to substitute one of the two continuous curves (the M or L curve). With the new pair of curves, the relative differences in the response of both cone types to any stimulus are reduced, as is the ability to discriminate colors—the more overlapping there is, the less ability

to discriminate. At the limit, a full overlapping of curves would make both cone types functionally equivalent, and these people would behave like dichromats.

### 3. CAUSES OF COLOR BLINDNESS

Most color blindness is genetic in origin. In this group, the alterations associated with problems in the L cones (protanopia and protanomalies) or in the M cones (deuteranopia and deuteranomalies) are collectively called “daltonisms” or “red–green problems.” These problems affect many more men (5–10%) than women (<0.9%) because they are associated with genetic alterations in recessively transmitted X chromosomes. Consequently, men will express red–green problems when there are anomalies in the only X chromosome they have. On the other hand, a woman must have problems in both X chromosomes to be daltonic. With only one affected chromosome, they will show a normal use of colors but can transmit the red–green disturbance to their offspring.

The most important gene in relation to the opsin characteristic of S cones is not located in sexual chromosomes; instead, it is located in the seventh pair of chromosomes. Consequently, problems associated with these cones have a similar incidence in men and women. These types of color blindness are transmitted dominantly, but they affect a very small number of people (<0.05%).

**TABLE I**  
**Comparison of Acquired and Congenital Color Vision Deficiencies**

<i>Acquired color vision defect</i>	<i>Congenital color vision defect</i>
Onset after birth	Onset at birth
Monocular differences in the type and severity of the defect occur frequently	Both eyes are equally affected
Color alterations are frequently associated with other visual problems such as low acuity and reductions in the useful visual field	The visual problems are specific to color perception; there are no problems with acuity (except in rod monochromatism) or visual field
The type and severity of the deficiency fluctuate	The type and severity of the defect are the same throughout life
The type of defect might not be easy to classify; combined or non-specific defects occur frequently	The type of defect can be classified precisely
Predominantly either protan or deutan	Predominantly tritan
Higher incidence in males	Same incidence in both sexes

Although the chromosomes associated with rod monochromatism have not yet been identified, this type of problem is known to be transmitted recessively. Consequently, the affected gene must be located in both chromosomes to produce problem expression.

Defective color vision can be acquired by any kind of agent with the capacity to disturb the normal functioning of the visual system—the result of ocular pathology, intracranial injury, the use of therapeutic drugs, aging, and so on. Frequently, color disturbances are initial symptoms of pathologies and can be used for early detection, indicating when therapeutic measures should begin or when they should be discontinued. Table I reviews the most important differences between acquired and congenital color deficiencies.

## 4. COLOR BLINDNESS AND PSYCHOLOGICAL INTERVENTION

### 4.1. Problems and Solutions: A Case of Protanopia

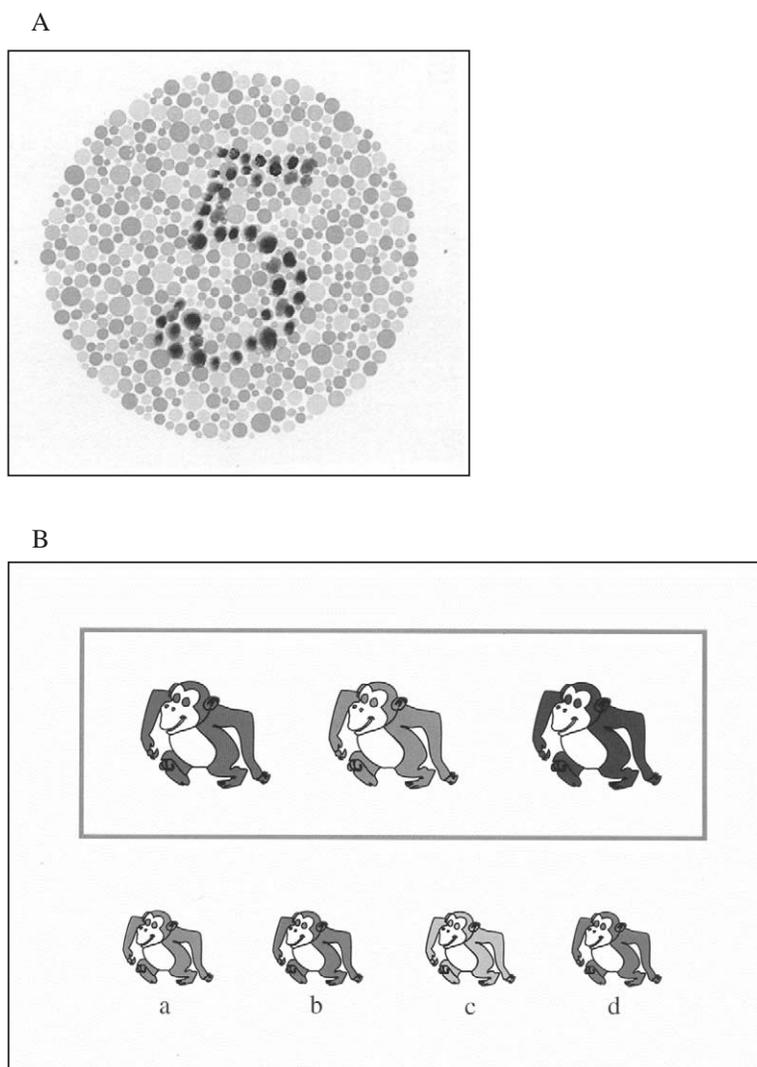
Mr. X works as an editor for a famous Spanish scientific dissemination journal. This activity is not severely handicapped by color blindness, according to various studies. Mr. X was diagnosed as a protanope some years ago. Although his perceptual limitations made it difficult for him to perform tasks in the same way as other people do, he had no problems after carrying out some modifications in his work environment. His past difficulty in correcting page galley proofs and the way

in which he “solved” this problem are an example. This protanope used to receive printed pages on which the errors to be corrected were indicated by red pencil marks. For a normal-sighted person, locating red marks is a very easy task because they “pop up” over the black letter background. However, for Mr. X, dark red and black colors are very similar (nearly metamers), so red marks did not “pop up” from his perspective. Consequently, he missed many error marks. How did he prevent this situation? Very simply: by using a color that would stand out over the background of black letters. In his case, blue ink was a very appropriate solution.

The kind of difficulty described in the previous paragraph has been experienced for centuries. To avoid them, people usually adopt one of two alternatives. Either color-blind people are denied access to some tasks and/or activities, or some task modifications are carried out to make them accessible to color-blind people. In any case, neither of these alternatives is viable in the absence of an accurate chromatic vision diagnosis.

### 4.2. Detecting and Diagnosing Color Blindness: A Two-Step Strategy

Pseudoisochromatic tests (Fig. 4) use stimuli that are very similar in color (“isochromatic”) for color-blind people but not for normal-sighted people (“pseudo”). For example, in the famous Ishihara test (Fig. 4A), color-blind people have difficulty in reading the number presented in each plate. This is because they do not perceive enough differences between the points that form the number and the background. Another similar



**FIGURE 4** Two examples of pseudoisochromatic plates (originals are colored). (A) Ishihara test (observer's task is to identify the number). (B) TIDA test (observer's task is to select the little monkeys similar in color to the big monkeys).

example is provided by a test, called the "TIDA," to detect color-blind children. In this case (Fig. 4B), color-blind individuals of the red-green type select some greenish and reddish monkeys from the bottom line as being similar ("belonging to the same family") to the larger gray monkeys presented in the upper line.

Because they are easily constructed and applied, there are many pseudoisochromatic tests available for using during the first stage of the diagnostic process to detect the presence of color blindness. However, their use must be complemented, during a second stage, with an anomaloscope to get the most accurate

diagnosis. Among the pseudoisochromatic tests capable of detecting a red-green anomaly are the following: the Ishihara test; the Dvorine test; the Hardy, Rand, and Rittler (HRR test); the D-15 test; and the City University Color Vision Test. The latter three tests are also useful for detecting tritan anomalies and for grading the severity of the color perception problem.

Anomaloscopes show users a circle divided into two halves. One half is used to present the reference stimulus and must be metamered in the other half by a stimulus mixture. As indicated previously, to detect and differentiate red-green problems, a Nagel match must be used.

This kind of match uses a reference stimulus of 589 nm (yellow for a normal-sighted person) and an adjustable stimulus mixture made up of 546 nm (green for a normal-sighted person) and 670 nm (red for a normal-sighted person). To detect and differentiate among tritan problems, the Moreland match must be used. This uses a reference stimulus made up of a mixture of 480 nm (greenish blue for a normal-sighted person) and 580 nm (yellow for a normal-sighted person) and an adjustable stimulus mixture made up of 436 nm (blue for a normal-sighted person) and 490 nm (bluish green for a normal-sighted person).

### 4.3. After Diagnosis: Trying to Compensate for Perceptual Limitations

Gaining awareness of people's color blindness and perceptual limitations is the first step in the process of obtaining psychological help. With this knowledge, color-blind people can avoid incorrect attributions about the causes of their perceptual difficulties, and just as important, they obtain a first cue to adopt compensatory interventions. More concretely, and returning to the case discussed earlier, Mr. X's awareness of his perceptual peculiarities prevented him from attributing his difficulty in locating red marks to cognitive or motivational factors. He never thought that his difficulties were due to an attentional deficit. Eventually, he realized that his problems were the consequence of his difficulty in differentiating the dark red ink marks from the black letters, and so he was able to suggest the use of blue ink.

Knowing that a person is color-blind, the psychologist should look for, and try to compensate for, the following two kinds of problems:

- Difficulty in performing figure-background segmentations
- Difficulty in using the basic chromatic categories

#### 4.3.1. Figure-Background Segmentation Difficulties

Many computer users are familiar with the following situation. Initially, they used a common application (e.g., Microsoft PowerPoint) to present a colored written message on a colored background (e.g., yellow letters on a green background). Of course, this message is easy to read on the original screen because the letters used have an adequate lightness contrast with the background.

However, when another screen is used to present the same message (e.g., a large screen used in a classroom), the message becomes very difficult to read due to a reduction in the lightness contrast. The cause of this problem is simply that the same document, as well as the same software specification, produces different colors on different screens.

A problem very similar to the one just described is relatively frequent for some types of color-blind people because, when looking at some stimuli, they experience lightness contrasts that are different from those experienced by normal-sighted people. For example, protanopes perceive reddish backgrounds as darker and, consequently, have difficulty in reading black letters on reddish backgrounds. They also have difficulty in detecting when a red light is on because they perceive it as less brilliant than do normal-sighted people.

Color-blind people not only have figure-background segmentation problems derived from their peculiar lightness perception, but a similar type of problems also may be a consequence of their restricted gamut of chromatic experiences. To understand the origin of this type of difficulty, consult Fig. 3 again and remember that, for some dichromats, the three stimuli used by the anomaloscope affect the responses of only one cone type. Consequently, when presented with the precise physical intensity (see Eq. 1), the exact same color experience is produced in these people, although normal-sighted people would perceive very different hues (green, yellow, and green) in response to these stimuli.

#### 4.3.2. Problems Related to the Use of Basic Chromatic Categories

The expression "basic chromatic categories" is used to name the linguistic categories consistently used among speakers of a certain language. For the languages in most developed countries (e.g., English, Spanish, Japanese), there are 11 categories. These are identified in English by the following terms: red, green, blue, yellow, white, black, brown, orange, purple, pink, and gray. To say that these categories are used consistently means that, for example, if an observer is required to select the prototype of one of these categories, he or she will select the same one, or a very similar one, on different occasions. These choices will also be very similar to the equivalent choices performed by other normal-sighted people.

Previously, the singular adjustments performed by color-blind people in a Nagel anomaloscope were described. Specifically, it was indicated that stimuli

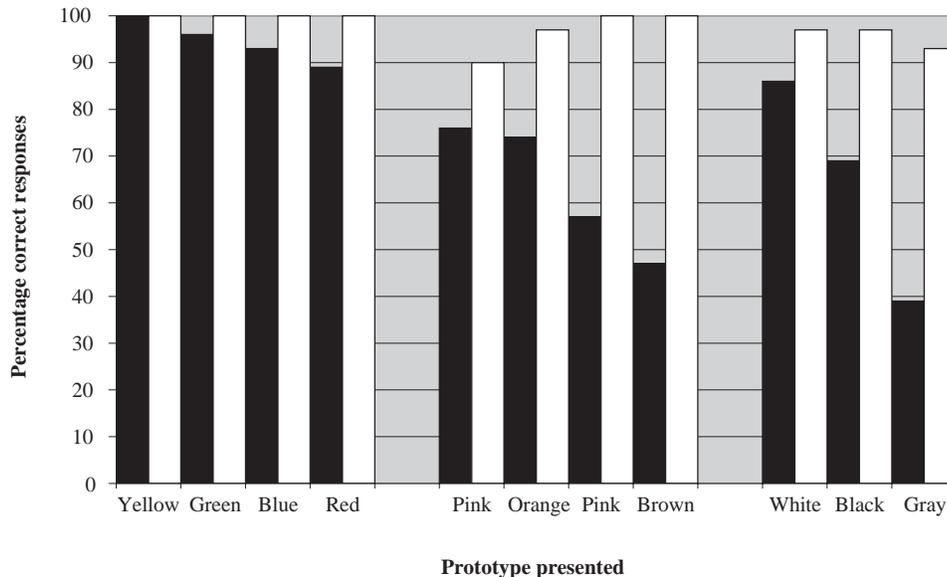
belonging, for normal-sighted people, to the red, green, and yellow categories can be metamers for red–green dichromats. Taking this into account, and also considering that color-blind people generally have an increased number of metamers, it is logical to expect color-blind people to have severe difficulties when using basic chromatic categories. In contrast to this expectation, research carried out with 30 dichromat children (protanopes and deuteranopes) provided the results presented in Fig. 5. The children achieved an important level of “hits” when required to name prototypical exemplars of the 11 basic Spanish chromatic categories. Moreover, for five basic categories (yellow, green, blue, red, and white), their hit percentage was more than 85%.

Why is there such disparity between, on the one hand, the low discriminative level shown by color-blind people in response to the stimuli used by diagnostic tests and, on the other, their relatively high capacity to name prototypes? What is the practical relevance of this fact? These questions are addressed in turn.

The main difference between the stimuli used in an anomaloscope and the stimuli provided by everyday chromatic surfaces is the range of available intensities. For example, when using a Nagel anomaloscope, the wavelengths that a normal-sighted person perceives as green (546 nm), yellow (589), and red (670 nm) can be presented with a broad range of physical intensities. Consequently, differences in the capacity to activate

cone responses can be compensated by intensity variations (see Eq. 1). This compensation is not possible with surface colors in which the response intensity (perceived lightness) is very closely related to the specific wavelengths reflected. So, for red–green dichromats, the yellow prototype is always perceived with a higher lightness than is the green or red prototype. Of course, this also helps these individuals to avoid naming confusions. This does not occur when no prototypical exemplars of the basic categories are used (e.g., pale green), in response to which more naming errors are produced.

Because many everyday chromatic stimuli are close to the basic chromatic category prototypes, color-blind people can frequently name a color correctly (Fig. 5). Consequently, for example, many teachers frequently do not agree with the results provided by the standardized chromatic vision evaluation procedure because these results are always worse than children’s real capacity to name and differentiate everyday chromatic materials. Taking this into consideration, should we conclude that chromatic evaluation is useless? The answer is clearly no because color blindness detection alerts one to the problems with figure–background segregation discussed previously and, logically, because these individuals continue to have more trouble in naming and using color categories than do normal-sighted people, even when prototype exemplars are used (Fig. 5).



**FIGURE 5** Percentage of correct naming responses to basic Spanish category prototypes. Black bars, dichromats’ responses; white bars, controls’ responses.

**See Also the Following Articles**

Coping ■ Psychoneuroimmunology

**Further Reading**

- Birch, J. (1993). *Diagnosis of defective color vision*. Oxford, UK: Oxford Medical Publications.
- Fletcher, R., & Voke, J. (1985). *Defective color vision: Fundamentals, diagnosis, and management*. Boston: Adam Hilger.
- Hsia, Y., & Graham, C. H. (1997). Color blindness. In A. Byrne, & D. R. Hilbert (Eds.), *Readings on color: The science of color*. Cambridge, MA: MIT Press. (Original work published 1965).
- Kaiser, P. K., & Boynton, R. M. (1996). *Human color vision* (2nd ed.). Washington, DC: Optical Society of America.
- Mollon, J. D., Pokorny, J., & Knoblauch, K. (2003). *Normal and defective colour vision*. Oxford, UK: Oxford University Press.
- Newton, I. (1952). *Optics* (4th ed.). New York: Dover. (Originally published 1730).



# Community Psychology

Kenneth I. Maton

University of Maryland, Baltimore County, Baltimore, Maryland, USA

1. Introduction
2. Guiding Values
3. Ecological Multilevel Framework
4. Community Research Methods
5. Community Research Areas
6. Prevention and Promotion
7. Social Action and Community Change  
Further Reading

## GLOSSARY

*ecological* In psychology, an emphasis on the importance of the environmental context in understanding individual behavior and social problems.

*prevention* Interventions that aim to reduce the occurrence of physical or psychological disorders and social problems.

*promotion* Interventions that aim to enhance psychological well-being and the quality of community life.

*social action* Interventions that strategically exert influence at the grassroots level to enhance access to resources for disadvantaged groups.

Community psychology is concerned with the relationship between individuals and their social and community environments. Community psychologists contribute to the public welfare through collaborative interventions that prevent psychological and social problems, promote personal and community well-being, and empower disadvantaged citizens and groups.

## 1. INTRODUCTION

The field of community psychology developed during the mid-1960s as a reaction against the limitations of traditional intrapsychological approaches to research and social problems. In contrast, the developers of the field emphasized the importance of social contexts and social environments in understanding behavior and social problems, envisioned a strengths-based approach rather than a deficits-based approach to research and action, and emphasized the importance of the prevention of problems rather than the treatment of problems after the fact. Community psychology shares with sociology an emphasis on the influence of structural features of the environment on human behavior, but the former differs from the latter research discipline in its greater emphasis on action and individual experience. Community psychology shares with social work a focus on helping individuals in a community context, but the former differs from the latter applied field of practice in its greater emphasis on empirical research, systems-level change, and prevention.

Community psychologists are involved in research and action focused on the entire spectrum of contemporary issues facing citizens and communities. These include educational reform, homelessness, domestic violence, intergroup conflict, substance abuse, teenage pregnancy, poverty, child abuse, youth violence, racial and sexual discrimination, coping with life stress, community capacity building, and enhancing psychological sense of community. Applied research on these issues draws on a wide

range of research methods, including needs assessment, qualitative interviews, quantitative field research, focus groups, evaluation research, and action research. Similarly, the range of intervention approaches through which community psychologists address social problems is broad and includes program development, consultation, community coalition building, social action, community development, and social policy advocacy.

Distinctive aspects of community psychology include its guiding values, its ecological multilevel framework, and its focus on prevention, promotion of well-being, group empowerment, and community change. Each of these is described in this article.

## 2. GUIDING VALUES

Community psychologists emphasize the distinctive nature of the values that guide their work. The following seven guiding values can be discerned.

### **2.1. Individual Wellness: The Physical, Psychological, Social, and Spiritual Health of Citizens**

Unlike clinical psychologists, community psychologists do not focus on the etiology and remediation of mental illness; rather, they focus on understanding the factors that contribute to well-being and healthy development and on the associated development of programs that prevent the development of physical and psychological problems.

### **2.2. Sense of Community: Citizens' Sense of Belonging to a Larger Group**

Community psychologists view the decline of community cohesion as a major contributor to individual and social problems and the enhancement of a sense of community as a key contributor to individual wellness and community vitality.

### **2.3. Social Justice and Empowerment: The Equitable Distribution of Economic, Political, and Psychological Resources in Society**

Community psychologists are concerned with addressing the pressing social issues of the day and in particular with empowering marginalized groups through enhancing

their access to economic, political, and psychological resources.

### **2.4. Citizen Participation: The Active Involvement of Citizens in All Aspects of Community Life**

Community psychologists are wary of expert-driven, top-down solutions for the problems that communities face, viewing such approaches as too often ineffective and as contributors to citizen alienation and dependence. In contrast, community psychologists view citizen involvement as the antidote necessary to bring about community revitalization and solutions to pressing social problems.

### **2.5. Collaboration and Community Strengths: Working Together with Citizens and Groups in the Community and Building on Their Strengths**

As proponents of a strengths-based perspective, community psychologists view citizens and community groups as the sources of solutions to community problems. Working together as partners allows the respective strengths of both community members and community psychologists to contribute to community betterment.

### **2.6. Respect for Human Diversity: Genuine Appreciation of Human Differences Such as Differences in Ethnicity, Gender, Sexual Orientation, Religion, Age, Physical Disability, and Social Class**

Community psychologists are critical of social paradigms that view those who are "different" from the mainstream as the cause of the problems they face in society. In direct contrast, community psychologists affirm diversity as a public good and an invaluable resource.

### **2.7. Empirical Grounding: The Research Basis for Efforts to Make a Positive Difference in Communities and the Larger Society**

Community psychologists view systematic evidence as the key to understanding individual and social problems and view such understanding as central to

effective program development, community action, and social policy.

Taken as a whole, these seven values reflect the field's joint commitment to research and action, its focus both on bettering the lives of individuals and on bringing about broader social change, and its emphasis on working with, for, and in the community. Individual community psychologists, however, vary greatly on the relative importance placed on each of the guiding values, with resultant differences in preferred approaches to community work. For example, academic researchers emphasize the importance of research and publication, whereas applied community practitioners emphasize the value of action and hands-on community involvement. Community psychologists oriented to individual wellness emphasize the development of local programs to prevent problems and assist those in need, whereas those oriented to social justice and empowerment often articulate the need for higher level social change efforts to transform social norms and to redistribute social and economic resources. Strongly quantitative community psychology researchers emphasize traditional research values of rigorous design and measurement validity, whereas those more oriented to social justice issues emphasize discovery-oriented qualitative, participatory, and action research methods.

### 3. ECOLOGICAL MULTILEVEL FRAMEWORK

The guiding conceptual framework of community psychology is social ecology, which emphasizes that an understanding of human behavior must always encompass the social environment in which it is embedded. Within the social environment, multiple levels of analysis can be distinguished, including the individual, the microsystem (e.g., family, classroom, social network), the organization, the community, and the macrosystem (e.g., cultural norms; the larger society). Research at the individual level of analysis might focus, for example, on the relationship between psychological well-being and level of perceived social support in a sample of inner-city youth. At the microsystem level of analysis, a researcher might study the relationship between the social climate of self-help groups and group effectiveness. Evaluating the impact of a comprehensive, community-wide teenage pregnancy prevention program on incidence rates of teenage pregnancy in communities that either received or did not receive the

intervention is an example of research at the community level of analysis. Additional examples of research topics at various levels of analysis are depicted in Table I.

Central to the ecological paradigm is the principle of ecological interrelatedness. Lower levels of analysis are embedded within higher levels and are directly influenced by, and can also influence, these higher levels. Thus, the role of the environment in causing individual and social problems is emphasized. Indeed, community psychology arose in part as a reaction against views that "blamed the victims," such as lower income populations and ethnic minorities, for problems that community psychologists view as stemming from larger community and societal forces.

Indeed, given the complex social problems with which the field is concerned, community psychologists believe that multiple levels of analysis should ideally be encompassed in community research and action. Take, for example, the case of youth violence in urban areas. A community psychology approach to addressing that problem, based on the ecological perspective, would naturally call for interventions at multiple levels. These might include interpersonal problem-solving and

TABLE I  
Ecological Levels of Analysis in Community Research

Level	Sample research topics
Individual	Psychosocial predictors of well-being, competency-enhancing prevention programs, psychological empowerment, resilience, individual indicators of community problems
Microsystem	Family cohesion, social networks, peer groups, classroom social climate, teacher expectations, self-help group processes, neighborhood effects
Organization	Organizational characteristics of empowering community settings, whole school reform, faith-based organizations, voluntary associations, block associations
Community	Community coalitions for drug abuse prevention, community-wide health promotion, psychological sense of community, community readiness for change
Macrosystem	Social-political forces, cultural beliefs, social movements, strengths-based social policy

Adapted from Dalton, Elias, & Wandersman (Table 1.1, p. 13).

conflict resolution skills training at the individual level, family support programs at the microsystem level, whole school reform of urban schools at the organizational level, a campaign to alter cultural norms about violence and masculinity at the community level, and a strengths-based antipoverty initiative at the federal level. Although a given research or intervention project of any individual community psychologist will not realistically be able to encompass all ecological levels in any social problem area, the field as a whole aspires to conduct research and interventions that encompass the entire spectrum of ecological levels.

#### 4. COMMUNITY RESEARCH METHODS

Community psychology researchers employ a wide range of research methods to examine community phenomena of interest. These encompass both qualitative and quantitative methods as well as research designs ranging from experimental program evaluation to action research.

Qualitative methods are especially useful for generating in-depth descriptions of a phenomenon, an understanding of context, new insights, and opportunities for the “voice” of marginalized populations to be heard. Methods include qualitative interviewing, participant observation, and focus groups. Interviews can be coded systematically using sophisticated qualitative software programs. Participant observation allows in-depth examination of a particular group or community setting, with a focus on uncovering the worldview and meaning system of members. Focus groups facilitate the generation of common themes that emerge as a result of discussion in the group context.

Quantitative methods are used for varied purposes in community psychology, including hypothesis testing, needs assessment, and program evaluation. Hypothesis testing research may focus on community samples of individuals, on samples of groups or organizations, or (increasingly in multilevel research) on individuals, families, and groups nested in a sample of organizations, neighborhoods, or communities. Needs assessment provides systematic descriptive information concerning the nature and level of local needs in a setting or community and serves as a basis for collaborative planning of interventions. Program evaluation examines the extent to which interventions are achieving their intended outcomes and the factors

contributing to observed outcomes, using experimental or quasi-experimental research designs.

Increasingly, community researchers combine qualitative and quantitative methods within a single study. This approach allows for building on the strengths of each research method. Furthermore, to address community phenomena, there is increasing focus on interdisciplinary research that incorporates research methods from other fields. Examples include community ethnography (anthropology), geographic information systems (geography), community case study (community sociology), and epidemiology (public health).

Consistent with the community psychology values of citizen participation and collaboration as well as building on community strengths, partnerships are often developed between community researchers and populations under study. Investigators engage those being studied as active collaborators in defining the research question, helping to develop appropriate research instruments, designing and implementing research procedures, and interpreting and disseminating research findings. In some cases, helping local groups or organizations to develop their own research capacities may be one of the intended outcomes of the research process, as is the case in empowerment evaluation.

Community psychology espouses adventuresome research methods appropriate to the research question at hand, even if these methods are not conventional in nature. In many research contexts, the researcher is viewed not as an objective, fully detached observer but rather as an individual whose values influence the scientific enterprise in important ways. Furthermore, consistent with this perspective, the researcher’s relationship with those studied is seen as part of the larger social ecology of the phenomenon under investigation.

#### 5. COMMUNITY RESEARCH AREAS

Community psychologists examine a wide range of topics in community-based research. Four illustrative research areas are briefly described in what follows.

##### 5.1. Social Problems

One broad area of community research centers on pressing social problems such as discrimination, poverty, education of minority students, substance use, HIV/AIDS, and homelessness. Much of this research focuses on marginalized groups in society such as ethnic minorities and lower income populations. Community

psychologists examine factors that contribute to the specific social problem under focus. Individual, family, and environmental variables are encompassed in this research. For example, in the urban education area, community psychology researchers have examined the roles of family support, teacher expectations, the transition to middle school, the social climate and culture of schools, and neighborhood poverty. The results of research in each social problem area provide a basis for intervention approaches to address the particular social problem.

### 5.2. Stress and Mental Health

A second long-standing focus of community psychology research examines the interrelationships among stress, coping, social support, and psychological well-being. Life stressors examined run the gamut from acute stressors (e.g., death of a parent) to ongoing stressors (e.g., poverty). Research in this area has revealed various personal and environmental factors that help individuals to cope with stress, including spirituality, interpersonal problem-solving skills, and social support from friends, family, and mentors. This body of research has suggested important foci for psychological and environmental interventions that aim to prevent negative outcomes of stress and enhance resilience in the face of adversity.

### 5.3. Community and Quality of Life

A third area of research examines the role of community settings in contributing to the quality of individual and community life. This research encompasses settings such as voluntary associations, block associations, faith-based organizations, and self-help groups. These settings provide a meaningful niche in which citizens find meaning, support, and opportunities to contribute to the well-being of others and the larger community. Research on community settings indicates that certain organizational characteristics, such as cohesion, support systems, opportunity role structure, and leadership, have been found to be related to member well-being and psychological empowerment.

### 5.4. Sense of Community

A fourth illustrative area of community psychology research centers on psychological sense of community. This refers to the sense of belonging or connectedness to a larger group. Sense of community is important both in relational communities defined by shared

goals and activities (e.g., church, professional association) and in geographical communities (e.g., neighborhood, city, nation). Researchers have delineated four underlying dimensions to sense of community: membership, influence, integration and fulfillment of needs, and shared emotional connection. The individual and environmental correlates of sense of community, subgroup differences, and changes over time represent important topics of research in this area.

The importance of context and the importance of diversity are cross-cutting themes in many areas of community psychology research. Depending on the dangerousness of neighborhood context, for example, parenting style or psychological sense of community has been found to be more or less adaptive. Ethnic groups differ in terms of life contexts and challenges, culturally embedded values and meaning systems, and access to resources and services. Findings related to distinctiveness of context and diversity provide a critical foundation for the development of context-sensitive and culturally competent interventions.

In addition to theory generation and hypothesis testing research, many community psychologists conduct evaluation research. This research sometimes focuses on interventions developed by the investigator but more frequently examines various community-based programs developed by others. Evaluation research examines program outcomes, the quality of implementation of programs, and the theory-based processes through which program efforts occur.

Importantly, research in community psychology often builds on theory and findings from other subfields of psychology. Developmental psychology provides critical underpinnings for community psychology research on children's coping, support, and well-being. Social psychological theory contributes to community psychology work in the areas of discrimination, oppression, and empowerment. Health psychology models help to guide community psychology work in the areas of health promotion and disease prevention. Cultural psychology provides important theoretical models and perspectives for community psychology work on ethnic identity, coping, and ethnic minority populations.

In comparable fashion, disciplines outside of psychology influence community psychology theory and research. Public health has a major influence in the area of prevention and health promotion. Anthropology's focus on cultural norms, practices, and worldviews has broadened community psychology models to encompass the culture of settings and cultural narratives of community members. More generally, across the gamut of social

problem domains, there are major theoretical paradigms and accumulated research findings from other fields of direct relevance (e.g., sociology of education, community sociology, urban affairs, criminology, economics, law, social history).

Conversely, many community psychology models and perspectives have influenced other subareas of psychology and allied researchers in other disciplines. For example, prevention has become an important domain of investigation in other applied areas of psychology, including developmental psychology, applied social psychology, health psychology, and clinical psychology. Similarly, the importance of community-based research, ecological context, the strengths-based paradigm, qualitative research, and varied community psychology findings in specific social problem areas has influenced investigators in multiple areas of psychology as well as those in allied social science disciplines.

## 6. PREVENTION AND PROMOTION

In contrast to helping individuals after serious problems have developed, community psychologists seek to prevent the initial occurrence of problems. Primary preventive interventions aim to reduce the number of new cases (i.e., incidence) of a wide range of physical illnesses and psychological problems, including HIV/AIDS, substance abuse, and depression. Universal preventive interventions seek to do so while focusing broadly on a general population (e.g., all students in a school or school system). Selective preventive interventions focus strategically on the subset of individuals at above average risk for developing a disorder or problem due to either environmental factors (e.g., alcoholic parent, concentrated poverty) or personal factors (e.g., school difficulty). Indicated prevention programs focus specifically on individuals at high risk for developing a disorder or problem, usually those showing early symptoms of the disorder or problem.

Prevention programs have been developed for every stage of the life span. However, most prevention programs focus on infants, children, and youth, based on the principle that early intervention is the most likely means to prevent later programs. For example, Nurse Home Visitor programs provide low-income, single new mothers with support, child-rearing information and skills, and linkages to community-based resources as a means to prevent child abuse and poor child outcomes. Head Start and related preschool programs help low-income preschoolers to learn basic school-

related social and cognitive skills commonly present in their middle-class counterparts, thereby preparing those served to make a positive transition to formal schooling.

Prevention programs in public schools provide at-risk children with important coping skills and social support. Programs for middle and high school youth often focus on prevention of specific social problems such as substance abuse, school dropout, and teenage pregnancy, for example, by providing refusal skills to counteract negative peer pressure, information that underscores the implications of poor choices, and meaningful positive activities (e.g., community service) in which to engage.

Prevention specialists increasingly emphasize the importance of comprehensive, community-wide, and multilevel interventions. In the case of a program to prevent school dropout among urban youth, for example, a comprehensive preventive intervention might address multiple risk factors related to dropout, such as students' academic skills and knowledge, coping skills, engagement in creative or prosocial activities, and adult social support, rather than only one of these (and other) contributing factors. A multilevel approach in the academic performance domain, for example, might include an individual tutoring component, enhancement of the cultural competence of teachers, and schoolwide changes to provide a more supportive and caring school climate rather than an intervention at only one of these levels. Finally, a community-wide intervention could mobilize students, teachers, parents, nonprofit organizations, businesses, and the media. Given the difficulty of making a difference in pressing social problems, comprehensive, multilevel, and community-wide programs are more likely to provide the intensity and multiplicity of impact necessary to counter the multiple factors contributing to problem development.

Promotion of wellness interventions seeks to enhance psychosocial competencies and development, as well as the overall quality of individual and community life, rather than aim specifically (as preventive interventions do) at reducing rates of a specific disorder or problem. Intervention programs that contribute to normative child development, enhance children's resilience to adverse circumstances, and/or strengthen parenting skills are examples of promotion interventions at the individual level of analysis. Efforts to enhance psychological sense of community, intergroup relationships, and/or neighborhood revitalization are examples of promotion interventions at the microsystem or community level.

Outcome research has demonstrated the effectiveness of a range of prevention and promotion programs. However, these programs often are developed by university-based investigators under the ideal conditions of large amounts of grant funding and a singular focus on program success. A major challenge in the prevention area has been to effectively sustain programs after grant funding ends and to effectively disseminate programs to new sites and communities lacking the resources and special conditions that were present for initial program development. Given these challenges, the study of factors leading to effective program implementation, dissemination, and adoption is receiving additional focus. Some of the key variables found to be important include strong agreement on program goals among the various stakeholders, linkage of the program goals to the basic mission of the setting, clear and strong program leadership, and adequate implementation of core program components and principles.

## 7. SOCIAL ACTION AND COMMUNITY CHANGE

An additional important domain of community psychology interventions relates to social justice and empowerment. Social action and community change aim to enhance levels of economic, social, and psychological resources for individuals and groups lacking power in society. A range of methods are employed.

Grassroots organizing efforts bring together citizens and community groups in low-income neighborhoods to pressure elected officials to allocate additional funding for basic individual and community needs. Organizing efforts may involve recruiting individual citizens or working with extant community groups, such as inner-city churches, that have preexisting access to large numbers of citizens. Community psychologists involved in grassroots organizing efforts bring to bear their group skills and empowerment perspective along with linkages to resources and research bases that can be used by those leading the organizing efforts.

Community coalition development brings together groups from multiple sectors of the community to devise plans for addressing a specific community need. Groups represented often include local government, business, human service agencies, citizen associations, and faith-based organizations. Community

psychologists can serve as coalition conveners, resource experts in specific substantive areas, technical support providers, and/or evaluators.

Action research addresses local community problems through collaboration with community groups centered on local community needs. Projects may be initiated by community psychologists interested in a particular community problem or by community groups that approach community psychologists for help in addressing a need specific to their groups or the community. Consistent with an empowerment perspective, the research and action pursued have direct utility for self-defined local community needs.

Finally, community psychologists have become increasingly involved in social policy advocacy. Some community psychologists bring policy-relevant research to the attention of policymakers through congressional hearings, dissemination of reports, and/or congressional briefings. Others work full-time in the policy sector as congressional aides or policy analysts in the executive branch of government. Finally, a small number of community psychologists have become policymakers in elected positions, including local school board members and city council members.

In their social action and community change work, community psychologists necessarily work with a wide range of citizens, community groups, organizational staff and directors, and elected officials. The values of the field, including the focus on empowerment, collaboration, and community strengths, serve as important guides in this arena.

## See Also the Following Articles

Social Networks ■ Stress

## Further Reading

- Dalton, J. H., Elias, M. J., & Wandersman, A. (2001). *Community psychology: Linking individuals and communities*. Belmont, CA: Wadsworth/Thompson.
- Levine, M., & Perkins, D. V. (1997). *Principles of community psychology: Perspectives and applications* (2nd ed.). New York: Oxford University Press.
- Rappaport, J., & Seidman, E. (Eds.). (2000). *Handbook of community psychology*. New York: Plenum.
- Revenson, T. A., D'Augelli, A. R., Hughes, D., Livert, D., French, S. E., Seidman, E., Shinn, M., & Yoshikawa, H. (Eds.). (2001). *A quarter-century of community psychology*. New York: Plenum.

- Rudkin, J. K. (2003). *Community psychology: Guiding principles and orienting concepts*. Upper Saddle River, NJ: Prentice Hall.
- Sarason, S. B. (1974). *The psychological sense of community: Prospects for a community psychology*. San Francisco: Jossey-Bass.
- Shinn, M., & Toohey, S. M. (2003). Community contexts of human welfare. *Annual Review of Psychology*, *54*, 427-459.
- Weissberg, R. P., & Kumpfer, K. L. (2003). Prevention that works for children and youth [special issue]. *American Psychologist*, *58*(6/7), 425-490.



# Compensation

Henk Thierry

Tilburg University, Tilburg, The Netherlands

1. Psychological Perspectives
  2. The Design of Compensation
  3. Job Evaluation
  4. Performance-Related Pay
  5. Conclusion
- Further Reading

## GLOSSARY

**cafeteria plan** A system in which organization members get the opportunity to trade some labor conditions for others to design a package with the highest personal utility; work time may be traded for leisure time, cash may be traded for additional health insurance, and so forth.

**fringe benefits** A term that refers to secondary and tertiary labor conditions; primary conditions cover direct compensation and provisions on work time, holidays, and vacations, secondary conditions concern health insurance and other types of social security, and tertiary conditions are specific to a company (e.g., car lease rates).

**incidental pay** A system that provides a supervisor with the opportunity to rapidly express recognition for a job extremely well done, for effort beyond the call of duty, and the like; rewards are often small (e.g., a free lunch, a dinner for two, a paid conference).

**job evaluation** A systematic procedure for describing the standard content of a job as well its standard conditions; depending on the system's particular characteristics, job descriptions are analyzed and evaluated.

**merit rating** A system of performance appraisal based on qualitative characteristics of individual workers (e.g., creativity, diligence, care for clients); the appraisal outcome

is rewarded (usually annually) with an extra step on the job's salary scale or with a separate bonus.

**performance-related pay** A category of pay systems in which work outcomes are tied to pay; outcomes may bear on individual or team contributions to work, the achieved results, and/or departmental or organizational effectiveness indexes (e.g., sales volume, profits, customer satisfaction).

**reflection theory on compensation** A theory in which compensation is considered to be a major domain relevant to a person's self-identity; compensation reflects information in four areas and resulting in four different meanings: motivational properties, relative position, control, and spending.

Compensation is a major domain within human resource management to which several disciplines contribute. Psychological approaches to financial compensation focus on the motivational force of pay, its potential to get organization members to learn new abilities and skills, and the conditions for perceiving pay systems and practices as fair. In industrialized countries, the structure of compensation in profit and nonprofit organizations is quite comparable. Two important components of compensation—job evaluation (which often determines base pay) and performance-related pay—are considered more closely. Research indicates, for instance, that job evaluation tends to grasp one common factor (extent of required education), whereas employees may perceive the evaluation outcome as a harness. Performance-related pay appears to be a rather vulnerable kind of system, particularly when performance appraisal is based on qualitative characteristics. Frequently, employees do not

perceive a relationship between the appraisal result and the awarded bonus amount. When quantified critical performance factors are applied, this relationship becomes more clear. Also, rather small differences in pay affect the behavior at work more often than do large differences in pay.

## 1. PSYCHOLOGICAL PERSPECTIVES

Managers and employees in an organization often wonder whether its compensation systems and strategies affect the work behaviors of the organization's members, individually as well as in teams, and the performance results of the organization at large. Psychologically, this theme can be addressed in more than one way. One approach is in terms of the conditions that make compensation more or less motivating, often relative to the motivating potential of other work characteristics. Characteristically, the motivational force of compensation is inferred from the organization members' level of effort, achieved performance, degree of satisfaction, and opportunities for personal development as well as from the effectiveness of cooperative efforts. A slightly different approach concerns the extent to which compensation facilitates the learning of new abilities and skills rather than stimulating employees to maintain their former habitual working patterns. This second approach is also basic to the decision, in cases of complex processes of organization change, as to whether alterations in compensation should precede or follow these. A third approach is distinguishable as managers face, for instance, the need to cut costs and are considering whether another compensation package might help them to achieve this. In such cases, an important perspective is the fairness with which this may be brought about, both in the procedures followed and in the quality of the end result. Here again, it is vital to consider the conditions under which compensation affects performance results.

These (and other) psychological perspectives are becoming more meaningful as they are considered within the frameworks of particular compensation systems. Although there are huge differences among pay forms and systems, there is a comparable compensation design across industrialized countries. This design is outlined in the next section. Then, two main categories of systems—job evaluation and pay for performance—are discussed from the viewpoint of how psychological insights may help to better understand recurring problems.

## 2. THE DESIGN OF COMPENSATION

Although there are great differences among industrialized countries in the amount of pay provided for particular jobs, the structure of compensation is quite comparable throughout the major part of the world. It is designed on the basis of most or all of the components described in what follows.

### 2.1. Base Pay

The base pay component reflects the value of a job to an organization often with respect to the particular labor market of a country or region. In many cases, systems of job evaluation are used to appraise and order the value of most or all available jobs within an organization. These relative job values are subsequently categorized in classes to which particular wage and salary scales are tied. In other cases, managers monitor the salaries that are paid in referent organizations for particular jobs and apply these to achieve "market conformity." Job values embody, to a large extent, how scarce particular abilities and skills—or competencies—are within a (national) labor market. Thus, base pay recognizes the (scarcity of) qualifications of a worker to fulfill a specific set of tasks.

### 2.2. Performance-Related Pay

The performance-related pay component bears on the appraisal of job performance. Characteristically, specific targets or standards are set, and these have to be met or passed before a particular bonus is awarded. When the quantity of work performance was dominant, targets applied to the number of "pieces" made (piece rate system) or the amount of time required to perform a particular task (tariff system). Current standards relate more often to performance quality, team effectiveness, sales value, client satisfaction, and the like. Through performance-related pay, employees are recognized for what they do (and don't do) in their work and are often thought to be stimulated to improve their performance.

### 2.3. Fringe Benefits

The fringe benefits component reflects the labor conditions that apply to members of an organization. In Europe, fringe benefits usually relate to a collective labor agreement (within the framework of a country's social security policy) and include both secondary (e.g., health insurance programs) and tertiary conditions (e.g., a

company's car lease program). In the United States, fringe benefits packages are much more organization specific, mirroring entrepreneurial and managerial views on appropriate provisions. Usually, their coverage is much more modest than in European countries. When a cafeteria plan is applied, organization members get the opportunity to select, on a regular basis (often annually), alternative labor conditions, trading less attractive options for more attractive ones. Fringe benefits mirror the views of entrepreneurs and union leaders (and society as well) on the required level of minimal security in work and life.

#### 2.4. Incidental Pay

When deciding on matters of pay, managers in most countries must take into account various laws, regulations, and the specific organization's tradition of established practices. This host of rules might make it very hard to quickly express recognition to employees who have performed beyond expectations, put in many more hours than agreed on to speed up deliveries, and so forth. Incidental pay provides managers with the opportunity to reward some employees directly when applicable. Usually, the awards are small (e.g., a lunch for two, a dinner party). However, incidental pay might also involve larger awards (e.g., a paid conference).

Countries differ to a large extent regarding the proportional amount of base pay within their organizations. In some countries, base pay may amount to more than 90% of an employee's salary, whereas in other countries, base pay may cover approximately 50% of the salary. Moreover, opinions vary concerning the degree to which organization members' incomes should be based on their performance and sensitive to the volatility of (consumer) markets. Because job evaluation (base pay) and pay for performance have an incisive effect on managers' and employees' behaviors at work, these two components are addressed in more detail in what follows.

### 3. JOB EVALUATION

There are many different systems of job evaluation in use. Common to most systems is that key jobs are described according to their content and accompanying standard work conditions and are subsequently analyzed and appraised in terms of a job evaluation system's characteristics (e.g., required knowledge, problem solving, supervision). Frequently, a point rating procedure, in which numerical values are assigned to the main job features, is

applied. The sum total of these values constitutes the "job value." This approach is, to some extent, derived from the technique of job analysis, a well-known subject area within work psychology. Through job analysis, the work elements of a job (e.g., cutting; sewing) are carefully analyzed in terms of those worker characteristics that are required for successful job performance (e.g., abilities, skills, personality factors). The objective of any job analysis system is to grasp the essence of a job in terms of a variety of distinct "worker" and/or "work" qualifications.

Would the latter objective also hold for job evaluation? Research shows that these systems usually measure one common factor: extent of required education. In practice, several (professional) groups in organizations tend to stress the relevance of particular job qualities (e.g., leadership behavior, manual dexterity and sensitivity for specific tools in production work) that are believed to have been underrated in a particular job evaluation system. But factor-analytic studies show that even when a job evaluation system is "extended" to include such particular job qualities, second and/or third factors—if they can be identified at all—have much in common with the first (general education) factor. In other words: job evaluation measures the educational background required for adequate decision making in a job. Although the concepts and wording of job characteristics in various job evaluation systems might seem to differ from one another, they tap into the same construct. This is also exemplified through the custom, in several countries, of using a simple arithmetic formula to translate job value scores from one system of job evaluation into the scores of another system.

Obviously, these procedures for determining job value and base pay level do have important implications. First, the process of translating work descriptions or characteristics into worker attributes is quite vulnerable because it requires psychological expert knowledge to determine the variety of attributes (e.g., abilities, aptitudes, skills) relevant to adequately perform a particular job. Research has shown repeatedly that ratings may be biased in various respects due to the "halo effect," gender discrimination, implicit personality theory, and the like. This theory refers to a pattern among attributes that a particular rater believes exists. One example might be that somebody scoring high on emotional distance is thought to excel in analytical thinking, have a "helicopter view," and engage in a structuring leadership style because the rater assumes such a pattern to be "logical."

Second, regardless of an individual worker's personal conception of his or her job's main content, that job's

value is determined mainly by the level and nature of the required educational background. Consequently, a worker may perceive the evaluation of his or her job as not representing the worker's ideas about the job's actual content but rather as being the outcome of a bureaucratic exercise. Related to this is a third issue: Job evaluation may be experienced as a "harness" (because it measures one factor nearly exclusively), especially when a fine-grained salary structure is tied to the job value structure. This characteristic often makes it difficult and time-consuming to adjust job value and base salary to changed conditions. Flexible procedures have been designed to accommodate these needs.

Fourth, the larger the base pay proportion, the more organization members may perceive their pay as meaningful to satisfying important personal needs, according to the reflection theory on compensation. This theory holds that pay acquires meaning as it conveys information relevant to the self-concept of a person. Four meanings are distinguished: motivational properties, relative position, control, and spending. However, meaningfulness often diminishes as employees reach the high end of their salary scale without having much expectation of improving their level of pay. This implies that although base pay may motivate employees initially to stay in their jobs and to maintain at least an acceptable level of performance, this motivational force may wear out the longer employees remain in their jobs without noticeable changes in their level of pay.

#### 4. PERFORMANCE-RELATED PAY

Research data and practical experience reveal that the introduction of a performance-related pay system is a time-consuming and rather complicated exercise. Evidently, performance-related pay may be successful in an organization, but there are also many instances where it fails to produce its intended effects. Many factors have to be considered, the most salient ones of which concern yardsticks of performance appraisal, antecedent conditions, size of pay differences, and some contingency variables (e.g., secrecy).

##### 4.1. Performance Appraisal: Qualitative Versus Quantitative Yardsticks

In many companies, qualitative characteristics are used for the appraisal of individual performance such as

creativity, innovation, and customer orientation. These are often alternatively called behavioral yardsticks, personality statements, or competencies. A major problem is their abstract and often heterogeneous nature, without a direct relationship to observable work behaviors. They require a supervisor to "translate" daily and less frequent work events and actions into the framework of abstract person concepts without paying due attention to the interaction between the work situation and individual attributes. Thus, the use of these qualitative yardsticks is very much subject to bias. Because performance appraisal often occurs rather infrequently—(e.g., semiannually, annually), this method of appraising individual performance is vulnerable in practice. Because this performance appraisal result is tied to pay (either as a step on the salary scale or as a bonus), each department's supervisor is usually assigned a particular budget to control salary costs. The supervisor may award each employee with an equal share in the budget. However, performance-related pay is intended to reward differences in performance differentially. Characteristically, a normal distribution curve is applied to effectuate this, as Fig. 1 shows.

Figure 1 reveals that the great majority of employees earn an average bonus, whereas a few receive either a low bonus or a high bonus. Now assume that a supervisor wants to assign a relatively high bonus to some employees whose performance was beyond expectations. The restrictions posed by the budget necessitate that the supervisor assign a relatively low bonus to several other employees even though their performance often would not require this kind of intervention. How

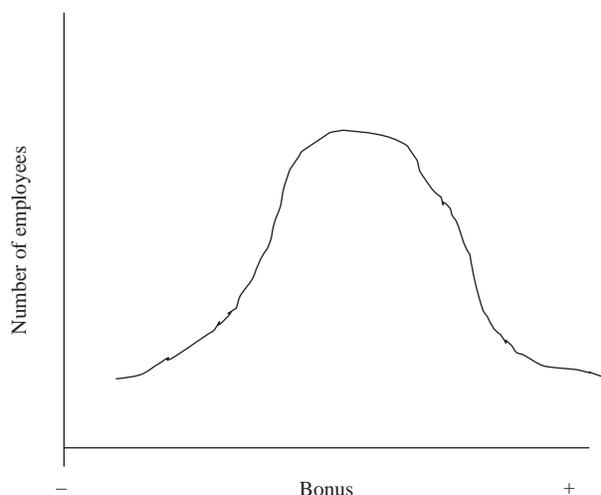
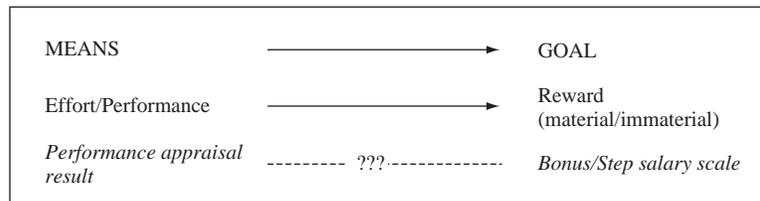


FIGURE 1 Normal distribution curve.



**FIGURE 2** Instrumentality perceptions according to expectancy theory and an absent “line of sight” as merit rating is applied.

should the supervisor get this message across? This example illustrates why research data on this system of merit rating (or competence pay) show time and again that employees hardly perceive a relationship between their performance appraisal result and their bonus. There is hardly any “line of sight,” as Fig. 2 illustrates.

Figure 2 shows a core component of the expectancy theory on work motivation. Accordingly, the motivation to perform occurs as an individual perceives a relationship between his or her effort and performance level (expectancy) as well as a link between his or her performance level and one or more outcomes (instrumentality) that are attractive (valence). The figure highlights the principle of instrumentality and the often lacking line of sight—between the appraisal result and the resulting pay amount—as merit rating is applied. Not surprisingly, employees appear to be dissatisfied with their pay and consider the system (as well as the results it produces) to be unfair. The bonuses hardly recognize employees’ performance contributions, and supervisors often wonder whether the bonuses can ever motivate employees to perform better.

There is no recipe for success. But when quantitative yardsticks are used, performance appraisal is a much less vulnerable process. Such yardsticks relate to outcomes of performance or to the pursuit of particular behavioral actions (e.g., prescribed behavior patterns in risky work situations). Performance outcomes can be set through defining a job’s critical performance factors—two or three characteristics that represent the essence of a job (or job family) such as the number of billable training hours (for professional trainers) or the rate of recurring errors (for equipment repairers). Defining these factors is time-consuming; however, they can also be used for education and training, personal development, and career planning and can also be the focus of leadership behavior. Because critical performance factors should be vital to a job and sensitive to the job incumbents’ decisions and actions, they lend themselves to performance pay as well. Research and practical experience show that their line of sight is usually adequate.

## 4.2. Antecedent Conditions

A performance-related pay system requires a careful phase of preparation. At least the following four conditions should be met.

### 4.2.1. Study of Work

Through analyzing work processes within a unit or department, it should become clear whether work activities are adequately tuned to each other, work methods are well chosen, and the workforce is sufficiently trained. Within this context, critical performance factors may be set.

### 4.2.2. Norm Setting

For each performance factor, a specific target ought to be set. Research on goal-setting theory has revealed that difficult and specific goals (for individual tasks with single goals) lead to high performance as employees accept these targets and are provided with frequent feedback.

### 4.2.3. Performance–Pay Link

Performance-related pay may be awarded when a target has just been met, is clearly within reach, or has been surpassed. The performance–pay link may be proportional or either less or more than proportional. Expectancy theory would recommend rewarding achieved results as frequently as possible.

### 4.2.4. Control by Employees

Mutual trust appears to be one of the cornerstones for getting performance-related pay under way. Employees (and managers) concerned should be involved in implementing a new system—in its design, in its introduction, and/or in its daily administration. The more employees believe that they are in control of a system, the more they tend to evaluate the system as fair.

### 4.3. Size of Pay Differences

Suppose that the salary an employee earned this month is several euros or dollars less than the amount he made last month. Chances are great that the employee will notify the salary administration department. However, if the employee's most recent paycheck would have shown a slightly higher salary than before, he probably would not have inquired about the difference. Upward and downward pay differences are not each other's reverse in their behavioral effects. Some studies show that a pay increase of less than 3% is usually not perceived as significant by individual employees (regardless of the level of their previous earnings). A pay increase of 5% or more is considered to be worthwhile "to go for," for instance, through better performance. However, if performance-related pay is applied, the size of the bonus appears not to be as important as the performance–bonus link per se.

Yet people at work often tend to compare themselves with others regarding their contributions to the work and the rewards they receive in return. Social comparison theory and equity theory have revealed that persons tend to compare their contributions–reward ratios with those of others who are in the most comparable situations (i.e., a referent). An employee who is comparing a valuable, relatively scarce attribute (e.g., his or her pay) with a referent wants to come out just a little bit higher. Such an outcome is quite motivating. But it hurts when the outcome is that the employee comes out a little bit lower. It is small differences that count.

### 4.4. Some Contingencies

Compensation systems and strategies operate within the social and cultural climate of an organization (and within the society at large). Consequently, many factors have been shown to affect compensation, one of which is pay secrecy. Many employers are quite open about their pay practices, whereas many others prefer to keep these secret. Whatever the arguments for such secrecy, it has an impact on the climate of trust and does not prevent organization members from making earnings comparisons among themselves. Research data show that such comparisons are unavoidably based on incomplete information, which may result in the misperception that the difference with the earnings of immediate subordinates is too small and that the difference with the earnings of superiors is too large. This may cause pay dissatisfaction. However, it is important

to differentiate among three levels of "progressively open" information: the compensation system, its salary scales, and the level of individual earnings.

Also, differences in personality variables may be important. For example, individuals with higher scores on self-efficacy are more in favor of performance-related pay. Also, the higher the negative affect (i.e., the tendency to focus on negative life experiences that causes feelings of guilt and shame), the more disadvantageous effects a performance-related pay system will have.

## 5. CONCLUSION

Compensation in organizations provides a good example of multidisciplinary concerns. Compared with sociological, business administrative, and (especially) economic contributions, more psychological expertise would be most welcome. Recent work on the relationship between an organization's general business strategy and its compensation strategy calls for better theories (and more research) on the choice of particular pay patterns. Psychology may profit greatly from the groundwork laid in other disciplines.

### See Also the Following Articles

Job Analysis, Design, and Evaluation

### Further Reading

- Dickinson, A. M., & Gillette, K. L. (1993). A comparison of the effects of two individual monetary incentive systems on productivity: Piece rate versus base rate pay plus incentives. *Journal of Organizational Behavior Management*, 14, 3–83.
- Gomez-Mejia, L. R., & Balkin, D. B. (1992). *Compensation, organizational strategy, and firm performance*. Cincinnati, OH: South-Western Publishing.
- Lawler, E. E. (2000). *Rewarding excellence: Pay strategies for the new economy*. San Francisco: Jossey-Bass.
- Lawler, E. E. (2003). *Treat people right*. San Francisco: Jossey-Bass.
- Rynes, S. L., & Gerhart, B. (2000). *Compensation in organizations: Current research and practice*. San Francisco: Jossey-Bass.
- Thierry, H. (2001). The reflection theory on compensation. In M. Erez, U. Kleinbeck, & H. Thierry (Eds.), *Work motivation in the context of a globalizing economy*. Mahwah, NJ: Lawrence Erlbaum.
- Thierry, H. (2002). Enhancing performance through pay and reward systems. In S. Sonnentag (Ed.), *Psychological management of individual performance*. Chichester, UK: Wiley.



# Competence at Work

Angelo S. DeNisi

Texas A&M University, College Station, Texas, USA

1. The Concept of Competence at Work
  2. Measurement and Assessment of Competence
  3. Barriers to Effective Assessment
  4. Reaction to Assessments
  5. Developing and Improving Competence
  6. Conclusion
- Further Reading

## GLOSSARY

**contextual performance** Behaviors exhibited by employees that are not part of their formal job requirements but that contribute to overall organizational effectiveness; examples include volunteering for assignments, working with enthusiasm, and helping others at work.

**frame of reference training** A type of training in which raters are taught the definitions of various scale points on the rating scale rather than being taught to eliminate rating errors; the goal is to increase the accuracy of the ratings provided.

**halo error** A rating "error" manifesting itself as a high degree of consistency across ratings in various aspects of the job; this is assumed to occur due to a belief that good performance in one part of the job influences performance in other parts of the job, but eventually it became clear that this consistency might not be an error at all.

**incident diary** A log or record kept by the rater, noting all incidents of good and poor performance exhibited by an employee during a period of evaluation.

**leniency error** Another rating "error," in this case manifesting itself as ratings consistently above the rating scale mean value; eventually, it became clear that this too

might not be the result of a rating error but rather might reflect true levels of performance on the job.

Competence at work refers to the general evaluation of how individual employees (or teams of employees) contribute to the overall goals of the organization. Most typically, this evaluation is based on the judgments of some person (most often a direct supervisor), and formal evaluations are accompanied by documented ratings as well as feedback and discussion about ways in which to improve. Assessments of competence are often the basis for decisions about pay raises and promotions, but they can also be used as criterion measures against which to validate selection techniques. Underlying many of the purposes is the need to assess competence so as to help employees improve their performance on the job.

## 1. THE CONCEPT OF COMPETENCE AT WORK

An organization, whether private or public sector, hires employees to help the organization reach its goals. It is unusual, however, for every person hired to contribute to those goals equally, so it is important that we be able to assess the contribution of each employee (or group of employees) in the organization. There are a number of reasons why it is important that we are able to assess contributions in this way. One reason that is critical for employee selection is that we must be able to develop a

criterion measure against which to validate selection techniques. That is, for legal as well as practical reasons, it is important to know that the persons selected with a given technique will be those most likely to contribute to the organization's goals. But from a broader organizational perspective, we must be able to assess competence so that we can help employees to improve and expand the contributions they make to organizational goals. This is possible if we can assess relative strengths and weaknesses and then provide feedback about how to correct those weaknesses, and if we can demonstrate how improvements are tied to rewards such as pay increases, so that we can reward the employees accordingly. This requires that (a) we can define competence and the nature of that contribution and (b) we can assign some score or value to the contribution of each employee such that we can make some type of comparison across time and across employees.

The most traditional approach to defining competence at work is based on job analysis. That is, the purpose of job analysis is to specify what an employee is expected to do on the job, and the definition of competence on the job should probably be related to those requirements. But it is difficult to define competencies on the basis of specific tasks or operations carried out by employees. Such an approach would become extremely cumbersome given that we rarely have multiple employees on different jobs performing similar tasks. Thus, it makes more sense to define competencies in terms of more broadly defined contributions that underlie multiple tasks. As a result, in many cases, competencies are defined in terms of traits or characteristics of individuals that are believed to lead to success on a number of jobs. Furthermore, although a major portion of an employee's contribution will probably come in areas related to his or her job requirements, an employee can certainly make other types of contributions as well.

Any organization has certain tasks that must be carried out for the good of the organization such as employees being willing to help each other solve problems and employees volunteering to put in extra hours when crises arise. Any employee behaviors that make the workplace more pleasant are valuable. Yet none of these things is part of anyone's formal job description. Some scholars refer to these kinds of behaviors as organizational citizenship behaviors, whereas others simply refer to them as examples of contextual performance. Regardless of what they are called, they represent part of an employee's potential

contribution to the organization's goals, and they are part of any complete definition of competence. Although there has been disagreement over whether these things should be part of any evaluation of competence, most experts agree that these behaviors are considered when evaluation decisions are being made.

Finally, the operational definition of competence, usually expressed in terms of an evaluation instrument, can be stated in several different ways. The assumption that competence should be based on job analysis would lead to evaluation systems that focus on tasks that are to be carried out by the employees or on behaviors that employees should exhibit to be effective on the job. Others argue that the behaviors themselves are not as important as the outcomes of those behaviors. For example, we may be interested in whether salespeople follow up on sales calls, but what we really care about is whether or not they make the sales. Therefore, some have suggested that we define competency and assess contributions in terms of observable outcomes such as units produced and sales volume.

This approach can work only when the outcomes involved are under the direct control of the employees, and systems based on outcomes tend to work better when they are tied to specific goals that the employees must accomplish. That is, the exact nature of the contribution should be specified and should serve as a goal the employees try to reach during some time period. We could, for example, set a sales goal of 10,000 units or of \$1 million. This would then define what we mean as competence at work; it would mean meeting that goal.

This approach has the benefit of defining expectations in the clearest of terms, and it allows employees to judge how well they are doing. It also makes the entire process somewhat less subjective (although the nature and level of goals are typically negotiable), and this should make it more acceptable to all parties involved. But this approach is not without its problems. It is critical that the organization state goals and objectives in the right terms and not just in terms that can be easily quantified. Otherwise, the organization runs the risk of having all of its employees meet their goals while the organization itself fails to meet its goals.

The final way in which competency has been defined at work is based on the presumed underlying traits and abilities needed to carry out the job. For example, regardless of the exact job duties, we might desire employees to demonstrate initiative on the job. Furthermore, we might believe that initiative will help those employees to carry out their assigned tasks and meet any meaningful goals. Defining competency and

assessing performance in this way allows the organization to use one measurement instrument across a wide variety of jobs, and this is the most commonly found approach to assessing performance. Although this approach leads to the most subjective of evaluation systems with standards that are difficult to define, there is reason to believe that there are certain underlying traits (e.g., dependability) that truly do differentiate outstanding employees from merely competent employees.

Each of these approaches to defining competence at work has its advantages and disadvantages. Although no one approach is clearly the best, the definition used will drive many of the management initiatives undertaken by the organization, so it is a critical decision. Thus, the key is to understand the implications of each approach and then decide on which one is best for the particular organization.

## 2. MEASUREMENT AND ASSESSMENT OF COMPETENCE

After the best working definition of competence is determined, the next major issue is determining how competence is to be measured. As noted previously, it is often not feasible to assess competence based directly on tasks performed. In addition, although there are instances where we can count units produced or monitor scrap rates, these instances are relatively rare, and as noted previously, the use of more objective measures can have the effect of focusing the employees' attention on the wrong things. As a result, most organizations use some type of rating scale for assessing competence, and many organizations spend a great deal of time experimenting with new forms.

Unfortunately, there is little evidence that any one type of rating instrument provides consistently better outcomes than other types of instruments. Therefore, it makes sense for an organization to use what works best for the organization and not to pay a lot of attention to what others are using. This issue is further complicated by the fact that there is no clear consensus as to what should be the criterion in any evaluation of measurement techniques. Furthermore, it is probably more important to focus on performance management techniques to try to improve competence than to focus on appraisal techniques designed to better assess competence.

Some of the efforts to develop better rating instruments have focused on the nature of the actual rating scales or on the number of points on those scales and

the anchors used. Others have focused on exactly what is being assessed—whether behaviors, traits, or outcomes. But these efforts were hampered, in part, by the fact that there was no consensus as to how one should go about evaluating appraisal systems.

For a long time, the focus of improving appraisals was on the reduction of rating errors. These errors include things such as halo error (i.e., when someone who is rated highly on one aspect of a job is rated highly on all other aspects of the job) and leniency (i.e., when employee evaluations are inflated to be higher than they should be), and it was assumed that if there were less halo error and leniency in a set of evaluations, these would better reflect the true levels of competence of the employees being evaluated. But there is no reason to believe that a set of ratings that are lower, or are less intercorrelated, are necessarily better. It became clear that these “errors” might not be errors at all and that, in any case, they were generally unrelated to the accuracy of the evaluations provided.

Some scholars then began to assess rating accuracy directly. This made sense because it seemed reasonable to say that measures of competence were better if they accurately reflected the “true” levels of competence displayed by the employees. However, this approach requires a “true score” measure against which actual ratings can be compared, and such true scores are available only in laboratory settings.

More recently, it has become apparent that all of these efforts to reduce errors or improve accuracy are somewhat misplaced. In most organizations, the primary purpose of assessing competence is to allow employees to improve their levels of competence. In other words, the main purpose is to help employees improve the contributions they make to their organizations. Employees are more likely to improve if they perceive the assessment system as fair and accurate. To the extent that rating accuracy and the absence of what have been called rating errors can lead to perception of fairness and accuracy, these efforts are important. But they are important only as a means of getting to the ultimate goal—improvement.

Finally, there is the problem of how to deal with contextual performance when assessing competence. As noted previously, contextual performance refers to behaviors that benefit the organization but are not part of any employee's formal job requirements. The challenge here is not in deciding how to measure contextual performance (there are several scales available) but rather in deciding whether or not contextual performance should be assessed. If these behaviors

are included as part of a formal evaluation system, they will no longer be examples of contextual performance and instead will be part of the job duties. Would that be fair? Unfortunately, we do not have an answer yet.

### 3. BARRIERS TO EFFECTIVE ASSESSMENT

The solution to better assessment probably does not lie with the nature of the rating system used. Instead, the key component is the rater, that is, the person carrying out the evaluation of the employee's competence. Specifically, there are two issues involved. First, the rater must have the ability to provide fair and accurate evaluations. Second, and perhaps more important, the rater must be motivated to provide fair and accurate ratings. In other words, the rater must be able to recognize truly outstanding employees and then to make that recognition part of the public record.

Of course, the entire issue of what constitutes "fair and accurate ratings" is not a simple matter. We can define fair and accurate ratings as those that correctly describe employees' strengths and weaknesses and that assign the highest rating to the best performer, the next highest rating to the next best performer, and so on. There has been a great deal of research and discussion about how to determine rating accuracy, and many complicated formulas have been developed to help assess accuracy. However, the perception that a set of ratings is accurate is probably more important than the actual formula used to assess accuracy. It is this perception that will be critical for employees' belief as to whether a set of ratings is fair.

Attempts to improve rater ability have generally focused on training raters to be better judges. Consistent with the earlier view that reducing rating errors would lead to better evaluations, there was a great deal of interest in developing training programs to help raters reduce these errors. This interest gave way to developing training programs designed to help raters make more accurate ratings, usually referred to as "frame of reference" training. These programs emphasize helping raters to understand what constitutes good and poor performance in each area of competence being assessed. They are generally well received and have been found to improve rating accuracy in carefully controlled settings.

A different approach to improving rater ability deals with keeping performance logs or diaries. Formal appraisals are usually conducted no more than once

or twice a year. Therefore, a rater must be able to recall the behaviors of each employee over a period of at least 6 months to provide accurate and meaningful evaluations. Clearly, if a rater "forgets" an incident, that incident cannot become part of the employee's evaluation. If the rater forgets enough of these critical behaviors, the evaluation will become less accurate. Furthermore, if there were some tendency to forget incidents involving good performance, the employee who received the evaluation would perceive it as unfair even though the rater was providing the fairest evaluation possible. A simple solution to this problem is to have the rater keep a diary or incident log over the evaluation period. The rater writes down (or somehow notes) every incident of good or poor performance and then uses this record to make an evaluation later. Although raters do not always keep these diaries as conscientiously as they should, there is evidence that such diaries are effective when they are used.

But even if the rater has the ability to provide fair and accurate evaluations, he or she might not always choose to act on that ability. Rater motivation is much more complex and less well understood than is rater ability. Why would a rater choose to be inaccurate? There are many reasons. Raters are human and so are prone to human biases. Many will favor people they like over people they dislike. Other forms of potential bias are along the lines of race, gender, and even age. Also, a rater may assign a lower rating than should be assigned just to establish his or her power to do so. Organizational policies and bureaucratic procedures may also make it easier for the rater to assign an inaccurate rating in some cases.

Raters will do what is in their own best interests. Assuming that they have the ability to provide fair and accurate ratings, they will actually do so only when they believe that either they will be rewarded for doing so or, at least, that they will not be punished for doing so. Organizational policies play a large role in this decision. For example, in some organizations, raters are required to provide excessive documentation if they give someone a truly outstanding evaluation. Faced with this requirement, raters may simply decide that it makes more sense to give a slightly lower evaluation.

In conclusion, raters are probably much more critical to the process of evaluating competence than is the instrument used. Raters must be trained, or otherwise helped, so that they have the ability to provide fair and accurate ratings, and policies must be such that raters will also be motivated to provide those ratings. If raters are capable of being fair and accurate as well as

motivated to be fair and accurate, the employees will be more likely to perceive the ratings as fair and accurate. Employees' reactions to these ratings, however, are critical to the effective management of performance at work.

#### 4. REACTION TO ASSESSMENTS

As noted previously, the ultimate purpose of assessing competence at work is to motivate employees to improve their performance on the job and so increase their contributions to the organization. The issues of rater ability and motivation were discussed, and these factors are relevant for the employees as well. However, the question of employees' ability to improve performance relates more to selection and training activities in the organization and is not discussed in this article. The issue of employee motivation to improve, on the other hand, is central to any discussion of assessing and improving competence at work.

Following an evaluation, the desired outcome is for the employee to take the feedback he or she receives and to work to improve. If the feedback is positive (the employee is generally doing fine), this may require maintaining effort coupled with some refinement of where to allocate effort. This is typically not a problem. However, when the feedback is more negative (the employee needs real improvement), there are a number of problems that can occur.

A fairly typical reaction to negative feedback is simply to ignore the feedback. An employee might decide that the feedback is not fair or that the poor performance is somehow not the employee's fault. In any case, no improvement will follow this reaction. For the chances of acceptance to be high, the evaluation on which the feedback is based must be perceived as fair. That is, the evaluation itself must be seen as fair, and the procedures used by the rater to develop that evaluation must be seen as fair. In addition, the feedback must be delivered in a constructive way that does not threaten the self-esteem of the employee. Finally, the feedback should focus on the behaviors at work rather than on the person. Research suggests that feedback is not nearly as effective in changing behavior as is typically assumed, so it is critical that the employee accepts the feedback and works to improve.

During recent years, there has been a great deal of interest in a particular approach to competence assessment that is concerned with the issue of feedback. Multisource assessments, more commonly referred to

as "360 degree appraisals," were designed to provide employees with feedback from peers, subordinates, and superiors and to compare this feedback with their own self-assessments. The assumption underlying this approach is that various individuals who have different relationships with an employee will focus on different behaviors. By providing feedback from all of the various sources, the employee should have a clearer picture of his or her strengths and weaknesses. There is actually little research about the effectiveness of 360 degree appraisals, and the results of the research that does exist are somewhat mixed. Nonetheless, the notion that individuals from various perspectives might have different views of an employee's competence, and that all of these may be useful, is an interesting idea.

#### 5. DEVELOPING AND IMPROVING COMPETENCE

The penultimate goal of assessing competence is to improve competence. In truth, the ultimate goal of the process is to improve the performance of the entire organization. This must begin, however, with helping employees to accept and act on feedback. Once the employee is motivated to improve, the organization can take steps to aid in that improvement. In fact, if the organization does not take some steps in this direction, it is unlikely that real change will occur.

For example, if it is important to build and develop competence, feedback needs to be a regular thing. That is, feedback about performance should not occur just once or twice a year; it should occur constantly. Whenever an employee exhibits desirable or undesirable behavior on the job, someone should note this; in the case of undesirable behavior, the employee should also be told how to improve. Having supervisors set improvement goals with employees and regularly monitor progress toward those goals is also essential for real improvement. It is also important that supervisors and managers be constantly aware of cases where they might actually be punishing employees who perform well. A classic example is asking an employee to perform a "rush job" and then "rewarding" that employee by assigning him or her an additional job to perform.

It is hoped that if all individual employees improve their level of competence on the job, the organization as a whole will somehow be more effective. The truth is that this upward aggregation of improvement is not inevitable and requires close monitoring. Management, at the highest levels, must ensure that the goals and

directions set for individual employees at every level are consistent with, and supportive of, the organization's strategic goals. Therefore, it is critical that organizational goals be set with some recognition of the human resources available to meet them and that tactics for improving competence consider how improvement can help to serve those strategic goals. It is also important for high-level managers to recognize that they must be able to leverage the competence they have. That is, if an organization has extremely creative employees, its managers should pursue strategies that emphasize the importance of creativity. In this way, an organization can ensure that improvements in individual competence can translate to improvements in organizational effectiveness.

## 6. CONCLUSION

An organization improves its effectiveness by improving the competence of individual employees at work so that these individuals can make greater contributions to the organization's strategic goals. This process begins with the definition of competence at work and what exactly is included in that definition. Once a definition is selected, the organization must develop the means by which to determine the current level of employee competence versus where the employee needs to be. Although this might sound simple, there are a number of barriers to effective assessment. Some of them stem from the characteristics of the persons doing the assessment, whereas others stem from organizational policies that may (unintentionally) make this assessment more difficult.

But it is not enough to develop sound assessment techniques. Those assessments must be fed back to the employee in ways that will encourage him or her to work toward improvement. Organizational policies

related to goal setting and regular feedback can help to ensure that this occurs. In addition, it is important to coordinate goals and incentives at every level in the organization so that improved competence at the individual level can translate to improved organizational effectiveness.

## See Also the Following Articles

Job Analysis, Design, and Evaluation ■ Organizational Participation

## Further Reading

- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include assessments of contextual performance. In N. Schmitt, & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco: Jossey-Bass.
- Colquitt, J., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology, 86*, 425–455.
- DeNisi, A. S. (2000). Performance appraisal and control systems: A multilevel approach. In K. Klein, & S. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 121–156). San Francisco: Jossey-Bass.
- Kluger, A. N., & DeNisi, A. S. (1996). The effects of feedback interventions on performance: Historical review, meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, 72–107.
- Murphy, K. R. (1991). Criterion issues in performance appraisal research: Behavioral accuracy versus classification accuracy. *Organizational Behavior and Human Decision Processes, 50*, 45–50.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.



# Competition in Sport

**Daniel Gould**

Michigan State University, East Lansing, Michigan, USA

**Cristina Rolo**

University of North Carolina, Greensboro, North Carolina, USA

1. Introduction
  2. Competition and the Competition Process
  3. A Model for Understanding Competition
  4. Positive and Negative Effects of Competition
  5. Competitive Stress in Sport
- Further Reading

## GLOSSARY

**burnout** A psychological, physiological, and/or emotional withdrawal from sport due to exhaustion, depersonalization, and low accomplishment feelings resulting from chronic stress.

**competition** According to Martens, a process whereby an individual's performance is compared with some standard of excellence in the presence of at least one other person who is aware of the criterion for comparison.

**state anxiety** Feelings of apprehension and tension associated with activation of the organism.

**stress** An athlete's perception of the imbalance between the environmental demands placed on him or her and the athlete's response capacity and resources for meeting those demands.

**trait anxiety** A personality disposition that predisposes individuals to perceive evaluative contexts as more or less threatening and to respond with varying levels of state anxiety.

Competition is a process that has important positive and negative consequences for athletes. Positive consequences can include increased confidence, motivation,

and satisfaction, whereas negative consequences can include stress and burnout. The degree to which competition has positive or negative consequences depends on the competitive context (e.g., the importance placed on performance and uncertainty) and the personality (e.g., trait anxiety, self-esteem) and perception of the athlete involved. Positive coaching strategies and stress management techniques can be used to reduce or cope with competitive stress.

## 1. INTRODUCTION

### 1.1. Competitive Sport: A Worldwide Phenomenon

Competitive sport is a worldwide phenomenon of enormous popularity. Epitomized by the Olympic Games and World Cup soccer championships, millions of individuals watch these events with tremendous enthusiasm. Moreover, countless others participate in various levels of competitive sport every year.

Competitive sport is not without controversy. Some argue that it is an excellent training ground for psychological attributes such as leadership, confidence, and teamwork, whereas others suggest that it leads to excessive anxiety, immoral behavior, and aggression.

Because of the persuasiveness of competitive sport, better understanding competition and the competitive process has been an important task for those in applied psychology. Three decades of sport psychological

research on competition have provided a solid understanding of the area.

### 1.2. Purposes

This article has three purposes. First, competition is defined and the competitive process is discussed. Second, the issue of whether sport competition has positive or negative effects on participants is examined. Finally, competitive stress, its sources, and its consequences are discussed.

## 2. COMPETITION AND THE COMPETITION PROCESS

Based on the groundbreaking work of Deutsch in 1949, early definitions of competition focused on situations where rewards were distributed unequally based on the performance of participants. Although this definition did much to guide early research in the area, its shortcomings (e.g., the fact that individuals do not always define rewards in the same way and, hence, that a person might lose an event but define his or her personal improvement as rewarding) led to the development of other definitions.

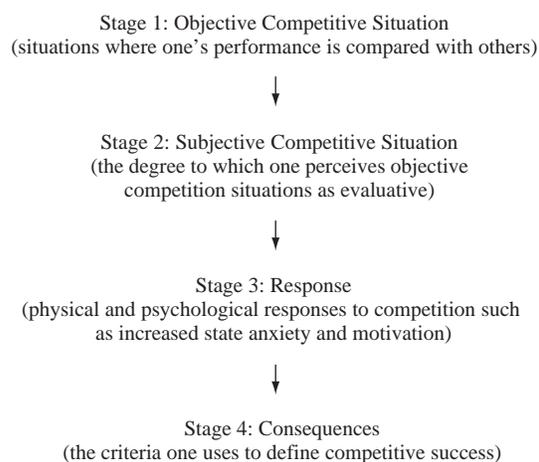
Today, competition is defined as a social process that encompasses a sequence of stages rather than a single event. Martens did much to advance the scientific study of competition when, in 1975, he defined it as a process where “an individual’s performance is compared with some standard of excellence in the presence of at least one other person who is aware of the criterion for comparison.” This differs from the reward definition in that the person is the focal point of the competition process and can influence the relationship among the various stages of competition. Primary emphasis is also placed on the individual’s perception of the competitive context and on how personality factors influence one’s competitive experience.

Standards of excellence involved in the competition process can be self-referenced (e.g., set a personal best in a 100-meter swim competition) or other-referenced (e.g., beat a particular individual in the competition). Considerable attention in contemporary sport psychology focuses on the utility of making self- versus outcome-oriented comparisons. Self-referenced standards are often emphasized because they are more in an individual’s control, and have been found to be associated with long-term motivation and greater satisfaction, compared with other-referenced comparisons.

## 3. A MODEL FOR UNDERSTANDING COMPETITION

Guided by the process definition of competition, Martens also developed a four-stage model of competition (Fig. 1). An inspection of this model reveals that the objective competitive situation (Stage 1) begins the competitive process and consists of a situation where an individual’s performance is compared with others in the presence of at least one other individual. For example, for most individuals, entering a gymnasium and seeing a basketball at midcourt and one’s name on the scoreboard would signal a competitive situation. However, not everyone would perceive this situation as competitive, for example, an elderly person who has no interest in or knowledge of basketball. Hence, Stage 2 of the process involves an individual’s subjective perception of the objective competitive situation. Research has shown that an individual’s personality orientations, particularly his or her levels of competitiveness or achievement goals (i.e., orientation to judge ability relative to oneself or a comparison with others), influence the degree to which the person sees situations as competitive.

After an individual appraises a situation as competitive, he or she may approach or avoid it or may become motivated, anxious, or excited. Thus, Stage 3 of the model focuses on an individual’s psychological-physiological response (e.g., anxiety, increased muscle tension, enhanced motivation). Interestingly, research also shows that it is not the absolute level of anxiety



**FIGURE 1** Martens’ model of competition. Reprinted from Martens (1975).

that influences an athlete's performance; rather, it is whether the individual views that increased anxiety as facilitative and helpful or as debilitating and a hindrance to performance. The athlete's perception, then, not only influences whether he or she views competition as anxiety provoking but also influences whether that anxiety is seen as helpful or hurtful.

Finally, the positive or negative consequences are the fourth stage in the competition process. An individual wins or loses on the scoreboard or perceives that he or she performed well or not. However, as with the three preceding stages, an athlete's perception of the consequences is more important than the objective outcome.

This model of competition is important because it shows that competition can best be understood as a series of stages. Moreover, it places an athlete at the center of the competitive process, with his or her personality dispositions interacting with environmental considerations.

#### **4. POSITIVE AND NEGATIVE EFFECTS OF COMPETITION**

The competitive ethic is a driving force in contemporary sport. It is common to hear coaches, athletes, and sports journalists say positive things such as "competition brings out the best in people" and "she is a fiery competitor." At the same time, an overly competitive, win-at-all-costs mentality is blamed for violence, rules infractions, and unsportsmanlike behavior. Hence, sports competition appears to be a double-edged sword, cutting both ways and having both positive and negative effects on participants.

Most researchers do not view competition as inherently good or bad today. Research shows that it can have both positive and negative effects on participants. For instance, in some studies, competition has been shown to lead to negative effects such as increased aggression and decreased performance. In other investigations, competition has been found to facilitate motivation and lead to improved performance. Whether these effects are positive or negative depend greatly on the competitive context and the emphasis that sport leaders and coaches place on competition and its meaning. For example, the quality of adult leadership has been shown to be a crucial determinant of whether competition has positive or negative outcomes for children. In their classic youth sports coaching research, Smoll and Smith found that children playing for coaches trained in a positive approach (i.e., who focused on encouragement and giving praise) exhibited

higher self-esteem and lower anxiety and were less likely to drop out of sport. When coaches were not trained to be positive and encouraging, children involved in competitive sport did not experience increased motivation, lower anxiety and enhanced self-esteem.

Much of the sport psychological research focusing on competition has been conducted with young athletes in entry-level programs. Less research on competition has been conducted at different levels of competition (local, national, and international). However, recent research with elite athletes shows that these outstanding performers are highly competitive. Their standards of comparison are both self- and other-referenced, they are both intrinsically and extrinsically motivated, and they have high perceptions of ability.

Champion athletes have also been found to approach competition differently across various phases of their careers. For instance, in the classic research by Bloom in 1985, elite athlete development was retrospectively traced across the careers of champion athletes. Bloom's research, as well as more recent research, found that these individuals did not begin their careers with an emphasis on competitive outcome (i.e., winning medals and defeating others). Rather, the focus was on having fun, learning fundamentals, and being active. It was only later that they approached competition in a very serious fashion. In addition, for much of their careers, they focused on long-term development rather than short-term competitive results.

#### **5. COMPETITIVE STRESS IN SPORT**

One of the most studied aspects of competitive sport has been its association with competitive stress. Levels of stress experienced by young athletes, sources of stress, and managing stress all have been topics of considerable interest to sport psychology researchers.

##### **5.1. Stress in Sport**

Stress has been defined as an athlete's perception of the imbalance between the environmental demands placed on him or her and the athlete's response capacity and resources for meeting those demands. For example, a golfer may face a situation where he or she needs to sink a crucial 15-foot putt to win a tournament. If the golfer perceives the demands as exceeding his or her capabilities, the result is increased competitive state anxiety, that is, feelings of apprehension and tension accompanied by physiological activation. Furthermore,

heightened levels of state anxiety, especially if they are perceived as debilitating, have been associated with poor performance and lower levels of enjoyment and satisfaction.

## 5.2. Sources of Stress in Athletes

Identifying sources of stress experienced by athletes engaging in competition has been a topic of interest to sport psychologists. This research has generally shown that although there are a variety of specific stress sources that an athlete can experience (e.g., self-doubts about performance capabilities, team selection), these fall into two general situational categories: (a) the importance placed on performance and (b) uncertainty. Specifically, the more importance that an athlete perceives is placed on an event, the more state anxiety is experienced. Similarly, the greater the degree of uncertainty (whether about performance or nonperformance issues) that the athlete perceives, the more competitive state anxiety is experienced.

However, situational factors are not the only class of factors influencing competitive stress responses. Consistent with Martens' model of competition, personality dispositions have been associated with elevated stress responses. Specifically, an athlete's level of trait anxiety (i.e., a personality orientation that predisposes the individual to view social evaluation and competitive contexts as threatening) influences his or her level of state anxiety. Performers with high trait anxiety consistently respond with greater levels of state anxiety in competitive situations. Self-esteem has also been consistently associated with levels of competitive state anxiety experienced in socially evaluative sport contexts. In competitive situations, athletes with low self-esteem experience higher levels of state anxiety than do those with high self-esteem. Finally, more recent studies have found relationships between increased anxiety and high levels of hardiness, perfectionism, and self-presentation concerns as well as low social support, although these findings need to be further verified in additional studies.

## 5.3. Are Young Athletes Placed under Too Much Stress?

For a number of years, researchers have been concerned with the levels of stress experienced by athletes, especially young athletes engaged in competitive sports. Thus, researchers have compared the amounts of state anxiety experienced during practices with those experienced during competitive sport situations. The thinking

behind these comparisons is that competition is more anxiety provoking than are practices, but it is important to note that at times practices can include considerable social evaluation (e.g., team selections). Although the results of these studies have consistently revealed that most young athletes experience more anxiety during competitions than during practices, the most important finding is that the vast majority of children do not experience excessive levels of state anxiety during competitions. However, certain children in certain situations do experience high levels of competitive stress and anxiety.

For example, in a set of classic studies of competitive youth sport participants, Scanlan and colleagues assessed the state anxiety levels of youth sport participants during practices (i.e., nonevaluative contexts) and compared these levels with those during competitions (i.e., evaluative contexts). A variety of personality and background factors were also assessed. Results of this research led to the general conclusion that most young athletes did not experience excessively high levels of state anxiety during competitions. Certain children in certain situations did experience high levels of state anxiety. These children were characterized by high competitive trait anxiety, low self-esteem, less fun, less satisfaction with performance, low personal performance expectancies, and worries about failure and adult evaluation.

In a related area of research, "burnout" or withdrawal from competitive sport has been studied. In these studies, burnout has been defined as a psychological, physiological, and/or emotional withdrawal from sport due to exhaustion, depersonalization, and low-accomplishment feelings resulting from chronic stress. Results reveal that a small percentage of sport participants experience burnout of sport and that chronic stress plays an important role in the burnout process. Moreover, the chronic stress results from physical overtraining and/or psychological factors such as overbearing parents or coaches, a lack of athlete multidimensional identity development, a perfectionistic personality orientation, high trait anxiety, and low self-esteem.

## 5.4. Managing Competitive Stress

Because competitive sport can be stressful, and high levels of stress have been associated with inferior performance and decreased enjoyment and satisfaction, researchers have been very interested in helping athletes to manage stress.

One important but often overlooked class of techniques for managing competitive stress is environmental engineering. With environmental engineering

techniques, coaches, adult leaders, and significant others can influence athlete stress levels by increasing the importance placed on competition (e.g., giving a fiery pregame pep talk, repeatedly emphasizing the importance of victory) or by creating uncertainty (e.g., not announcing starting lineups for games, basing liking for a child on performance). Or, they can reduce competitive stress by reducing the importance placed on performance (e.g., not giving a pregame pep talk, not repeatedly emphasizing the importance of victory) or by increasing certainty (e.g., announcing starting lineups early, not basing liking for a child on performance). Training youth sport coaches to adopt a positive and encouraging coaching orientation, rather than a negative or critical one, can reduce the levels of competitive stress experienced.

A number of techniques, mirroring general stress management research in psychology, have been used successfully by athletes to manage their competitive stress. These include cognitive anxiety reduction strategies (e.g., imagery, meditation), somatic anxiety reduction techniques (e.g., progressive muscle relaxation, breath control, biofeedback), and multimodal techniques (e.g., stress inoculation, cognitive-affective stress management training). Competitive athletes have also been found to use a variety of problem-focused (e.g., time management, goal setting) or emotion-focused (e.g., controlled breathing, relaxation training) coping techniques to control the stress of competition.

Relative to the utility of teaching stress management to athletes, sport psychologists emphasize the need for long-term training/practice and the incorporation of these techniques into the actual competitive setting (e.g., learning how to relax while an athlete actually runs vs using relaxation techniques the night before a race). Finally, research has also revealed that there is an optimal recipe of emotions that lead to superior sport performance, so that using stress management techniques to eliminate all stress is counterproductive. Thus, competitive athletes must know their optimal levels of emotional arousal and arousal-related emotions needed for best performance and then use stress reduction and enhancement strategies accordingly.

## See Also the Following Articles

Anxiety and Optimal Athletic Performance ■ Decision Making in Sport ■ Goal Setting and Achievement Motivation in Sport ■ Overtraining and Burnout in Sports ■ Performance Slumps in Sport: Prevention and Coping ■ Psychological Skills Training in Sport ■ Self-Confidence in Athletes ■ Sport Psychology, Overview ■ Stress

## Further Reading

- Bloom, B. S. (Ed.). (1985). *Developing talent in young people*. New York: Ballantine.
- Deutsch, M. (1949). An experimental study of the effects of cooperation and competition upon group processes. *Human Relations*, 2, 199–231.
- Gill, D. (2000). *Psychological dynamics of sport and exercise* (2nd ed.). Champaign, IL: Human Kinetics.
- Gould, D. (1993). Intensive sport participation and the prepubescent athlete: Competitive stress and burnout. In B. R. Cahill, & A. J. Pearl (Eds.), *Intensive sport participation in children's sports* (pp. 19–38). Champaign, IL: Human Kinetics.
- Hanin, Y. L. (Ed.). (2000). *Emotions in sport*. Champaign, IL: Human Kinetics.
- Horn, T. (2002). *Advances in sport psychology* (2nd ed.). Champaign, IL: Human Kinetics.
- Jones, G., Hanton, S., & Swain, A. (1994). Intensity and interpretation of anxiety symptoms in elite and non-elite sports performers. *Personality and Individual Differences*, 17, 657–663.
- Martens, R. (1975). *Social psychology of sport*. New York: Harper & Row.
- Scanlan, T. K. (1986). Competitive stress in children. In M. R. Weiss, & D. Gould (Eds.), *Sport for children and youth* (pp. 113–118). Champaign, IL: Human Kinetics.
- Smith, R. E., Smoll, F. L., & Curtis, B. (1979). Coach effectiveness training: A cognitive-behavioral approach to enhancing relationship skills in youth sport coaches. *Journal of Sport Psychology*, 1, 59–75.
- Smoll, F. L., & Smith, R. E. (Eds.). (2002). *Children and youth in sport* (2nd ed.). Dubuque, IA: Kendall/Hunt.
- Weinberg, R. S., & Gould, D. (2003). *Foundations of sport and exercise psychology* (3rd ed.). Champaign, IL: Human Kinetics.



# Conflict within Organizations

Michael Ryan and Stephanie L. Castro

Florida Atlantic University, Boca Raton, Florida, USA

1. Introduction
  2. The Conflict Process
  3. Conflict Management Strategies
  4. Structural Conditions Affecting Conflict
  5. Conflict Outcomes
  6. Third-Party Intervention
  7. Conclusion
- Further Reading

each individual's) interests are violated, or in danger of being violated, by the other individual.

**judgment conflict** Conflict between parties where one party's assertion about something is inconsistent with another party's assertion about it.

**normative conflict** Conflict between parties where one party perceives that the other party has exhibited behavior that is believed to be wrong.

## GLOSSARY

**conflict episode** A single engagement between two parties in conflict over some issue.

**conflict process** The series of activities from initiating conflict to resolving it.

**distributive outcome** Zero-sum resolution to a conflict in which the outcome is perceived as fixed and the purpose of the conflict is to distribute the outcome in some proportion between the parties (e.g., 50/50, 60/40, 100/0).

**external conflict** Conflict existing between an individual or group and another individual or group.

**goal conflict** Conflict between parties where one party has a goal that is mutually exclusive with a goal of another party.

**integrative outcome** Resolution to a conflict in which the parties work to create an outcome that is compatible with both parties' goals.

**internal conflict** Conflict existing within an individual who is experiencing mutually exclusive desires or beliefs.

**interpersonal conflict** A state that exists between two individuals with valued interests in which one individual's (or

Organizational conflict is a state that exists between two parties within an organization or between organizations, each with valued interests, in which one party's (or each party's) interests are violated, or in danger of being violated, by the other party.

## 1. INTRODUCTION

Both within and between organizations, individuals interact regularly and almost as regularly experience conflict with one another. Sometimes the conflict between individuals is minor and unnoticed, whereas other times it is heated and drawn out. Whatever the case, the characteristics of conflict between individuals are important to understand for the sake of the individuals and the organizations where they work. This article provides a general understanding of interpersonal conflict in organizations and highlights some of the most recognized ideas that have been presented in research on the subject.

Interpersonal conflict is a state that exists between two individuals with valued interests in which one individual's (or each individual's) interests are violated, or in danger of being violated, by the other individual. It is an external form of conflict, as opposed to internal conflict where two or more of an individual's interests are mutually exclusive (e.g., role conflict, conflict of interest).

Interpersonal conflict occurs between parties with opposing interests and can take three forms: normative, judgment, and goal. Normative conflict exists when one individual is offended by another individual's behavior and believes that the latter should have behaved differently. Judgment conflict occurs when the individuals disagree over a factual issue (e.g., how a particular event transpired). Goal conflict is a situation in which two individuals are seeking mutually exclusive outcomes (e.g., how a slice of cake will be divided between them).

In 1992, Thomas, an extremely influential figure in conflict research, divided conflict into four factors that affect one another directly or indirectly: structural conditions, the conflict process, outcomes, and third-party intervention. Structural conditions make up the context of the situation in which conflict occurs and include the relatively stable factors of the parties and their surroundings. When conflict is triggered, the conflict process ensues with interactions between the parties involved. Both experiences and behaviors of the parties affect the process. Conflict outcomes are the result of the conflict process (specific episodes or series of episodes) and can include task outcomes (relevant to the matter of the conflict) and social system outcomes (relevant to the organization, the relationship between the parties, etc.). During the process, a third party may intervene. Third-party intervention is the involvement of an individual or a group whose primary concern is the resolution of the conflict as opposed to the substantive concerns of the conflicting parties.

Some third parties may, on resolution, change the structural conditions of the context, thereby influencing the resolution of future conflict. An example might be a manager who helps to resolve a dispute between two employees and then creates a policy to handle similar disputes in the future. Aside from these primary effects, feedback effects are likely to occur. For instance, the social system outcomes of a conflict episode are likely to alter structural conditions such that no two conflict episodes occur in identical environments.

## 2. THE CONFLICT PROCESS

When two individuals are experiencing conflict, the conflict process is the most apparent and explicit aspect of all that goes on between the parties. They are likely unaware of the contextual factors and their individual characteristics that set the conflict in motion. Although they have an idea of what they want out of the conflict, they are probably not considering all of the effects that the conflict will have beyond their individual goals. It is likely that each party is most interested in how to behave, and in how the other party should behave, in dealing with the conflict. The series of activities that run from initiating conflict to resolving it is referred to as the conflict process. The most popular research on the conflict process suggests that the conflict process is best observed at the level of the conflict episode. A conflict episode is a single engagement between two parties in conflict over some issue.

Two researchers developed and presented influential models of the conflict process: Pondy in 1967 and Thomas in 1992. Although both models have similar aspects, Thomas's model gives attention to additional characteristics of the process such as the interaction of emotions, rational reasoning, and normative reasoning. The main aspects of this model are awareness that conflict exists, a series of thoughts and emotions leading to intentions that determine how the conflict will be handled, and interaction between the parties beginning with the first party's behavior and continuing with the other party's reaction. This leads to a continuous exchange cycle between the parties through thoughts/emotions, intentions, and behavior that ends with resolution and outcomes.

Thomas defined conflict as "the process that begins when one party perceives that the other has negatively affected, or is about to negatively affect, something that he or she cares about." At some point before an individual becomes involved in handling conflict with another person, that individual must first recognize that conflict exists. This point of first recognizing conflict is referred to as awareness. There will often be an event that triggers awareness such as an action or a statement made by the opposing party.

On becoming aware of conflict, an individual then begins to make sense of the situation by considering what is at stake and how the process should be handled. One aspect of this sense-making is defining the issue of the conflict. In addition to recognizing the issue at stake, an individual will begin to consider what

outcomes are likely to result in a resolution that is acceptable to both parties.

Having defined the issue and considered possible settlements, an individual will then begin to reason both rationally and normatively in considering ways in which to handle the conflict that (a) are likely to result in efficiently and effectively achieving a beneficial resolution and (b) are perceived as acceptable and fair by the social environment. For example, if two individuals are in conflict over purchasing the last of a particular toy in a store before Christmas, the issue is apparent: Which of the two will get to purchase the toy? Rationally, one party may consider that attacking the other party and stealing the toy would efficiently and effectively result in a beneficial outcome. However, normatively, the social environment would not approve of this behavior. Therefore, the party should consider a more acceptable strategy for acquiring the toy.

The writer Oscar Wilde once said, “Man is a reasonable animal who always loses his temper when he is called upon to act in accordance with the dictates of reason.” Accordingly, individuals engaged in conflict are often affected by their emotions in addition to considering rational and normative reasoning. Both positive and negative emotions can affect how the opposing party’s behavior is interpreted and can amplify or dampen the influence of rational and normative reasoning. All of this will filter down to influence behavior. For instance, anger can result in aggression and negative interpretations of behavior, anxiety can lead to withdrawal, and innocent humor can decrease the negative effects of aggression and anger.

The interaction of an individual’s reasoning and emotions leads to a set of strategic and tactical intentions for handling the conflict episode. Strategic intentions refer to the overall goal one party has for the situation such as cooperating with or competing against the other party. Research on strategic intentions often refers to them as conflict management strategies and is discussed in the next section. Tactical intentions, on the other hand, are more specific goals or plans for achieving a beneficial outcome such as using a calm tone in presenting a case to keep the other party from becoming angry and focusing on the benefits that a proposed solution has for the other party. These intentions serve as guides for an individual’s behavior in managing conflict.

The intentions that an individual forms for dealing with a conflict episode will influence that party’s behavior. After the first party expresses behavior, the other party experiences a similar process of reasoning and experiencing emotions, forming intentions, and

expressing behavior as a response to the interpretation of the first party’s behavior. This is not necessarily a turn-based activity, although it can be in certain negotiation situations. Each party responds to the other party through the stages of thoughts/emotions, intentions, and behavior in a continuous interaction between the parties until some outcome is reached. This might not be the end of the conflict between these parties; however, it is the end of the conflict episode. The outcomes of conflict are discussed in a later section.

### 3. CONFLICT MANAGEMENT STRATEGIES

Five strategies for managing conflict have been used throughout numerous research studies on conflict: avoiding, accommodating, competing, compromising, and collaborating. These strategies are coordinated on a two-dimensional grid with the axes labeled either “concern for own interests” and “concern for other’s interests” or “assertiveness” and “cooperativeness” (Fig. 1). Thus, avoiding would represent either a low concern for self-interests and for other’s interests or low levels of assertiveness and cooperation, respectively.

The first conflict management strategy is accommodating, also referred to as obliging. Figure 1 shows that this strategy is low in assertiveness and high in cooperativeness—a lose–win strategy. In this case, one party chooses to allow the other party to satisfy its interests

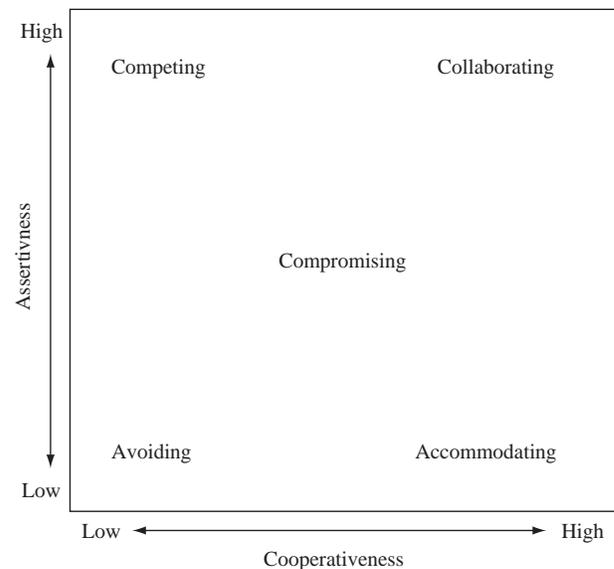


FIGURE 1 Conflict management strategies.

completely. One party may choose accommodation for several reasons. For example, the party may be unwilling to compete, may want to minimize losses, may consider the other party's interests as more valuable, or may seek to prevent damage to the relationship between the parties. Accommodating can be an active strategy in which one party helps to satisfy the other party's interests. It can also be a passive strategy in which inaction results in the satisfaction of the other party's interests.

The second conflict management strategy is competing or dominating. This is the opposing strategy to accommodating and is classified as high in assertiveness and low in cooperativeness—a win-lose strategy. Competing is an active strategy in which one party attempts to satisfy its own interests, usually by preventing the other party from satisfying its interests when both parties' interests are perceived as mutually exclusive. Competing may be more effective when quick and decisive action is vital, unpopular but important decisions must be made, an organization's welfare is at stake and the right course of action is known, and/or the other party takes advantage of noncompetitive behavior.

Compromising or sharing is the third strategy shown in Fig. 1. In this case, one party is moderately assertive and cooperative. In contrast to either party winning or losing, compromising seeks a "middle ground" in which each party wins in part and loses in part. Compromising may be perceived as a weaker form of collaborating, but the use of compromising assumes a fixed amount of resolution that must be distributed, whereas collaborating seeks to determine both the size of the resolution and how it is to be divided. Compromising may be effective when opponents with equal power are committed to mutually exclusive goals, the issue is complex and at least a temporary settlement must be reached, interests cannot be fully sacrificed but there is no time for integration, collaborating and competing fail, and/or goals are important enough to fight for but not so important that concessions cannot be made.

The fourth conflict management strategy is avoiding. This strategy reflects low assertiveness and low cooperation—a lose-lose situation. With this strategy, one party chooses not to engage the other party in resolving the conflict in such a way that neither party is able to satisfy its interests. Avoiding may be a good strategy when an issue is trivial, there is no chance to satisfy a party's own interests, the other party is enraged and irrational, and/or others can resolve the conflict more easily.

The last of the conflict management strategies is collaborating or integrating. As the term implies, this strategy has an integrative focus. Parties using the

collaborating strategy are classified as being both highly assertive and cooperative—a win-win attitude. In this case, the two parties attempt to work with each other to develop a resolution that will completely satisfy the interests of both parties. Although much of the research on conflict has proposed that collaborating is always the ideal conflict management strategy, others have suggested that there may be situations in which other strategies are more effective. Collaborating may be a more effective strategy when each party's interests are too important to be compromised, full commitment to the resolution by both parties is desired, and/or there is sufficient time to resolve the conflict by integrating both parties' interests.

Research on conflict management strategies has comprised a large portion of the total research on interpersonal conflict. The five strategies outlined here have become somewhat canonical in conflict research. Much research has validated the existence of these strategies and shown the advantage of a two-dimensional model over a one-dimensional (two-strategy) model, but there have been no remarkable attempts to improve on or add to this model of conflict management. Although there is much room for more research to be done in this area of interpersonal conflict, it has not suffered from a lack of attention as other areas of conflict have.

#### **4. STRUCTURAL CONDITIONS AFFECTING CONFLICT**

The conflict process does not occur in a void; there are structural conditions that affect the process. Both the contextual situation and the parties involved have characteristics that may affect the various factors in the conflict process such as the choice of a conflict management strategy, inclusion of a third party, and norms for behavior. In 1992, Thomas suggested, "Whereas process refers to the temporal sequence of events that occurs in a system, structure refers to the parameters of the system—the more or less stable (or slow-changing) conditions or characteristics of the system that shape the course of those events." Research has revealed several structural factors in both individuals and organizations that may have an effect on the conflict process. For instance, Wall and Callister's review of conflict literature in 1995 discussed several individual and issue-related factors affecting conflict.

### 4.1. Individual Characteristics

Individuals constitute the primary parties in interpersonal conflict; therefore, it is likely that characteristics of those individuals will influence how the conflict episode is played out. It could be that differences in characteristics between individuals are the very reason for a conflict episode, for example, when two team members get into an argument because one is task oriented and the other is more relationship oriented. This difference can be the root of a series of events that result in conflict. Similarly, these differences may lead to each party handling the episode in a different way. Thus, individuals have stable or slow-changing characteristics that can influence conflict and the conflict process.

One characteristic that affects conflict is an individual's culture (i.e., societal norms). Conflict is seen by some cultures as beneficial but is seen as detrimental by others. Thus, more assertive strategies are likely to be preferred by individuals from cultures seeing conflict as beneficial. Other individual characteristics affecting conflict are goal characteristics, stress, and the need for autonomy. An individual with high aspirations or extremely rigid goals is more likely to engage in conflict due to increased awareness of possible impediments to his or her goals. In addition, an individual whose goals are interdependent with another person's goals is more likely to engage in conflict because the latter person's behavior has an impact on the former person's ability to achieve a goal. Stress may also lead to greater conflict due to the tension that an individual experiences, much like a rubber band pulled tight and ready to snap. Finally, if an individual has a high need for autonomy and another party makes demands that infringe on that autonomy, there is an increased likelihood of conflict.

Another individual characteristic that affects conflict is personality. For example, "Type A" and "Type B" personalities are likely to manage conflict in different ways given that Type B individuals are less competitive, less temperamental, and more patient. Studies by Antonioni in 1998 and Moberg in 2001 examined the effects of the "Big Five" personality factors—Extraversion, Openness, Agreeableness, Conscientiousness, and Neuroticism—on preference of conflict management strategy. For instance, avoiding seems to be preferred by individuals with low Extraversion and low Conscientiousness, collaborating is preferred by those with high Extraversion and high Conscientiousness, and competing is associated with low Agreeableness.

Self-esteem has received considerable attention in conflict research. Research has shown that individuals

with high self-esteem value satisfying both their own concerns and their counterparts' concerns. These individuals are also more confident that collaboration is possible and are less likely to become defensive, a condition that negatively affects collaboration. Thus, individuals with high self-esteem are more likely to use collaboration in managing conflict. Other individual factors affecting conflict include locus of control (internal may provide greater confidence for resolving conflict), dogmatism (high levels may limit the flexibility of resolution possibilities), and stage of moral development (higher stage individuals are more likely to use collaboration). The individual characteristics presented here do not represent a comprehensive list of individuals' structural variables that may affect conflict. However, these are some that have been addressed in research. Primarily, the discussion presented here suggests that it is important to consider more than the issue when trying to understand a conflict episode. The individuals themselves may have characteristics that contribute to both the triggering of the conflict episode and its management.

### 4.2. Organizational Characteristics

Conflict does not occur in a vacuum; rather, it occurs in some sort of interactive setting such as the workplace. Much of the research presented here (and in general) has concentrated on conflict occurring in an organizational context. Thus, much research has focused on examining the characteristics of organizations that influence conflict and its management.

Task characteristics may affect how a party manages conflict. When performing competitive tasks, an individual may be more likely to either compete or avoid. However, when performing a cooperative task within a group setting, collaborating and compromising are more likely to be used to manage within-group conflict. In addition, if the individual is performing a task that is highly interdependent with the other party's task, he or she is more likely to prefer a collaboration strategy.

In addition to these, organizational culture and norms are likely to affect both parties' behaviors during conflict. For instance, some conflict management strategies may be taboo within an organization. Rules and procedures are likely to affect conflict management as well because these can constrain certain behaviors while requiring other behaviors.

Structural conditions are considered fixed in the short run during the conflict process. However, in the long

run, structural conditions may be manipulated to make the conflict process more effective for the parties and the organization of which they are members. Some structural conditions can be affected by individual or organizational intervention. That is, parties can make adjustments to these conditions to influence how conflict is managed when it occurs. For instance, if increased stress results in undesirable behaviors and outcomes of a conflict episode, an organization may benefit by fostering an environment of low stress for its members. Intervening in this way can improve the way in which conflict is managed throughout the organization. Understanding individual and organizational characteristics that influence conflict can help managers and organizations to deal with conflict more effectively.

## 5. CONFLICT OUTCOMES

During a conflict episode, the primary parties and any third parties are continually working toward some sort of resolution. The ideal or acceptable outcomes of a conflict episode are likely to be related to the type of conflict management strategy employed by a primary party or to the type of intervention used by a third party. However, a conflict episode may have outcomes that go beyond what is expected by the parties involved. Conflict has several distinct types of outcomes: task, performance, and structural. Task outcomes include the resolution of the issue and the direct repercussions of the specific episode. Conflict can also have an effect on individual or group performance, for better or for worse, in the case where the conflict occurs in an organizational context; thus, performance is related to the direct repercussions portion of task outcomes. Finally, the effects of the episode on structural characteristics are a third distinct outcome. Thus, the three categories of outcomes that are discussed here are issue resolution, performance outcomes, and structural characteristics.

Issue resolution outcomes are related to the types of strategies employed to manage conflict. Much of the negotiation literature suggests that an agreement can result in either a distributive outcome or an integrative outcome. In addition, a third possibility is an impasse, that is, no agreement at all. A distributive outcome is one in which there is a fixed resolution that is either taken wholly by one party or somehow divided between the parties. An integrative outcome, on the other hand, is one in which the parties create an arrangement in which both parties are able to fully or significantly

achieve their goals in resolving the conflict. This seems to be an ideal outcome, but in some situations integrative outcomes are not possible. In such cases, impasses and distributive outcomes are the only possibilities.

Jehn has contributed greatly to the research on performance outcomes of interpersonal conflict in organizations. Increased or decreased performance is a conflict outcome of which the primary parties might be unaware. That is, during conflict or as a result of conflict, each party's performance in an organization may either suffer or benefit. Two types of conflict with differing effects on performance in organizations are task conflict and relational conflict. Task conflict is concerned with the content and goals of the work itself such as a disagreement between managers on which market to enter next. Relational conflict focuses on the interpersonal relationship between the parties such as one party having a personal vendetta against the other party.

Moderate to high levels of task conflict appear to result in better performance in an organization. This is likely due to the fact that when bad or less effective ideas are opposed, countering with good or more effective ideas is likely to have a positive effect on performance. On the other hand, relational conflict has a negative effect on an individual's performance, probably due to the distractive effects of interpersonal feuding. In addition, each type of conflict is unique, with one type of conflict rarely transforming into the other (e.g., task conflict becoming relational conflict). This suggests two important points. First, an individual's work performance is affected by conflict. Second, whether an individual's work performance is affected positively or negatively by conflict depends on whether the conflict faced is task conflict or relational conflict.

When a conflict episode ends, the effects of the episode can affect future conflict episodes by influencing or changing characteristics of the parties or the organization. For instance, in 2000, Frone found that conflict with a supervisor can cause a worker to be less committed to an organization and can lead to lower job satisfaction and increase a worker's intention to quit. Similarly, a particularly influential conflict episode can lead to instituting or changing organizational policies on how conflict is handled such as creating an "open door" policy and requiring managers to oversee conflicts between coworkers. Thus, one unintended outcome of a conflict episode might be changes in individual or organizational characteristics that may linger and have important effects on future conflict episodes involving those parties or within that organization.

This suggests that conflict outcomes can be intended or unintended, direct or indirect, and beneficial or detrimental. It is likely that some of the outcomes of a conflict episode will be unintended such as performance outcomes and structural changes. The intended outcome, usually the issue resolution, may be the primary or only focus of a party engaged in a conflict episode. However, it is important for the two parties to recognize that conflict can lead to unintended detrimental outcomes that can outlast the intended beneficial outcomes they are trying to achieve. It is even more important for an organization to recognize that internal interpersonal conflict can result in detrimental outcomes for the firm, requiring third parties to intervene on behalf of the organization's interests and for the benefit of the two parties.

## 6. THIRD-PARTY INTERVENTION

In some situations, the two parties involved in conflict need assistance in reaching a resolution. For instance, if each party refuses to make any concession and neither party can exert influence over the other party, the result is an impasse. If a resolution is necessary, a third party may get involved to assist in resolution. In other instances, resolution of the conflict may have an effect on organizational interests and a third party may get involved to look out for those interests. Because third parties are often involved in managing conflict, it is important to understand the ways in which third parties can help (or hinder) the management of conflict. It is also helpful to understand what factors affect how third parties will intervene and the effects of such intervention on the organization and the primary parties.

### 6.1. Methods of Intervention

Two aspects of a conflict situation are relevant to third-party intervention: the conflict process and the resolution or outcome of the conflict episode. Process intervention is primarily concerned with managing the manner in which the primary parties interact. Outcome intervention is a matter of whether the third party has any control over the final decision regarding resolution. In 1985, Lewicki and Sheppard presented a set of third-party intervention methods, each of which exerts high or low process or outcome control. These intervention methods can be seen in Fig. 2.

Using high outcome and process control is referred to as inquisitorial intervention and has been shown

Outcome control	High	Inquisitorial intervention	Adversarial intervention
	Low	Mediation	Providing impetus
		High	Low
		Process control	

**FIGURE 2** Third-party intervention methods. Reprinted from Lewicki, R. J., & Sheppard, B. H. (1985). Choosing how to intervene: Factors affecting the use of process and outcome control in third party dispute resolution. *Journal of Occupational Behavior*, 6, 49–64. Copyright John Wiley & Sons Limited. Reproduced with permission.

empirically to be the most common form of intervention employed by managers in dealing with conflict between subordinates. Using this type of intervention, a third party is likely to direct the interaction between the primary parties, deciding what information should be shared, deciding what is relevant or irrelevant, and possibly instructing parties as to how they should behave during the process. The third party is then likely to decide on a good resolution independently and to enforce this on the primary parties.

High outcome and low process control by a third party is referred to as adversarial intervention and is also a common method of intervention used by managers. This method is compared to the American judicial system in which parties present arguments and evidence with little or no direction from the third party. After each party has made a case, the third party makes the decision of how the conflict will be resolved and may even enforce the decision.

A less popular yet still used method of third-party intervention is the low outcome and low process control technique providing impetus. When using this method, a manager may listen to the primary parties to get a basic understanding of what the conflict is about and then suggest that the parties figure out a resolution themselves quickly or else some sort of punishment will be given or enforced. This is similar to a parent telling two children that they had better work out their differences or else they will be punished. A style with similar control as this method, yet more constructive and less threatening, is a method in

which the third party listens to disputants' views, incorporates their ideas, and asks for resolution proposals from the primary parties.

The low outcome and high process control method, or mediation, has been normatively described as the ideal means of third-party intervention but is rarely, if ever, used by managers in handling conflict between subordinates. It is suspected that this type of intervention is not used due to the large time investment required to direct the conflict process yet allow the primary parties to decide on a mutually beneficial resolution. Managers may prefer the other three methods of intervention so as to bring the conflict to a swift conclusion.

A third party is more likely to exert greater outcome control when time pressure is high, when the two parties are not expected to interact frequently in the future, and when the conflict outcome will have an impact on other parties and/or the organization. Greater process control is more likely to be used when the parties are less likely to interact in the future. Thus, it is suggested that managers are likely to use inquisitorial intervention when there are disputes between subordinates who do not interact very often. In addition, managers as third parties are more likely to use more autocratic (high-control) methods than are peers as third parties.

## 6.2. Outcomes of Intervention

How does one determine to what degree third-party intervention has succeeded? The effectiveness of a third party's involvement can be interpreted according to the achievement and quality of a resolution or according to the primary parties' perceived fairness of the process and outcome. Much of the research on the effects of third-party intervention focuses on the latter. Because third parties can vary in the amount of control exerted over the conflict process and outcome, the primary parties can have different perceptions about the fairness of the process and the outcome.

Procedural justice refers to the degree to which a party perceives that the conflict process was handled fairly by a third party. A party is more likely to perceive the process as fair if a compromise is reached or if the outcome is otherwise favorable to that party. In addition, a party is likely to perceive greater procedural justice when a third party uses mediation, which is a more facilitating role. Providing impetus and inquisitorial intervention may lead to negative perceptions of fairness in the conflict process because these roles are more autocratic.

Distributive justice refers to the degree to which a party perceives the outcome of an episode as fair. In a similar reflection of procedural justice, a party is more likely to perceive the outcome as fair if a compromise is reached or if the outcome is favorable to that party. Therefore, it is no surprise that a party is more likely to perceive an outcome as unfair when the outcome favors the other party. Unlike procedural justice, however, the method of intervention does not appear to have an impact on perceived fairness of the outcome. Research has shown that a party is more likely to perceive the outcome as fair if the process is perceived as fair.

Another possible outcome of a conflict episode is the lack of reaching a resolution, that is, an impasse. The use of autocratic intervention by third parties often increases the chances of impasses or one-sided resolutions, whereas facilitating methods of intervention often increase the chances of compromise resolutions. In addition, an impasse is more likely if a peer, rather than a manager, takes on an autocratic third-party role.

## 7. CONCLUSION

Interpersonal conflict is an unavoidable aspect of organizational life and can have a substantial impact on individuals and organizations. As noted in this article, conflict can affect task outcomes, performance, and organizational structure. Thus, learning to manage conflict is a worthwhile investment for both individuals and organizations.

This article summarized information that may help to develop a general understanding of interpersonal conflict in organizations. The conflict process was first described in an effort to elucidate the sequence of a conflict episode. Five conflict management strategies were discussed along with suggestions as to when each might be more appropriate. Individual and organizational characteristics that may influence both the development and management of a conflict episode were presented. The effects or outcomes of conflict were noted, as were the possibilities for third-party intervention. Because interpersonal conflict is inevitable, the challenge for individuals in organizations is to use this information to manage conflict more effectively such that positive outcomes are obtained.

### See Also the Following Articles

Industrial/Organizational Psychology, Overview ■ Organizational Culture and Climate

### Further Reading

- Antonioni, D. (1998). Relationship between the Big Five personality factors and conflict management styles. *International Journal of Conflict Management*, 9, 336–355.
- Blum, M. W., & Wall, J. A., Jr. (1997, May–June). HRM: Managing conflicts in the firm. *Business Horizons*, pp. 84–87.
- Brockmann, E. (1996). Removing the paradox of conflict from group decisions. *Academy of Management Executive*, 10, 61–62.
- Frone, M. R. (2000). Interpersonal conflict at work and psychological outcomes: Testing a model among young workers. *Journal of Occupational Health Psychology*, 5, 246–255.
- Jehn, K. A. (1997). A qualitative analysis of conflict types and dimensions in organizational groups. *Administrative Science Quarterly*, 42, 530–557.
- Karambayya, R., & Brett, J. M. (1989). Managers handling disputes: Third-party roles and perceptions of fairness. *Academy of Management Journal*, 32, 687–704.
- Karambayya, R., Brett, J. M., & Lytle, A. (1992). Effects of formal authority and experience on third-party roles, outcomes, and perceptions of fairness. *Academy of Management Journal*, 35, 426–438.
- Lewicki, R. J., & Sheppard, B. H. (1985). Choosing how to intervene: Factors affecting the use of process and outcome control in third party dispute resolution. *Journal of Occupational Behavior*, 6, 49–64.
- Lynch, D. (1997, May). Unresolved conflicts affect bottom line. *HRMagazine*, pp. 49–50.
- Moberg, P. J. (2001). Linking conflict strategy to the Five-Factor Model: Theoretical and empirical foundations. *International Journal of Conflict Management*, 12, 47–68.
- Mushref, M. A. (2002, November). Managing conflict in a changing environment. *Management Services: Journal of the Institute of Management Services*, pp. 8–11.
- Oetzel, J. G. (1999). The influence of situational features on perceived conflict styles and self-construals in work groups. *International Journal of Intercultural Relations*, 23, 679–695.
- Pondy, L. R. (1967). Organizational conflict: Concepts and models. *Administrative Science Quarterly*, 12, 296–320.
- Pruitt, D. G. (2001). Conflict and conflict resolution, social psychology of. In N. J. Smelser and P. B. Baltes (Eds.) *International encyclopedia of the social and behavioral sciences*. San Diego: Elsevier Science.
- Rahim, M. A., Buntzman, G. F., & White, D. (1999). An empirical study of the stages of moral development and conflict management styles. *International Journal of Conflict Management*, 10, 154–171.
- Thomas, K. W. (1992). Conflict and negotiation processes in organizations. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3, pp. 651–717). Palo Alto, CA: Consulting Psychologists Press.
- Van Slyke, E. J. (1999, November). Resolve conflict, boost creativity. *HRMagazine*, pp. 132–137.
- Wall, J. A., Jr., & Callister, R. R. (1995). Conflict and its management. *Journal of Management*, 21, 515–558.



# Conformity across Cultures

Rod Bond

University of Sussex, Brighton, United Kingdom

1. Comparisons Across Cultures
2. Comparisons within a Culture at Different Periods in its History
3. Conformity or Harmony?
4. Conclusion  
Further Reading

## GLOSSARY

**collectivist cultures** Cultures in which the emphasis is on interdependence and the self is defined in relation to group membership.

**cultural group** A group sharing a common culture.

**developed society** A society with developed systems of political authority and use of technology in economic production.

**individualist cultures** Cultures in which individuals are independent of one another and social behavior is largely determined by personal goals rather than collective goals.

**pastoral society** A society in which grazing sheep and/or raising livestock form the primary mode of subsistence.

**subsistence economy** An economy that meets little more than the basic necessities and is typically characterized by low levels of production used for consumption rather than exchange.

Theories of conformity have long stressed the importance of cultural values in shaping people's responses to group pressure. Asch, for example, believed that his classic demonstrations of conformity—of people's willingness

to "call White Black"—raised questions about the values fostered by American society and about its educational practices. Others have argued that the early demonstrations of conformity in the United States, especially those of Asch, reflect the particular climate of the McCarthy era and would not be found in the more tolerant culture of post-cold war society. In line with these arguments, evidence for the cultural roots of conformity has come from two types of comparisons: those across cultures and those within a culture at different periods in its history.

## 1. COMPARISONS ACROSS CULTURES

Studies that have compared conformity across cultural groups can be divided into three broad types: comparisons of subsistence economies, comparisons of developed societies, and comparisons of different cultural groups within a society.

### 1.1. Comparisons of Subsistence Economies

Berry has argued that different modes of subsistence require different human qualities that are reflected in cultural values and socialization practices and that give rise to differences in conformity. It has been found that pastoral and agricultural peoples emphasize obedience and responsibility in their socialization practices because these economies need people who are

conscientious and compliant, whereas hunting and fishing economies emphasize independence, self-reliance, and individual achievement because these economies need people who are individualist and assertive. Berry argued that if this is the case, pastoral and agricultural peoples should also be more likely to conform than are hunting and fishing peoples. Consistent with this reasoning, he found greater conformity on a line judgment task among the Temne of Sierra Leone, an agricultural society with strict disciplinarian socialization practices, than among the Eskimo of Baffin Island, a hunting and fishing society whose socialization practices are lenient and encourage individualism. However, his later research in Australia and New Guinea, as well as among North American Indians, obtained weaker support for this theory. He also found that where a subsistence economy had been exposed to Western values, there was less cross-cultural variation in conformity.

### **1.2. Comparisons of Developed Societies**

Comparisons of developed societies have yielded a mixed set of findings. Several studies have reported cross-cultural differences in conformity that are explained by the relative values attached to conformity in the societies concerned. One of the earliest comparisons by Milgram found greater conformity on a perceptual task among Norwegian students than among French students, and this was taken as reflecting the fact that Norway is less tolerant of deviance than is France. Some investigators have replicated the Asch experiment, in which participants judge which of three lines is the same length as another line after hearing the unanimous, but manifestly erroneous, judgment of a group of other participants who are in fact confederates of the experimenter. Replications in Brazil, Zimbabwe, Ghana, and Fiji, as well as among the Chinese, found high levels of conformity when compared with the original results, and these findings were interpreted as reflecting the higher value attached to conformity in these societies.

Anticipated differences have not always emerged, however, and further studies of conformity from countries as diverse as Brazil, Lebanon, Zaire, France, Kuwait, Portugal, and Hong Kong all have reported comparable results, leading some authors to remark on the cross-cultural stability of the classic demonstrations of conformity. Studies in Japan, a country that at

first sight might be expected to have higher levels of conformity, have also been inconclusive and point to the importance of taking into account the relationship between the individual and the group. When the majority are strangers to the individual, there is some evidence for a lower level of conformity (relative to the United States) among Japanese students, but other evidence suggests a higher level of conformity when the majority are friends of the individual.

### **1.3. Comparisons of Cultural Groups within a Society**

Some investigators have compared cultural groups within a society as a way in which to study the cultural roots of conformity, although this has not proved to be a fruitful line of inquiry. There is some indication that conformity is higher among minority groups than among majority groups. For example, some studies have found higher conformity among Blacks than among Whites in the United States, and another study found higher conformity among Puerto Ricans than among Whites. A study in Britain found higher conformity among West Indians than among Whites. A study in Fiji found greater conformity among Fijian Indians than among native Fijians. However, there is no consistent pattern of results, nor is it clear why such differences have been found.

### **1.4. Conformity and Individualism–Collectivism**

Cross-cultural comparisons have yielded a mixed set of findings. This may be due, in part, to the methodological limitations of many of the studies in that frequently poor control is exercised over differences in sample and method between the cultures that are compared. Also, comparisons typically are made with single samples from each culture, and evidence from other studies carried out in a culture is disregarded. The explanation for cross-cultural differences, when they are found, is usually post hoc, and there has been little attempt to directly assess the values presumed to mediate responses to group influence. To address these concerns, Bond and Smith in 1996 conducted a meta-analysis of 133 Asch-type conformity studies from 17 countries that was able to provide tighter control over extraneous factors and combine all available evidence from various cultures. Moreover, they proposed that cross-cultural differences in conformity would depend on the extent to which a culture subscribed to

collectivist values as opposed to individualist values. Individualism–collectivism has emerged as the most important dimension on which cross-cultural differences in values have been described, albeit one that continues to be a central topic of debate. In individualist cultures such as European Americans, individuals are independent of one another and social behavior is largely determined by personal goals rather than collective goals. Where personal goals conflict with collective goals, it is personal goals that tend to be prioritized. In contrast, in collectivist cultures such as East Asian societies, the emphasis is on interdependence and the self is defined in relation to group membership. Social behavior is largely determined by collective rather than personal goals, and collective goals are given priority whenever they are in conflict with personal goals. Conformity to the majority would be more likely in collectivist cultures than in individualist cultures given the higher value placed on maintaining harmony in group relations. This was what Bond and Smith found. Using measures of collectivist values derived from three independent cross-national value surveys, they found that studies conducted in collectivist countries tended to report larger conformity effects than did those conducted in individualist countries.

## **2. COMPARISONS WITHIN A CULTURE AT DIFFERENT PERIODS IN ITS HISTORY**

A second line of evidence for the cultural roots of conformity comes from the observation that replications of conformity experiments within a society produce different results at different points in time. Larsen conducted three replications of the Asch experiment and interpreted the different levels of conformity as reflecting sociopolitical changes in American society. The more questioning attitude of students of the Vietnam War era explained the lower level of conformity found in 1974, whereas the decline of student activism explained the rise of conformity levels by 1979. In 1988, conformity had declined again somewhat, and this was seen as reflecting the increase in protest activities. Similarly, Perrin and Spencer were unable to replicate the Asch experiment with British students in 1980, and they attributed this to their more questioning attitude compared with Asch's respondents during the 1950s. They concluded that the Asch experiment was not an enduring social

psychological phenomenon but rather one rooted in its historical context. A few years later, another study found some conformity among British students that was attributed to the greater cohesiveness brought about by the Falklands war.

Not all investigators have found it difficult to replicate the Asch experiment, however, and several other recent studies of university students conducted in Britain, Belgium, Holland, and Portugal all have found reasonably high levels of conformity comparable to those found in Asch's research. Another line of argument, proposed by Lamb and Alsifaki, is that levels of conformity have been steadily on the increase, drawing on the hypothesis that modern industrial societies are characterized by increasing numbers of "other-directed" types who are more easily influenced by peer pressure.

This line of inquiry is highly speculative given that there are many potential reasons, apart from changes in the sociopolitical context, why two studies conducted at different points in time would give different results. Nevertheless, Bond and Smith's meta-analysis of Asch-type conformity studies enabled a synthesis of a number of studies and controlled for differences in study design and method. Using only those studies carried out in the United States, of which there were 97 in total, Bond and Smith found that the level of conformity was related to the time when the study was carried out. Studies carried out more recently have tended to report lower levels of conformity than did earlier studies, perhaps suggesting that there has been a greater value placed on independence in U.S. society during the post-World War II era.

## **3. CONFORMITY OR HARMONY?**

The social psychological literature, by and large, portrays conformity in negative terms. The individual is pitted against the group, and to yield to group pressure is an act of submission and a sign of weakness. Individuals are expected to express their own opinions, and those who continue to do so in spite of pressures to do otherwise are admired. Such a view of conformity is rooted in Western individualist values in which self-expression and independence from the group are highly valued. Viewed from the standpoint of collectivist values, however, such behavior may seem selfish and immature. In such societies, individuals should seek to maintain harmony with the group, and open disagreement can be a source of embarrassment. Thus, the same behavior might be viewed as conformity in an

individualist culture (and be judged negatively) but be viewed as maintaining harmony in a collectivist culture (and be judged positively).

In 1999, Kim and Markus demonstrated the different values placed on conformity and uniqueness in collectivist and individualist cultures. For example, they found that magazine advertisements in Korea were more likely to have conformity themes, emphasizing respect for collective values and beliefs as well as harmony with group norms, than were those in the United States. In contrast, advertisements in the United States were more likely to have uniqueness themes consisting of freedom, choice, individual uniqueness, and rebelling against collective values and beliefs. These authors also studied preferences for unique or common elements of abstract geometrical figures and found that European Americans preferred unique elements, whereas East Asians preferred common elements. Thus, conformity takes on different meanings depending on the network of values and beliefs that comprise the cultural context within which the behavior occurs, and these differences in meaning in turn will entail that conformity or independence will carry different affective consequences.

#### 4. CONCLUSION

The literature on cross-cultural differences in conformity has provided a mixed set of results. This is due, in part, to the difficulty in ensuring comparable methods and samples when conducting cross-cultural research, and conformity studies are in any case difficult to carry out because it is so easy to arouse suspicion among participants. Most studies have used the classic conformity paradigms of Asch or Crutchfield and have used mostly perceptual judgment tasks. The major limitation has been the lack of attention to underlying process or theoretical elaboration of relevant features of the situation. Recent work relating individualism–collectivism to conformity goes some way in redressing

this issue and suggests that people in collectivist cultures will conform more. There is also some evidence that conformity in U.S. society has declined since World War II. However, no studies have yet directly assessed how collectivist values relate to conformity. Moreover, the nature of collectivism suggests that the relation of the individual to the group is crucial and that it is not generally true that people in collectivist societies will conform more. In collectivist cultures, individuals will seek to maintain harmony with in-group members but may be less concerned about maintaining harmony with out-group members. Finally, care must be taken in interpreting behavior within its cultural context. Conformity may reflect a different capacity to resist group pressure or a different determination to avoid embarrassment. Similar behavior may take on different meanings and reflect different processes in different cultural contexts.

#### See Also the Following Articles

Acculturation ■ Cross-Cultural Psychology, Overview  
■ Cultural Complexity

#### Further Reading

- Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, 119, 111–137.
- Furnham, A. (1984). Studies of cross-cultural conformity: A brief and critical review. *Psychologia*, 27, 65–72.
- Kim, H., & Markus, H. (1999). Deviance or uniqueness, harmony or conformity? A cultural analysis. *Journal of Personality and Social Psychology*, 77, 785–800.
- Mann, L. (1988). Culture and conformity. In M. H. Bond (Ed.), *The cross-cultural challenge to social psychology* (pp. 184–187). Newbury Park, CA: Sage.
- Smith, P. B., & Bond, M. H. (1998). *Social psychology across cultures: Analysis and perspective* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.



# Connectionism

Yoshihisa Kashima

University of Melbourne, Melbourne, Australia

1. Introduction
2. General Features of the Connectionist Cognitive Architecture
3. Localist and Distributed Representational Systems and Their Properties
4. Dynamics of an Information Processing Unit
5. Dynamics of Interunit Connections
6. Examples of Connectionist Architectures
7. Applications of Connectionism  
Further Reading

## GLOSSARY

**delta rule (backpropagation rule)** A well-known algorithm used in a connectionist network for adjusting the strength of links among artificial neurons as a function of the amount of errors generated when the network predicts outputs based on its inputs.

**distributed approach to meaning** An approach where a meaningful concept is represented in a connectionist network by a pattern of activation over a number of nodes or artificial neurons.

**“exclusive or” operation** A logical operation where either one or the other alternative is selected and where both alternatives may not be selected.

**feedforward network** A type of connectionist network architecture in which input signals are transmitted forward, but not backward, to output nodes.

**generalized delta rule** An algorithm for adjusting the strength of links among artificial neurons so that a connectionist network can learn input–output relations.

**Hebbian learning rule** An algorithm for adjusting the strength of connections among artificial neurons where the amount of change in connection strength is a function of the amount of activation of the connected neurons; it is named after Donald Hebb, who first proposed the idea for the learning involving real neurons.

**localist approach to meaning** An approach where a meaningful concept is represented in a connectionist network by a corresponding node or artificial neuron.

**multilayer feedforward network** A type of connectionist network architecture in which several clusters (or layers) of artificial neurons are connected with each other and input signals are transmitted only forward through multiple layers of artificial neurons to an output layer.

**recurrent network** A type of connectionist network architecture in which input signals are transmitted forward from input nodes to output nodes but also backward from output nodes to input nodes.

**sigmoid function** A mathematical function that links input amount to output amount (typically of an artificial neuron) where output increases slowly when the amount of input is small or large but increases rapidly when the input amount is medium.

**supervised learning** A type of learning that occurs in connectionist networks where outputs that are to be associated with inputs are provided as standards or “teachers” from which to learn.

**synaptic link** A link between neurons in a nervous system through which neural signals are transmitted.

**tensor network** A type of connectionist network architecture in which several clusters of artificial neurons are connected with each other.

**threshold function** A mathematical function that links input amount to output amount (typically of an artificial neuron)

where an output occurs when an input reaches a certain level or a threshold.

**unsupervised learning** A type of learning that occurs in connectionist networks where outputs that are to be associated with inputs do not act as standards or “teachers.”

Connectionism is a way of conceptualizing psychological processes that gets its inspiration from the way in which the biological brain processes information. In this approach, psychological phenomena are explained in terms of activations of interconnected neuron-like units whose operations are governed by simple rules.

## 1. INTRODUCTION

Although ideas akin to connectionism may be attributed to past thinkers such as Aristotle and William James, modern connectionism began its tenure during the mid-20th century. Alongside the theoretical and actuarial development of digital computing based on von Neumann’s concept of serial computing (most current computers are of this type), there emerged a research tradition that took seriously the biological brain as a metaphor for information processing. In this view, just as the brain does its work through neurons and synaptic links, the mind too may be conceptualized as functioning through neuron-like simple units and connections among them. Connectionism, also known as neural networks, is so called due to the network of connections among the neuron-like units that are postulated to explain psychological phenomena.

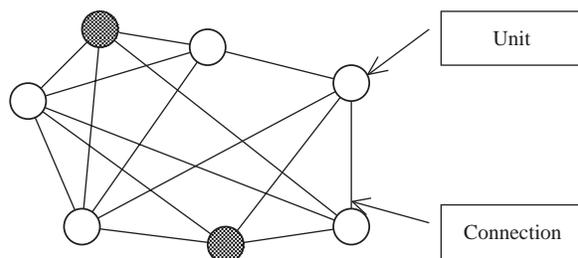
McCulloch and Pitts, Hebb, and Rosenblatt conducted their pioneering work during the 1940s and 1950s. Nevertheless, Minsky and Papert pointed out in 1969 that some of the simple logical operations such as “exclusive or” cannot be performed by the then available mechanisms. These limitations, together with the development of serial computers as a main technological tool of computation, resulted in a decline in the research on brain-inspired information processing mechanisms. While Anderson, Fukushima, Grossberg, Kohonen, and others continued to work during the 1960s and 1970s, it was the 1980s that witnessed the renaissance of connectionism. A mechanism called the generalized delta rule, which rectified connectionism’s problem identified by Minsky and Papert, was developed. A network of simple neuron-like units was shown to be able to model psychological processes such as perception, memory, and learning. Rumelhart and colleagues’

*Parallel Distributed Processing*, published in 1986, popularized the connectionist approach.

By this time, however, the approach that used the von Neumann serial computer as a metaphor of mind (a psychological process is considered to be a series of discrete steps of mental operations on some mental representations) had established itself as the dominant research program in psychology and cognitive science. Controversies ensued as to how connectionism differs from the serial processing approach, which approach explains psychological processes better, and whether connectionism provides a realistic model of brain processes.

## 2. GENERAL FEATURES OF THE CONNECTIONIST COGNITIVE ARCHITECTURE

Although connectionist architectures vary considerably, they can be characterized in general terms. First, a connectionist network typically consists of a large number of units (sometimes called nodes) that are connected to each other with varying strengths (Fig. 1). The operation of each unit is governed by a simple rule (discussed later) that determines its activation state (“firing” or “resting”) as a function of the inputs it receives from the connected units. The connection strength (often called weight) of a given unit to another unit indicates the amount of influence the first unit has over the second unit. Second, the pattern of activation of the units changes over time. The rules used in connectionist networks specify how the activation of a unit at one point in time ( $t$ ) affects other units at a later time ( $t + \Delta t$ ). Interunit effects occur in parallel, so that during a given time interval ( $\Delta t$ ), the effects of all units on other units (a unit may have an effect on itself) take place. As



**FIGURE 1** A hypothetical connectionist network. The shaded unit is activated.

a result, the state of a connectionist network can continue to change, potentially forever. However, under some circumstances, the state of a connectionist network shows a discernible temporal pattern. For instance, it may come to a stable pattern, oscillate between two or more patterns, or exhibit a chaotic pattern. Third, a connectionist network performs its function collectively. If one unit is activated, its activation spreads via its connections to other units in accordance with a rule. As these activations reverberate through the network, the network as a whole may then perform various functions such as the encoding and retrieval of information.

Connectionism differs from models of cognition based on the serial computer metaphor in several respects. A serial computer typically has a central processing unit (CPU), data, and procedures (i.e., software) that operate on the data, and the CPU executes the procedures to handle the data. One fundamental assumption of the serial computer metaphor of cognitive processes is that both data and procedures are written in terms of physical symbols. Physical symbols (e.g., the word “cat”) refer to or stand for some objects and events in the world (e.g., the cat on the mat) and are themselves physically coded on a physical medium such as an ink mark on paper, an electronic circuit, or a synaptic circuit in the brain. In contrast, connectionism does not assume the existence of a single CPU; rather, it assumes the existence of a multitude of simple processing units. What the CPU does in a serial computer is governed by a simple or complex set of procedures that can be programmed (e.g., a simple calculator, complex word processing software), but processing units in connectionist networks have a simple and fixed rule. In other words, connectionist networks are not programmable in the same way in which a serial computer can be programmed—although they can be trained or taught (discussed later). Finally, connectionism does not assume the explicit coding of information in physical symbols. Despite these differences, both metaphors have been fruitfully used to advance psychological science. One metaphor appears to be more suited to model some psychological processes than to model others, but both metaphors often do a similarly good job of explaining a set of phenomena. At this point in history, they may be best considered to be different research programs with different theoretical assumptions and conventions rather than two competing paradigms.

Although neurally inspired, connectionist networks are not models of brain functions. Put differently, the relationship between the connectionist network and the

brain is metaphorical rather than literal. To be sure, connectionist modelers often make use of design principles that are observable in the brain. For instance, just as the firing of a neuron is not a simple linear function of the excitation it receives, the activation of a unit in a neural network is often governed by a nonlinear function of the inputs it receives. Just as synaptic connections may be excitatory or inhibitory, connections among connectionist units may be excitatory or inhibitory. However, a learning rule such as the generalized delta rule, which plays an important role in neural network research, has no known corresponding mechanism in the brain. More generally, the brain anatomy is highly structured, but most neural networks do not have differentiated architectures that resemble the brain structure. In the long run, neural network research may become more informed by neuroscience (and, in fact, there is research that tries to model brain functions). Currently, however, connectionist networks are at best highly abstracted and stylized models of how the brain works. For this reason, connectionist networks are often called artificial neural networks.

### 3. LOCALIST AND DISTRIBUTED REPRESENTATIONAL SYSTEMS AND THEIR PROPERTIES

How do we understand objects and events in the world, and how do we experience our world as meaningful? Meaning is one of the most important questions for models of cognition. In cognitive models based on the serial computer metaphor, symbols gain their referential meanings by designation; a symbol is designated to refer to a certain object or event in the world. In this sense, meaning is built into the system of cognition. In neural networks, meaning is handled in two different ways. The localist approach is akin to the serial models of cognition. Here, a unit in a connectionist network is designated to have its referent. When this unit is activated, the person is considered to be in some sense “thinking about” the referent that the unit represents. In contrast, in the distributed approach, any given unit might not have a clear referent, and its activation cannot be interpreted as corresponding to any meaningful thought in and of itself. Rather, it is a pattern of activations of many units that imply a meaning. In other words, meaning is not localized to a unit but rather distributed across units. A unit in a distributed system then represents some microfeature of a meaningful

concept that may or may not be interpretable using symbols such as words and phrases. In this sense, a distributed approach models subsymbolic processes.

An advantage of the localist approach is its ease of interpretation. The activation of a unit means that the concept represented by the unit is being thought about; changes in activation of a network can be easily interpreted as changes in the content of thought. Although the localist approach may be sufficient to model many psychological processes, the distributed approach might be necessary when some of the following features are important. First, the distributed representational system has a built-in capacity to model similarity. Because a concept is represented by a pattern of unit activation in the distributed system, the similarity between one pattern and another pattern of activation can be naturally interpretable as a similarity in meaning. Second, the capacity to represent similarity enables the distributed system to have the capacity for psychological generalization as an inherent aspect of the system. If a system responds to a pattern in a certain way, it responds to a similar pattern in a similar way; that is, the system generalizes as a natural consequence of its distributed format. Third, these capacities often result in a graceful degradation if the system is damaged. For instance, if one unit in a localist network is damaged or removed, it loses the capacity to represent the corresponding concept. However, even if one unit in a distributed network is lost, it still retains most of the capacity to represent concepts even though the clarity of representation might be compromised to some extent.

#### 4. DYNAMICS OF AN INFORMATION PROCESSING UNIT

In a time interval ( $\Delta t$ ), an information processing unit in a neural network receives inputs from other units, computes its net inputs, gets activated according to an activation function, and sends out its output to other units according to an output function. Figure 2 shows a schematic picture in which Unit  $i$  receives inputs from its neighbors, Unit  $j$  and other connected units, and sends out its output. When a neighboring Unit  $j$  sends out its output,  $o_j$ , it spreads via the connection with its connection strength or weight,  $w_{ij}$ , to Unit  $i$ . The input from  $j$  to  $i$  then is the product of the output and weight, that is,  $w_{ij}o_j$ . The inputs from other units are also computed in a similar fashion. The Unit  $i$  then sums

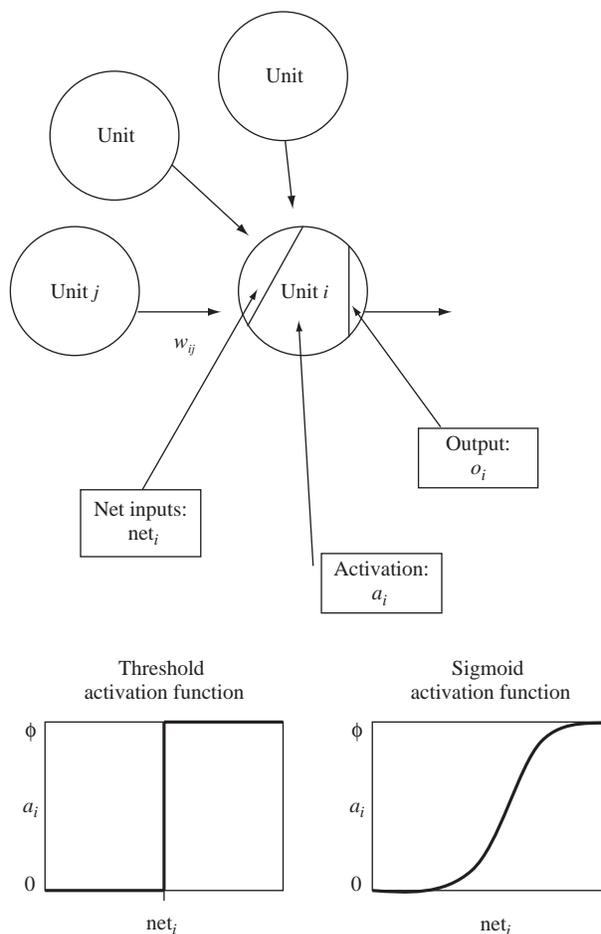


FIGURE 2 Information processing unit and activation functions.

all inputs to compute its net input, that is,  $net_i = \sum w_{ij}o_j$ , where the summation is over  $j$ , which indexes units that send out their outputs to Unit  $i$ .

Whether, or how much, Unit  $i$  is activated depends on the net input and the activation function of the unit. Typically, the resting state of a unit is indicated by the activation level of zero and the firing state is 1; however, depending on the architecture, a unit may take a value between  $-1$  and  $+1$  or between negative and positive infinity. The activation function determines activation level of the unit as a function of net inputs. Although activation function may be a simple linear function, it often takes a nonlinear form. One example is a threshold function (lower left panel of Fig. 2), where the activation level remains zero if the net inputs remain lower than a threshold value but its level becomes 1 when the net inputs exceeds the threshold.

Another example is a sigmoid function, where activation changes smoothly as a function of net inputs in the shape depicted (lower right panel of Fig. 2). These nonlinear activation functions are critical in some applications. The output of Unit  $i$  is determined by the activation level of the unit and its output function. Although output is often identical to activation (i.e.,  $o_i = a_i$ ), it may vary for some applications.

When activation spreads in only one direction (i.e., from Unit  $j$  to Unit  $i$ ), activation feeds forward from one set of units to another set in one time interval, whose activation then spreads to a third set in the next interval, and so on. This type of network, called a feedforward network, basically transforms one set of activations in the first set of units into another set of activations in the last set of units. Its main function is to convert one set of signals to another set of signals. In psychological terms, it may be useful, for instance, for modeling how a given situation or stimulus may induce a set of responses. However, activation may spread bidirectionally (from Unit  $j$  to Unit  $i$  as well as from Unit  $i$  to Unit  $j$ ), and the connection strength in one direction may or may not be the same as that in the opposite direction. In these cases, more complex dynamics emerge where activation may spread back and forth and reverberate through the network. This type of network, sometimes called a recurrent network, may tend toward a stable pattern of activation in which activation levels of the units change little or oscillate among several stable patterns. These networks may be used to model memory retrieval processes. Recurrent networks may sometimes exhibit an even more complex pattern known as chaos.

### 5. DYNAMICS OF INTERUNIT CONNECTIONS

Interunit connections in a connectionist network may be fixed or modifiable. In the latter case, connections are modified in accordance with a learning rule, so that the network may learn contingencies among events in its environment. For illustrative purposes, let us consider a simple example of a distributed feedforward network with two layers. In this network, each layer has three units and there is no connection among the units within each layer, but all of the units in one layer are fully connected to the units in the other layer (Fig. 3). Let us suppose that the network is to learn that when Event A happens, Event B follows. Further

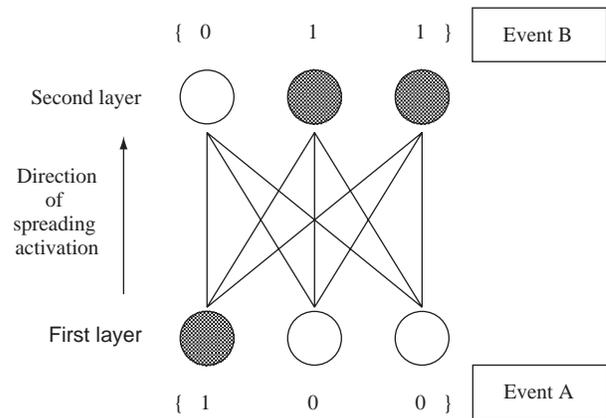


FIGURE 3 A two-layer distributed feedforward network. Shaded units are activated at +1. The bottom layer represents Event A, and the top layer represents Event B.

assume that Event A is encoded by the units in the first layer; when the first unit is activated (+1) but the other units are not activated (0), this means that Event A is represented in the network. Event B is encoded by the units in the second layer; when the first unit in this layer is not activated (0) but the second and third units are activated (+1), this means that Event B is represented in the network.

For the network to learn the association between Event A and Event B, the network needs to modify its connections so that when Event A is represented in the first layer, its activation would spread to the second layer and Event B should then be represented in the second layer. There are two basic types of learning rules that can accomplish this. One is called unsupervised learning, and the typical learning rule is called the Hebbian learning rule, named after the psychologist Donald Hebb. The rule is to strengthen the connection between the units that are activated at the same time. In this example, the connections between the first unit in the first layer and the second and third units in the second layer should be strengthened (e.g., from 0 to 1), whereas the other connections should remain unchanged (e.g., 0). In this case, assuming that input, activation, and output all are the same for a given unit, it is easy to see that the pattern of activation {1 0 0} in the first layer would spread to the second layer and activate the second and third units but would not activate the first unit in the second layer. More formally, this rule is more generally expressed by the following equation:  $\Delta w_{ij} = \epsilon a_i o_j$ , where  $\Delta w_{ij}$  is the amount of change in connection strength from Unit  $j$

to Unit  $i$ ,  $a_i$  is the activation level of Unit  $j$ ,  $o_j$  is the output of Unit  $j$ , and  $\varepsilon$  specifies a learning rate, that is, the parameter that governs the speed of learning.

The other type of learning is called supervised learning, for which the typical rule is the delta (or back-propagation) rule. Here, the connections are changed so that if Event A is represented in the first layer, it predicts Event B in the second layer. Event B acts as a “teacher” that is to be emulated by the network. Here is how the rule works. Assuming that all of the connection strengths are initially zero, all of the inputs to the units in the second layer would be zero. For the first unit in the second layer, then, the teacher is zero and the activation is also zero, that is, no error. When there is no error, the connection strengths remain the same. So, the connections coming into the first unit in the second layer remain zero. However, the teacher for the second and third units says that they should be activated (i.e., +1) even though they are not activated (i.e., 0). The difference between the teacher and the predicted activation is an error, which is then used to modify the connection strength. The rule is to change the connection strength to reduce the error so that the connection with the activated unit gets stronger but that with the resting unit remains the same. Therefore, the connection from the first unit in the first layer to the second and third units in the second layer is strengthened (e.g., from 0 to +1), but the other connections remain unchanged (e.g., 0). Again, this rule can modify the connection strengths appropriately so that the network predicts Event B when Event A is encoded. More generally, this rule is expressed by the following equation:  $\Delta w_{ij} = \varepsilon(t_i - a_i)o_j$ , where  $\Delta w_{ij}$  is the amount of change in connection strength from Unit  $j$  to Unit  $i$ ,  $t_i$  is the teacher’s activation level,  $a_i$  is the predicted activation level of Unit  $j$ ,  $o_j$  is the output of Unit  $j$ , and  $\varepsilon$  is the learning rate.

At one level, supervised learning differs from unsupervised learning in the presence of a teacher or supervisor in the learning process. Supervised learning is sometimes said to be less realistic because people learn without being told what they have expected is wrong by how much. However, even with supervised learning rules, there is no need to postulate the explicit presence of a human teacher. The only necessary assumption is that one of the events to be associated is regarded as a teacher. In other words, the “teacher” may be a human, a machine, or nature. More significantly, supervised and unsupervised learning rules differ in terms of their conceptualization of learning. On the one hand, unsupervised learning conceptualizes learning as the observation and learning of the co-occurrence of events in the world.

On the other hand, supervised learning presupposes a learning agent that implicitly predicts what follows what (e.g., Event B follows Event A) and learns from mistakes. In this sense, supervised learning may be able to model adaptation. However, networks with supervised learning may actually be too adaptive in that they might forget previously learned associations—sometimes even catastrophically—when new associations are learned. Unsupervised learning does not have this problem most of the time. Both the Hebbian and delta rules can be generalized to more complex forms.

## 6. EXAMPLES OF CONNECTIONIST ARCHITECTURES

A tensor network is an example of the connectionist architecture that embodies a generalized form of the Hebbian learning rule (Fig. 4). In the example in the previous section, the network learned the association between two events, but a tensor network is designed to learn to associate three or more events (Fig. 4 depicts a network that associates three events). One cluster of units represents one type of event, the units in one cluster are connected to the units in other clusters, and the connection among units (the connection among three units in Fig. 4) changes as a function of the activation levels of those units. This type of network is called a tensor network because the learning rule can be mathematically described by a mathematical representation called tensor. When one of the clusters of units is activated, its activation spreads to other units, which are then activated to retrieve stored information. A tensor network is suitable for integrating multiple pieces of information into a structured representation. It can bind together different types of information such as color, time of blooming, and

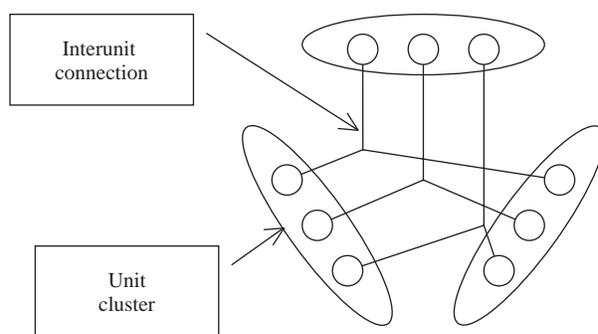


FIGURE 4 A tensor network that associates three events.

habitat; for instance, it can represent the name of a red flower that blooms during the early spring in wetlands. More generally, a tensor network can approximate the intersection of different sets.

A multilayer feedforward network is an example of the architecture that uses a generalized form of the delta rule (Fig. 5). In the previous example, there were two layers of units, but a multilayer network has three or more layers of units. The first and last layers are called input and output layers because they interface with the environment, encoding the incoming information (input layer) and decoding the outgoing response (output layer). The middle layers are called hidden layers because they are hidden from the environment. When the input layer units are activated, their activation spreads to the hidden layer units, which are then activated. The activation feeds forward so that the output layer units are activated. The output layer activation is compared to the relevant teacher, and the discrepancy (error) is then used to modify the connection strengths among the units. First, the connections between the output layer units and the units in the next last layer are modified, then the connections between the next last layer and the units before this layer (input layer in Fig. 5) are modified, and so on. Thus, the error information propagates backward, so to speak, to modify the connections. This is why the generalized delta rule is also called the backpropagation rule. Although a two-layer feedforward network cannot perform a simple logical operation of “exclusive or” (i.e., either/or), a three-layer network with a hidden layer can do so. More generally,

a hidden layer is useful for transforming an input representation to re-represent it.

Some architectures use the Hebbian or delta rule with further modifications to the basic network structure. In recurrent networks (sometimes called auto-associative networks), units may have connections that spread activation not only in one direction but also in the opposite direction. This type of architecture tends to amplify strong activation and dampen weak activation, sharpening the pattern of activation. It can be used for pattern completion; that is, when only a part of the original information is presented, this network can often fill in the missing part to complete the original pattern. Its practical uses include noise reduction and error correction.

Although these types of connectionist networks are based on somewhat different conceptions of learning and have different advantages and disadvantages, they may be combined to model more complex processes. A complex network that combines different types of network architectures is sometimes called a modular architecture. Each module is designed to perform a specified function, whose output is then used by another module for further processing. For example, a recurrent network may be used to clean noisy inputs so as to provide a clearer input to a tensor network, whose output may then be transferred to a multilayer feedforward network for modeling adaptive learning. Several modules may then function in tandem to provide a more complete model of human and animal psychological processes.

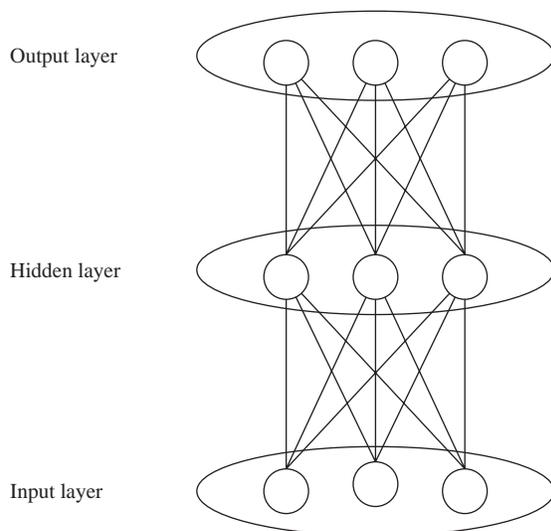


FIGURE 5 A three-layer feedforward network.

## 7. APPLICATIONS OF CONNECTIONISM

Connectionist networks provide versatile research tools for developing a model of human psychological processes. For example, they have been used to model some basic perceptual and cognitive processes such as pattern recognition, learning, memory, categorization, judgment and decision making, and natural language processing as well as more social psychological processes such as person impression formation, stereotyping, causal attribution, and stereotype formation and change. Some models have been used to discuss issues pertaining to developmental psychology. Some networks have been used to model neuropsychological issues; when parts of the interunit connections of a network are removed, the network exhibits behaviors that are similar to those of people with brain lesions. Other variants can model

animal behaviors as well. Although more applied areas of psychological research have been relatively slow to embrace connectionism, there are some signs of its use. Finally, the technology used in connectionist networks has been adopted not only to model psychological processes but also for engineering or other purposes. For instance, optic pattern recognition is useful for machine pattern recognition, and feedforward networks may be used for forecasting.

### See Also the Following Articles

Cognitive Skills: Training, Maintenance, and Daily Usage  
■ Cognitivism

### Further Reading

- Anderson, J. A., & Rosenfeld, E. (1988). *Neurocomputing*. Cambridge, MA: MIT Press.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1998). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Read, S. J., & Miller, L. C. (1998). *Connectionist models of social reasoning and social behavior*. Mahwah, NJ: Lawrence Erlbaum.
- Rumelhart, D. E., & McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.



# Conscientiousness

Robert R. McCrae

National Institutes of Health, Baltimore, Maryland, USA

1. Conscientiousness as a Dimension of Personality
  2. Origin and Development of Conscientiousness
  3. Applications
  4. Conclusion
- Further Reading

## GLOSSARY

**cross-observer agreement** Agreement between two or more observers on characteristics of a target person.

**exvia** In Cattell's system, a broad factor resembling introversion–extraversion.

**heritable** Determined in part by genes inherited from one's parents.

**intercorrelate** To vary together such that people who are high on one of the variables are likely to be high on the others as well.

**orthogonal** In factor analysis, referring to factors that are unrelated to each other.

**psychopathology** Mental and emotional disorders.

Conscientiousness is one of five major dimensions of individual differences in personality. It contrasts people who are methodical, purposeful, and deliberate with those who are disorganized, lazy, and hasty. Conscientiousness shows strong genetic influences. It is relatively low among adolescents and increases up to 30 years of age. Thereafter, mean levels and individual differences are generally stable throughout adulthood. Low levels of the Conscientiousness factor are associated

with several forms of psychopathology, whereas high levels are associated with academic achievement and superior job performance.

## 1. CONSCIENTIOUSNESS AS A DIMENSION OF PERSONALITY

Personality traits are tendencies to think, feel, and act in consistent ways. People in every culture learn to describe themselves and others in trait terms, with words such as *nervous*, *energetic*, *original*, *altruistic*, and *careful* referring to personality traits. The English language has several thousand such words. In addition, personality psychologists working in a variety of theoretical perspectives have identified traits such as ego strength, *exvia*, tolerance of ambiguity, dependency, and socialization.

All of these traits are similar in some respects. All are normally distributed; that is, a few people score low, most score in the average range, and a few score high. All traits are relatively stable across time. All of them affect behavior, but their impact is often weak and observable only when averaged across a wide sample of occasions.

It has become clear during the past 20 years or so that nearly all personality traits are related to one or more of five basic personality dimensions or factors, most commonly labeled Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Together, these factors and their constituent traits are known as the Five-Factor Model (FFM). Each of the five factors is a broad disposition defined by many narrower

traits. Furthermore, the factors are orthogonal in that people who are very high in Conscientiousness may be introverted or extraverted and may be agreeable or antagonistic.

Many personality theorists have described traits akin to Conscientiousness. These include Tupes and Christal's dependability, Tellegen's constraint, Gough's norm adhering, Cloninger's persistence, Digman's will to achieve, and Cattell's control. These various names reflect somewhat different conceptions of the Conscientiousness factor, with some suggesting a proactive purposefulness and others suggesting an inhibitive self-control. In an attempt to represent the full range of relevant traits, Costa and McCrae in 1991 assessed six facets of Conscientiousness in their Revised NEO Personality Inventory (NEO-PI-R): competence, order, dutifulness, achievement striving, self-discipline, and deliberation. Factor analyses have shown that these six facets intercorrelate and define a clear factor in both men and women, as well as in both adults and children, in many different languages and cultures.

## 2. ORIGIN AND DEVELOPMENT OF CONSCIENTIOUSNESS

Terms such as *hard-working*, *reliable*, and *persevering* describe desirable aspects of character. Indeed, Conscientiousness was omitted from some personality systems because it was once believed to be a moral evaluation rather than a real psychological attribute. The reality of individual differences in Conscientiousness has now been clearly established by studies of cross-observer agreement. Peer and expert ratings confirm the self-reports that people make about their degrees of Conscientiousness. Furthermore, both self-reports and observer ratings of Conscientiousness predict real-life outcomes such as academic success.

During most of the 20th century, psychologists believed that personality traits could be divided into two categories: temperament and character. Temperament traits were thought to be biologically based, whereas character traits were thought to be learned either during childhood or throughout life. With the advent of the FFM, behavior geneticists began systematic studies of the full range of personality traits, and it soon became clear that all five factors are substantially heritable. Identical twins showed very similar personality traits even when they had been separated at birth and raised apart, and this was equally

true for both character traits and temperament traits. Parents and communities influence the ways in which Conscientiousness is expressed, but they apparently do not influence its level.

Currently, little is known about Conscientiousness in young children because the self-report inventories typically used to assess it are not appropriate for that age group. It is obvious, however, that there are individual differences on this factor at an early age. We know, for example, that some children have attention deficit/hyperactivity disorder, which is characterized in part by problems with concentration, organization, and persistence—traits related to Conscientiousness. Longitudinal and cross-sectional studies suggest that Conscientiousness is relatively low among adolescents but increases between 18 and 30 years of age. Through most of the rest of adult life, there is relatively little change in the average level of Conscientiousness, although it is less clear what happens during extreme old age. Individual differences are strongly preserved, meaning that a careful, neat, and scrupulous 30-year-old is likely to become a careful, neat, and scrupulous 80-year-old.

## 3. APPLICATIONS

It matters greatly whether one is purposeful, diligent, and persistent or aimless, undisciplined, and impulsive. Consequently, Conscientiousness is associated with a variety of important life outcomes.

### 3.1. Mental Health and Psychotherapy

Individuals who are low in Conscientiousness are at risk for a number of psychiatric disorders. Because they act without due consideration of the long-term consequences of their actions, and because they lack self-control, they are prone to alcohol and substance abuse. Because they are chronically low in motivation to pursue goals, they are more susceptible to demoralization and depression. Because they do not have a strong sense of social obligations, they are more likely to be psychopaths or to have antisocial personality disorder. In 1998, Ransen and colleagues showed strong associations between Conscientiousness and adult attention deficit disorder.

In principle, very high levels of Conscientiousness might also lead to problems in living. Some theorists link high Conscientiousness to obsessive-compulsive

personality disorder and perfectionism. Highly conscientious people might also be workaholics whose task orientation interferes with family and social life.

The importance of Conscientiousness for clinical psychology and psychiatry was long underrated, in part because one of the chief measures of psychopathology, the Minnesota Multiphasic Personality Inventory (MMPI), does not include clear measures of this factor. It would appear that the assessment of Conscientiousness should be of considerable value in the diagnosis of psychopathology.

Conscientiousness has implications for psychotherapy as well. Individuals who are lax and careless might not show up for therapy sessions, and they are less likely to complete homework assignments. Clinicians whose patients are low in Conscientiousness might need to make special efforts or enlist the assistance of family members to compensate for these patients' low motivation and persistence.

### 3.2. Physical Health

For several reasons, Conscientiousness is an important determinant of physical health. Conscientious people have the self-discipline to develop and maintain good health habits. They exercise, stick to diets, and drink in moderation. Because they are thoughtful and deliberate, they avoid unnecessary risks in driving, and they are less likely to engage in unsafe sex. Conscientious individuals are likely to be better patients who are more willing and able to comply with prescribed medical regimens. Research has generally shown higher levels of adherence among highly conscientious patients, although many other variables are also involved. For at least some of these reasons, some studies have reported an association of Conscientiousness with longevity.

Conscientiousness may sometimes be affected by physical health, particularly brain injuries or diseases. In both Alzheimer's and traumatic brain injury patients, one of the most notable features is a precipitous decline in level of Conscientiousness. Observed declines in organized and purposeful behavior may prove to be early indicators of brain disease.

### 3.3. Academic Performance

Research on personality and academic performance can be traced as far back as Webb's 1915 study of character and intelligence. Webb identified a factor called persistence of motives or will that clearly anticipated the construct of Conscientiousness. Many

subsequent studies confirmed that the organization, diligence, neatness, and punctuality of highly conscientious students contribute to academic success.

Academic performance is also affected by Openness to Experience, and combinations of Conscientiousness and Openness define different learning styles. Students high on both factors are prototypical good students, pursuing a wide range of interests with careful and thorough study. Students who are conscientious but closed are more likely to excel in fields such as accounting and pharmacy that emphasize accuracy over imagination.

### 3.4. Job Performance

Conscientiousness is best known to applied psychologists as the dimension of personality most consistently associated with job performance. A series of meta-analyses have confirmed that highly conscientious people are rated as superior performers in nearly every occupation. Other personality factors are also related to job success, but usually only in specific job categories; for example, Openness is likely to be more useful to a novelist than to a banker. But organization, persistence, meticulousness, and achievement orientation are assets in any job.

Typically, Conscientiousness scores correlate roughly .20 to .30 with performance ratings. This correlation is small, in part because performance ratings are themselves imperfect measures and in part because vocational success also depends on a host of other factors, including intelligence, training and experience, and compatibility with coworkers. Nevertheless, other things being equal, employers would be better off hiring more conscientious candidates.

If test respondents are instructed to "fake good" on a personality inventory, they typically score high in Conscientiousness. Furthermore, job applicant samples typically score higher on desirable traits than do volunteers with no incentive to manage the impressions they are making. These facts have led some industrial and organizational (I/O) psychologists to question whether self-reports on personality inventories can be trusted. However, research to date has shown that Conscientiousness scores are valid predictors even when they are administered as part of a job selection battery.

## 4. CONCLUSION

Conscientiousness is the dimension of personality most directly relevant to the accomplishment of tasks. Individuals who score high on this factor have clear

goals and the ability to persevere in their efforts to attain them. It is possible, of course, that one's goals are inimical to society, and highly conscientious individuals can sometimes become dangerous criminals or entrenched dictators. For the most part, however, Conscientiousness is useful to both the individual and society, and anyone interested in success in life must take it into account.

### See Also the Following Articles

Intelligence Assessment ■ Intelligence, Emotional ■ Intelligence in Humans ■ Personality Assessment ■ Traits

### Further Reading

- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261–272.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*, 9–30.
- Conner, M., & Abraham, C. (2001). Conscientiousness and the theory of planned behavior: Toward a more complete model of the antecedents of intentions and behavior. *Personality and Social Psychology Bulletin, 27*, 1547–1561.
- Costa, P. T., Jr., & McCrae, R. R. (1998). Six approaches to the explication of facet-level traits: Examples from Conscientiousness. *European Journal of Personality, 12*, 117–134.
- Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12*, 887–898.
- De Raad, B., & Schouwenburg, H. C. (Eds.). (1996). *Personality, learning, and education* [special issue]. *European Journal of Personality, 10*(5).
- Han, K., Weed, N. C., & McNeal, T. P. (1996). Searching for Conscientiousness on the MMPI-2. *Journal of Personality Assessment, 67*, 354–363.
- Ranseen, J. D., Campbell, D. A., & Baer, R. A. (1998). NEO-PI-R profiles of adults with attention deficit disorder. *Assessment, 5*, 19–24.
- Trost, K. K., Herbst, J. H., Masters, H. L., III, & Costa, P. T., Jr. (2002). Personality pathways to unsafe sex: Personality, condom use, and HIV risk behaviors. *Journal of Research in Personality, 36*, 117–133.
- Webb, E. (1915). *Character and intelligence: An attempt at an exact study of character*. Cambridge, UK: Cambridge University Press.



# Conservation Behavior

Florian G. Kaiser

Eindhoven University of Technology, Eindhoven, The Netherlands

1. Conceptions of Conservation Behavior
  2. Performance and Consequences
  3. Shaping Performance
- Further Reading

## GLOSSARY

**conservation behavior** Actions such as recycling, energy and water conservation, and the use of public transportation that contribute toward environmental preservation and/or conservation.

**ecological footprint** A measure of the totality of the environmental resources used and waste produced by a single actor (i.e., a person, a company, a nation).

**face value action** Behavior defined from an observer's viewpoint by its apparent consequences.

**factor analysis/extraction** A set of mathematical models designed to search for one or more underlying latent (i.e., unobservable) variables, the so-called factors; factors account for the intercorrelations among a set of (observed) variables.

**goal-directed action** Behavior defined from an actor's viewpoint as a means to achieve a subjective goal.

**item** A task that provokes a response that is the basis for designing variables.

**item response theory (IRT)** A set of measurement models in which the observed variable is understood as a function of one or more latent (i.e., unobservable) traits and one or more item features (e.g., difficulty, discrimination, guessing chance).

**Rasch model** The most basic model within item response theory; in the dichotomous Rasch model, the probability of an approving or correct response in an observed

variable is expected to depend on two parameters only: the difficulty of the item and a latent (i.e., unobservable) trait of the respondent.

**true score model** The measurement model fundamental to classical test theory (CTT), the prevailing measurement model in psychology; in this model, the scores in the observed variable are a composite of the persons' (unknown) true scores and a random error.

**varimax rotation** A rotation method that generates orthogonal (i.e., noncorrelated) factor structures; rotation methods are designed to transform the solution from factor extraction into a solution that is easier to interpret.

Conservation behavior can be approached from either an observer's or an actor's viewpoint. When behavior is defined from the outside by its face value, even seemingly similar behaviors such as glass, paper, and battery recycling fall into distinct categories. When behavior is defined from the inside as a goal-directed performance, even diverse acts such as recycling and ownership of solar panels appear to belong to one class of actions. Correspondingly, there are two distinct research traditions that differ substantially.

## 1. CONCEPTIONS OF CONSERVATION BEHAVIOR

The impact of individuals' actions on the natural environment has led to ecological transformations of yet unknown gravity. Psychological analyses of

environmental problems have focused on effectively promoting more ecologically sustainable patterns and courses of action such as recycling, choice of transportation, and energy conservation. According to a recent *Journal of Social Issues* article by Stern, behaviors such as these, like other human performances, can be approached from two different angles: from an observer's viewpoint, when behavior is defined by its face value (which is evident in its consequences), and from an actor's subjective viewpoint, when behavior is defined as a motivated or goal-directed performance, namely, as the behavioral means by which people try to achieve a particular goal. Based on these two conceptions of behavior, there are two distinct research traditions that differ substantially in their behavior typology and measurement approach, in the ways in which they link behavior to its environmental consequences, and in the ways in which they tackle behavior change.

### 1.1. Face Value Action: A Multitude of Behavior Types

As a person drives to the grocery store, to the kindergarten, or to work, he or she is using a car that consumes fossil fuel and pollutes the air. In general, the environmental consequences—air pollution and resource consumption—are the basis for classifying a behavior as conservational or, in this case, nonconservational or environmentally harmful. An example of this view is Axelrod and Lehman's definition that categorizes "actions which contribute towards environmental preservation and/or conservation" as being conservational. Strictly speaking, however, all behaviors have environmental consequences given that every behavior uses the physical environment in which it takes place. Taking a leisurely walk down the sidewalk, for instance, involves breathing air and rubbing off microscopic pieces of concrete and rubber. Because behavior is generally consumptive, it is the relative extent and the particular quality of the environmental consumption that matter in this face value conception of conservation behavior.

When behavior is defined by its face value as it normally is, measurement is traditionally based on factor analysis, varimax rotation, and the true score (measurement) model. In this tradition, even seemingly similar behaviors such as glass, paper, and battery recycling are commonly found to fall into distinct and independent categories of conservation behavior. In turn, psychological research has focused on rather

specific behaviors, such as environmental activism, consumerism, and energy conservation, with a dominant preference for recycling behavior, as was pointed out by Vining and Ebreo. Studying behavior rather particularly, however, ultimately means that modification programs have to be specifically tailored to each type of conservation behavior.

### 1.2. Goal-Directed Action: One Class of Behavioral Means

Even when a behavior has environmentally harmful consequences, the individual might not necessarily intend to cause damage. At most, the individual accepts harm as a side effect of a particular performance. Thus, from a motivational viewpoint, air pollution and resource consumption often are irrelevant. For example, by taking the car to the grocery store, to the kindergarten, or to work, a person tries to increase his or her comfort level or to decrease transit time. The environmental consequences of the action are by and large ignored. However, if the person's intent is not to pollute the air or use fuel but rather to easily transport groceries home or give his or her child a safe ride, we cannot expect environmental values and attitudes to be the essential internal drives that determine car use. Instead, child-rearing and convenience-related considerations are presumably more crucial. Obviously, it is not a behavior's apparent face value that matters from a motivational viewpoint; rather, it is the personal reason, that is, the subjective goal or purpose that accompanies the act. By this logic, behavior cannot be meaningfully defined by its face value (i.e., by its obvious consequences); instead, it must be defined by its underlying motive (i.e., the goal for which a behavior is a means to an end).

When behavior is defined from a motivational viewpoint as a goal-directed performance (i.e., as the behavioral means necessary to achieve the conservation goal), research has unanimously found even seemingly diverse acts such as glass recycling and ownership of private solar panels to belong to one class of actions (based on the Rasch model within item response theory).

According to Kaiser and Wilson's model, a person's dedication to a conservation goal is most obvious in the face of increasingly demanding hurdles and progressively intolerable sacrifices. Similarly, the more obstacles a person overcomes and the more effort he or she expends along the way to the goal, the more evident

that person's commitment is to the particular cause. Likewise, if the tiniest difficulty is enough to stop a person from taking the goal-implied behavioral steps, devotion to this goal likely is rather low. In other words, the more demanding the behavioral tasks, the greater the commitment to achieving the goal implied by the performances (and vice versa).

In this second research tradition, measurement of conservation behavior essentially takes on the format of a performance test. Thus, motivationally relevant goal-directed behavior cannot be identified by the inspection of single acts. It can be measured only when several performances—all of which are necessary to achieve a particular goal—are considered and when each behavior can be characterized by the personal effort (i.e., the behavioral costs) involved in its realization. These costs are determined by the sociocultural boundary conditions (e.g., terrain, climate, customs, legal regulations, economy, societal infrastructure) in which an act takes place.

## 2. PERFORMANCE AND CONSEQUENCES

Regardless of whether one favors a consequences- or a goal-based conception of behavior in the conservation domain, it is the environmental impact that matters rather than conduct per se. Thus, psychology's ultimate aspiration must be to reduce people's ecological footprints in terms of pollution and energy and other resources consumed.

### 2.1. A Primary Focus on Consequences

In a recent *American Psychologist* article, Stern advised psychologists to justify their choice of behavior to study by its relative contribution to environmental preservation. In contrast, others decide to skip behavior entirely and to aim directly at predicting environmental consequences such as the amount of energy or water people consume. The former strategy generally diminishes psychology's significance by encouraging comparison of the impact of individual behavior with that of commercial sectors or industry. Individual households account for only approximately one-third of energy use and for only approximately 5% of solid waste. Such a focus necessarily points to the relative irrelevance of people's motives, values, and beliefs and ultimately to the relative insignificance of psychology.

Research that directly addresses the consequences rather than individual performances typically finds the boundary conditions of an act—the technology that is used, product price, and household size—to be by far the most significant determinants in affecting a person's ecological footprint. For example, regardless of any motivational differences, people consume relatively more energy for heating during the winter than during the summer. Thus, despite the fact that the absolute quantity might be decreased with an extremely conservational attitude compared with a more mundane position, the outside temperature will generally dominate the absolute amount of energy that is consumed. This relative supremacy of the external conditions is especially striking for the most environmentally crucial actions, that is, the expensive and/or high-effort performances such as retrofitting homes and purchasing energy-efficient cars.

### 2.2. A Deficient Relationship

Focusing on the environmental impact of people's actions obviously takes the face value definition of behavior to its extreme. However, with shared variances of 5 to 15%, the link between conservation behavior and its environmental impact is far from perfect. This finding also explains the diverse results of research based on the two discrepant criteria: behavior and its consequences. Concomitantly, if acting does not have the desired outcome, we can also conclude that people lack knowledge about the actual environmental consequences of what they themselves call conservation behavior. In other words, people may know about the general class of actions that contribute to environmental conservation and/or preservation, but at the same time they might remain ignorant of the relative environmental effectiveness of different behaviors. Logically, consequences are an unreasonable substitute for behavior.

### 2.3. No Goal, No Drive

Focusing too closely on the consequences ignores the actor's perspective in two essential ways. First, there are myriad environmental consequences that people should or could take into account—energy savings, biodiversity-relevant effects, greenhouse gases, and various air, soil, or water pollutants—all of which are worthwhile to consider but are far too numerous to reflect on altogether concurrently and constantly. Second, if a certain amount of energy to be saved is the ultimate goal, an actor needs to know about the

quantitative conservation effects of the various behavioral alternatives. For instance, what saves how much or relatively more energy: purchasing a fuel-efficient car or driving less? Without such knowledge, we cannot expect people to form a corresponding quantitative performance goal, and without an individual goal, there is no motive for a person to strive for it either. In other words, we cannot expect subjective reasons to be significant for a goal that does not exist. In summary, a focus on the environmental consequences makes it unlikely to find psychological determinants to be notable, and if we wish to assess psychology's factual significance in contributing to environmental conservation, psychologists must pay attention to behavior.

### 3. SHAPING PERFORMANCE

Despite environmentalists' and environmental scientists' arguments to the contrary, various conservation behaviors seem to fall into one mental category for a majority of people. This implies that various people who wish to act in a conservational way do not necessarily take on the same behavioral tasks to achieve their conservation goals. Whereas one person rides a bike to work, another person uses a car for daily commuting but compensates by avoiding the use of batteries. Whereas one person perceives his or her contribution to environmental preservation by switching to a vegetarian diet, another person sees some irregular recycling of bottles as his or her part. This explains why people excuse their unecological behavior by pointing to their conservational engagement. For example, an intercontinental holiday flight can subjectively be compensated by regularly buying seasonal produce, recycling, and so forth. According to Thøgersen, this interchangeability of various conservation behaviors, from a subjective viewpoint, has not received much attention yet.

#### 3.1. Multiple Behaviors, Determinants, and Complex Interventions

When behavior is defined by its face value as it normally is, conservation behaviors are found to fall into distinct mental categories. However, if avoiding household chemicals and energy conservation are unrelated, substituting one with the other is unreasonable. Not surprisingly, various behaviors have different determinants. And predictably, an integrated theory of the

antecedents of conservation behavior has become a remote goal as well. What psychologists normally find is a multitude of influences differentially significant with various behaviors. And as it turns out, with many specific behaviors, the sociocultural boundary conditions result in the most striking effects. In terms of effect size, sociocultural boundary conditions are clearly more prominent than environmental motives, attitudes, and values, and sociocultural boundary conditions often are even more pronounced than the numerous psychological factors combined. Axiomatically, the more powerful the external conditions are, the less influential the internal ones will be.

In line with Stern in his *American Psychologist* article, conservation psychologists infer that the presence of the appropriate psychological condition, such as the right attitude, explains behavior only when there are relatively insignificant external barriers present—and even then, to a modest degree at best. Predictably, increasing sociocultural opportunities or removing such barriers to performing conservation behaviors often achieves the most powerful effects, especially when the conditions are demanding.

The psychological measures and strategies that are thought to be effective in this line of research are complex because they need to incorporate the fact that people are diverse, live under a variety of different circumstances, and respond to various interventions differentially. In contrast, intervention strategies that consist of single measures, such as a particular persuasion technique that aims at changing people's environmental attitude, often fail to have the desired behavioral consequences. However, the more sophisticated interventions become, the less efficient they are and the more time, effort, and money that are needed to develop and employ them—and, consequently, the less appealing they are for policymakers to widely promote behavior change.

#### 3.2. Differentially Striving for a Single Goal

Because people can do different things to strive for a conservation goal, they can normally choose among various behavioral alternatives. Correspondingly, we cannot deterministically predict future behaviors; that is, we cannot foresee whether people will limit their leisure time mobility, recycle glass, or retrofit their houses. All we can safely expect is a prudent selection of the behavioral means to achieve the goal. Because any behavior being performed requires personal

resources such as time, money, and effort, we can anticipate that if two behaviors seem equally suited to achieve an objective, people will favor the relatively less demanding action over the more challenging one. Hence, if a conservation goal can be accomplished with a variety of different acts, people will most likely go with the less strenuous or less costly one.

Moreover, we can also predict that a person will engage in an act only if his or her motivation to achieve the conservation goal exceeds, or at least matches, the behavioral costs involved in its realization. In other words, the more central the conservation goal is to the person's life, the more demanding the behavioral tasks the person takes on. Predictably, Kaiser and colleagues repeatedly found that people's motivation (i.e., their intention to achieve a conservation goal) almost perfectly reflected in their goal-directed performances.

Behavior interventions, such as improving environmental motivation, generally leave people with a choice. Because the conservation goal can be accomplished with a variety of different acts, people can engage in several activities. They can buy energy-efficient light bulbs, recycle paper, refrain from using a dryer for their laundry, purchase solar panels for their own energy supply, boycott companies with poor conservation reputations, contribute to environmental organizations, and so on. Logically, because psychological measures always aim at the entire range of behavioral means, researchers cannot confirm successful behavior changes when they opt for specific and narrow performance criteria. In other words, the neglect of alternative behavioral means by which people can strive for the conservation goal makes it likely that researchers underestimate the effectiveness of psychological interventions.

Focusing on the entire range of the behavioral means implies that researchers' attention ultimately shifts away from single environmental consequences, such as the amount of heating energy or certain carbon dioxide emissions, and moves toward a person's overall environmental effect. Necessarily, with a goal-directed behavior perspective, psychology addresses the person's entire ecological footprint as well.

## Acknowledgments

I am grateful to Steven Ralston for his language support and to Terry Hartig, Gundula Hübner, Anneloes Meijnders, Jürgen Rost, and Wesley Schultz for their comments on earlier versions of this article.

## See Also the Following Articles

Motivation and Culture ■ Motives and Goals

## Further Reading

- Axelrod, L. J., & Lehman, D. R. (1993). Responding to environmental concern: What factors guide individual action? *Journal of Environmental Psychology*, *13*, 149–159.
- Becker, L. J., Seligman, C., Fazio, R. H., & Darley, J. M. (1981). Relating attitudes to residential energy use. *Environment & Behavior*, *13*, 590–609.
- Gatersleben, B., Steg, L., & Vlek, C. (2002). Measurement and determinants of environmentally significant consumer behavior. *Environment & Behavior*, *34*, 335–365.
- Kaiser, F. G., & Wilson, M. (2004). Goal-directed conservation behavior: The specific composition of a general performance. *Personality and Individual Differences*, *36*, 1531–1544.
- McKenzie-Mohr, D. (2000). Fostering sustainable behavior through community-based social marketing. *American Psychologist*, *55*, 531–537.
- Stern, P. C. (2000a). Psychology and the science of human–environment interactions. *American Psychologist*, *55*, 523–530.
- Stern, P. C. (2000b). Toward a coherent theory of environmentally significant behavior. *Journal of Social Issues*, *56*, 407–424.
- Tanner, C., Kaiser, F. G., & Wöfling Kast, S. (2004). Contextual conditions of ecological consumerism: A food-purchasing survey. *Environment & Behavior*, *36*, 94–111.
- Thøgersen, J. (1999). Spillover processes in the development of a sustainable consumption pattern. *Journal of Economic Psychology*, *20*, 53–81.
- Vining, J., & Ebreo, A. (2002). Emerging theoretical and methodological perspectives on conservation behavior. In R. B. Bechtel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 541–558). New York: John Wiley.



# Consultation Processes in Schools

Joseph E. Zins

University of Cincinnati, Cincinnati, Ohio, USA

Charlene R. Ponti

Hamilton County Educational Services Center,  
Cincinnati, Ohio, USA

1. Introduction
2. Definition
3. Key Components and Applications  
Further Reading

## GLOSSARY

**consultant** A professional who provides consultative assistance.

**consultee** A person who seeks assistance from a helping professional for someone for whom they bear responsibility (e.g., a student).

**ecobehavioral** Theoretical orientation that combines ecological (broad viewpoint that recognizes environmental, contextual, and individual factors as influencing behavior) and behavioral (focusing on elucidating and changing inappropriate or maladaptive behavior) perspectives to help understand people's actions.

**empower** Acquire the means or skills to obtain resources necessary for success.

**indirect service** Assistance provided through a third party rather than directly as in teaching or counseling.

Consultation is the process of establishing a partnership with a parent or professional to provide problem-solving assistance regarding a child or student for whom the parent or professional is responsible. The goal is to promote students' academic and/or social-emotional well-being and performance.

## 1. INTRODUCTION

Most educators find that they are better able to provide the highest quality education services for their students when they work as partners with other staff and families to address student-related problems and promote competence. School consultation is a systematic process for implementing this partnership and engaging in problem solving. This article describes the knowledge, beliefs, and procedures required for successful practice.

Interest in school consultation became prominent during the late 1960s and 1970s, and there continues to be a high demand for training and practice in this area today. The impetus was that many school practitioners found that demands on their time for direct one-on-one service exceeded their capacity to provide assistance, and they also observed that many problems could be solved or prevented from becoming more serious if teachers and parents were given relatively immediate support in addressing problems that arose. Consultation addressed both of these needs, and it continues to be an important job activity for many school professionals who provide support services, including school psychologists, special educators, guidance counselors, school social workers, speech and language pathologists, and school nurses. As a result, many universities that offer professional training in education and applied psychology now offer formal coursework and practicums in this area. Further

support for such services results from the mounting body of empirical evidence demonstrating the effectiveness of consultative services.

## 2. DEFINITION

School consultation refers to the process of providing preventively oriented, systematic, reciprocal problem-solving assistance to empower consultees, thereby promoting students' academic and/or social-emotional well-being and performance. Participants form cooperative partnerships using an ecobehavioral perspective, with the consultees retaining responsibility for the students. The process is distinguished from "collaboration" and other services in that students are helped indirectly through the consultative problem-solving process.

## 3. KEY COMPONENTS AND APPLICATIONS

### 3.1. Participants

Consultation can involve two or more individuals. The person who provides help (e.g., school psychologist, special educator) is the consultant, and those seeking assistance are consultees (e.g., teachers, parents). The majority of school consultation takes place on a one-on-one basis, but group consultation frequently occurs as well, as with intervention assistance teams and in individualized education program (IEP) meetings or when major school reform initiatives are introduced by an external consultant. A critical distinction between consultation and other services is that the student who is the focus of the problem solving generally is not directly involved in the consultation sessions. In other words, the student is helped indirectly rather than directly as in teaching or counseling.

### 3.2. Ecobehavioral Perspective

Students and problems must be examined from a broad perspective because many times the issues being discussed have their origins in factors outside of the students and students live in systems outside of the school (e.g., family, community). Frequent causes of problems include the manner in which instruction is delivered, limited opportunities provided for student engagement in class, methods used for behavior management, social relationships, district policies, and

family issues. However, sometimes problems are related to factors internal to students (e.g., those resulting from medical conditions such as attention deficit/hyperactivity disorder, autism, and traumatic brain injury) and to environmental factors (e.g., poverty, emotional neglect or abuse). A behavioral approach is commonly used to identify and define problems, understand the contingencies maintaining them, alter consequences to change the problem behavior, and assess the outcomes of the intervention.

### 3.3. Preventive Orientation

Consultation makes assistance more readily available rather than requiring a formal referral for an evaluation and then waiting for extended periods of time for it to take place as is traditionally done in many schools. Through consultation, problems can be solved in a timely manner and consultees are likely to seek assistance early before problems become more entrenched and habitual. Likewise, through the process, consultees improve their own skills and learn techniques for solving problem situations that they can apply to similar issues in the future.

### 3.4. Problem Solving

The essence and purpose of consultation is to solve problems. The starting point is problem identification. If problems are correctly identified, it is highly likely that they will be solved successfully. A structured systematic approach to problem solving is followed, as outlined in [Table I](#).

### 3.5. Relationship and Interpersonal Aspects

Consultants exhibit warmth, genuineness, and empathy, and the consultant-consultee relationship is characterized by mutual respect, trust, and cooperation. Establishing a partnership with colleagues and parents is critical for effective consultative problem solving. What this means is that the consultant not only must view consultees as partners who possess expertise but also must actively involve and engage them as the problem-solving process unfolds. As a result, consultees are more likely to support and carry out interventions that are developed.

**TABLE I**  
**Consultative Problem-Solving Process**

---

Establishment of cooperative partnership

- Promote understanding of each other’s roles and responsibilities
- Avoid the “egalitarian virus” (failure to challenge erroneous opinions)

Problem identification and analysis

- Define problem in behavioral terms and obtain agreement with consultee
- Collect baseline data regarding problem frequency, duration, and/or intensity and conduct task analysis as needed
- Identify antecedent determinants of the problem behavior
- Identify consequences that may maintain the behavior
- Assess other relevant environmental factors
- Identify all available resources

Intervention development and selection

- Brainstorm range of possible interventions
- Evaluate the positive and negative aspects of the interventions
- Select intervention(s) from the alternatives generated

Intervention implementation, sustainability, evaluation, and follow-up

- Clarify implementation procedures and responsibilities
- Implement the chosen strategy
- Evaluate intended outcomes and any side effects
- Program generalization, plan maintenance, and develop appropriate fading procedures
- Recycle and follow up as necessary

---

Source. From Zins, J. E., & Erchul, W. P. (2002). Best practices in school consultation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV* (pp. 625–643). Bethesda, MD: National Association of School Psychologists. Copyright 2002 by the National Association of School Psychologists. Reprinted with permission of the publisher.

### 3.6. External Consultants

In addition to having special services and other staff providing consultation services, some schools hire outside experts to assist with specific projects, for example, introducing a new program such as Positive Behavioral Intervention and Support Systems and evaluating district efforts to achieve No Child Left Behind Act goal attainment. The services these individuals offer can be highly specialized, but these individuals generally work with districts on a time-limited basis.

### 3.7. Other Approaches to Consultation

There are many approaches to consultation in schools. Although a behavioral approach, as described in this article, is used most often in schools, mental health and organizational consultation often can be found. The essential commonalities among these approaches far exceed the differences. The term “consultation” should not, however, be confused with the “expert advice-giving” approach typically found in medical settings. The latter approach is very different from the egalitarian process described in this article.

#### See Also the Following Articles

Home–School Collaboration ■ School–Community Partnerships

#### Further Reading

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation in applied settings*. New York: Plenum.

Caplan, G. (1970). *The theory and practice of mental health consultation*. New York: Basic Books.

Erchul, W. P., & Martens, B. K. (2002). *School consultation: Conceptual and empirical bases of practice* (2nd ed.). New York: Kluwer Academic.

Friend, M., & Cook, L. (2003). *Interactions: Collaboration skills for school professionals* (4th ed.). New York: Allyn & Bacon.

Gutkin, T. B., & Curtis, M. J. (1999). School-based consultation theory and practice: The art and science of indirect service delivery. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 598–637). New York: John Wiley.

Medway, F. J., & Updyke, J. F. (1985). Meta-analysis of consultation outcome studies. *American Journal of Community Psychology*, 13, 489–504.

Sheridan, S. M., Welch, M., & Orme, S. F. (1996). Is consultation effective? A review of outcome research. *Remedial and Special Education*, 17, 341–354.

Zins, J. E., & Erchul, W. P. (2002). Best practices in school consultation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. 1, pp. 625–643). Bethesda, MD: National Association of School Psychologists.

Zins, J. E., & Johnson, L. J. (1994). Prereferral intervention for students with special needs. In T. Husén, & T. N. Postlethwaite (Eds.), *The international encyclopedia of education, Vol. 8: Research and studies* (2nd ed., pp. 4657–4662). Oxford, UK: Elsevier Science.

Zins, J. E., Kratochwill, T. R., & Elliott, S. N. (Eds.). (1993). *Handbook of consultation services for children: Applications in educational and clinical settings*. San Francisco: Jossey–Bass.



# Consumer Psychology

Michael R. Solomon

Auburn University, Auburn, Alabama, USA

1. I Buy, Therefore I Am
  2. Consumers as Decision Makers
  3. Psychological Influences on Consumption
  4. Cultural and Interpersonal Influences on Consumption
- Further Reading

## GLOSSARY

**archetypes** Universally shared ideas and behavior patterns, involving themes such as birth, death, and the devil, that frequently appear in myths, stories, and dreams.

**brand loyalty** A form of repeat purchasing behavior reflecting a conscious decision to continue buying the same brand.

**consumer addiction** A physiological and/or psychological dependence on products or services.

**consumer psychology** The study of the processes involved when individuals or groups select, purchase, use, or dispose of products, services, ideas, or experiences to satisfy needs and desires.

**consumer satisfaction/dissatisfaction (CS/D)** The attitude a person has about a product after it has been purchased.

**consumption situation** Factors beyond characteristics of the product that influence a purchase decision.

**diffusion of innovations** The process whereby a new product, service, or idea spreads through a population.

**evaluative criteria** Dimensions used to judge the merits of competing options.

**evoked set** The options identified by a consumer that may satisfy a need.

**expectancy disconfirmation model** Model whereby consumers form beliefs about product performance based on

prior experience with the product and/or communications about the product that imply a certain level of quality.

**heuristics** Mental rules-of-thumb that simplify decision making.

**information search** A scan of the environment to identify the options available to satisfy a need.

**lifestyle** A pattern of consumption reflecting a person's choices of how he or she spends time and money.

**market segmentation** The process of identifying groups of consumers who are similar to one another in one or more ways and devising marketing strategies that appeal to the needs of one or more of these groups.

**need recognition** The stage in decision making when the consumer experiences a significant difference between his or her current state of affairs and some desired state.

**opinion leader** A person who is frequently able to influence others' attitudes or behaviors.

**perceived risk** The belief that a poor product choice will produce potentially negative consequences.

**psychographics** Data about a person's attitudes, interests, and opinions (AIOs) that allow marketers to cluster consumers into similar groups based on lifestyles and shared personality traits.

**reference group** An actual or imaginary individual or group that influences an individual's evaluations, aspirations, or behavior.

**self-concept** The beliefs a person holds about his or her own qualities and how he or she evaluates these qualities.

**self-image congruence models** Models that examine the process of cognitive matching between product attributes and the consumer's self-image.

**subculture** A group whose members' shared beliefs and common experiences set them apart from others in the larger culture to which they belong.

**symbolic self-completion theory** Theory that argues that people who have an incomplete self-definition tend to bolster this identity by acquiring and displaying symbols associated with it.

**values** Culturally determined general ideas about good and bad goals.

Consumer psychology is the study of the processes involved when individuals or groups select, purchase, use, or dispose of products, services, ideas, or experiences to satisfy needs and desires. The decision to consume typically is the culmination of a series of stages that include need recognition, information search, evaluation of alternatives, purchase, and postpurchase evaluation. However, in some cases (especially when involvement with the product or service to be chosen is low), this rational sequence is short-circuited as consumers make decisions based on “shortcuts” called heuristics (e.g., “Choose a well-known brand name”). In other cases (especially when involvement with the product or service to be chosen is especially high, as is the case with extremely risky decisions or when the object carries extreme emotional significance to the individual), subjective criteria also may cause the person’s choice to diverge from the outcome predicted by a strictly rational perspective on behavior. Indeed, many consumer behaviors, including addictions to gambling, shoplifting, and even shopping itself, are quite irrational and may literally harm the decision maker. The study of consumer psychology underscores the importance of individual and group variables that help to shape preferences for products and services. In addition to demographic differences such as age, stage in the life cycle, gender, and social class, psychographic factors such as personality traits often play a major role. A person’s identification with others who constitute significant reference groups or who share the bonds of subcultural memberships also exerts a powerful impact on his or her consumption decisions. These macro influences on behavior make it more or less likely that an individual will choose to adopt new products, ideas, or services as these innovations diffuse through a market or culture.

## 1. I BUY, THEREFORE I AM

Consumer psychology is the study of the processes involved when individuals or groups select, purchase, use, or dispose of products, services, ideas, or experiences to satisfy needs and desires. The field embraces many kinds

of consumption experiences, ranging from canned peas, a massage, or democracy to hip-hop music or a celebrity such as Madonna. Needs and desires to be satisfied range from physiological conditions such as hunger and thirst to love, status, or even spiritual fulfillment.

Consumers take many forms, ranging from an 8-year-old girl begging her mother for Pokémon cards to an executive in a large corporation deciding on a multi-million-dollar computer system. This article focuses on individual consumers, but with the caveat that many important issues relate to the psychology of group decision making involving dyads, families, and organizations. One must also recognize that the study of consumer behavior is extremely interdisciplinary. Although psychology is one of the core disciplines that have shaped the field, many other important perspectives from economics, sociology, and other social sciences also play a dominant role.

During its early stages of development, the field was often referred to as buyer behavior, reflecting an emphasis on the exchange, a transaction in which two or more organizations or people give and receive something of value. Most marketers now recognize that consumer behavior is in fact an ongoing process and not merely what happens at the moment a consumer hands over money or a credit card and, in turn, receives some good or service. This expanded view emphasizes the entire consumption process, which includes the issues that influence the consumer before, during, and after a purchase.

One of the fundamental premises of the modern field of consumer psychology is that people often buy products not for what they do but for what they mean. This principle does not imply that a product’s basic function is unimportant; rather, it implies that the roles that products play in our lives extend well beyond the tasks that they perform. The deeper meanings of a product may help it to stand out from other similar goods and services. All things being equal, a person will choose the brand that has an image (or even a personality) consistent with the purchaser’s underlying needs.

People’s allegiances to certain sneakers, musicians, or even soft drinks help them to define their place in modern society, and these choices also enable people to form bonds with others who share similar preferences. Following are some of the types of relationships a person might have with a product:

- *Self-concept attachment.* The product helps to establish the user’s identity.
- *Nostalgic attachment.* The product serves as a link with a past self.

- *Interdependence.* The product is a part of the user's daily routine.
- *Love.* The product elicits emotional bonds of warmth, passion, or other strong emotion.

Self-concept refers to the beliefs that a person holds about his or her own qualities and how he or she evaluates these qualities. People with low self-esteem may choose products that will enable them to avoid embarrassment, failure, or rejection. For example, in developing a new line of snack cakes, Sara Lee found that consumers low in self-esteem preferred portion-controlled snack items because they believed that they lacked the self-control to regulate their own eating.

A consumer exhibits attachment to an object to the extent that it is used to maintain his or her self-concept. Objects can act as a "security blanket" by reinforcing people's identities, especially in unfamiliar situations. For example, students who decorate their dorm rooms with personal items are less likely to drop out of college. This coping process may protect the self from being diluted in a strange environment. Products, especially those that serve as status symbols, also can play a pivotal role in impression management strategies as consumers attempt to influence how others think of them. Despite the adage, "You can't judge a book by its cover," in reality people often do.

The use of consumption information to define the self is especially important when an identity is yet to be adequately formed, as occurs when a person plays a new or unfamiliar role. Symbolic self-completion theory suggests that people who have an incomplete self-definition tend to bolster this identity by acquiring and displaying symbols associated with it. For example, adolescent boys may use "macho" products, such as cars and cigarettes, to augment their developing masculinity; these items act as a "social crutch" during periods of self-uncertainty.

Self-image congruence models assume a process of cognitive matching between product attributes and the consumer's self-image. Research tends to support the idea of congruence between product use and self-image. One of the earliest studies to examine this process found that car owners' ratings of themselves tended to match their perceptions of their cars; for example, Pontiac drivers saw themselves as more active and flashy than Volkswagen drivers saw themselves. Some specific attributes useful in describing matches between consumers and products include rugged/delicate, excitable/calm, rational/emotional, and formal/informal.

## 2. CONSUMERS AS DECISION MAKERS

Traditionally, consumer researchers have approached decision making from a rational perspective. In this view, consumers calmly and carefully integrate as much information as possible with what they already know about a product, weigh the pluses and minuses of each alternative painstakingly, and arrive at satisfactory decisions.

### 2.1. Stages in the Decision-Making Process

This process implies that marketing managers should carefully study the stages in decision making shown in Fig. 1 to understand how product information is obtained, how beliefs are formed, and what product choice criteria are specified by consumers. This knowledge will enable them to develop products that emphasize appropriate attributes and to tailor promotional strategies to deliver the types of information most likely to be desired in the most effective formats.

#### 2.1.1. Need Recognition

The decision-making process begins with the stage of need recognition, when the consumer experiences a significant difference between his or her current state

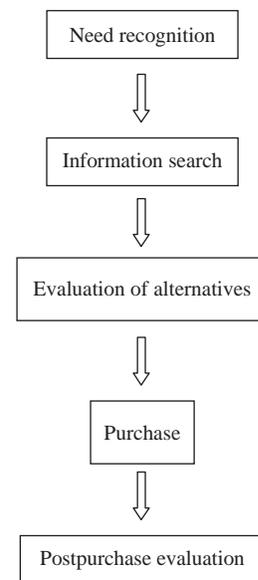


FIGURE 1 Stages in consumer decision making.

of affairs and some desired state. A person who unexpectedly runs out of gas on the highway recognizes a need, as does the person who becomes dissatisfied with the image of his or her car even though there is nothing mechanically wrong with it.

Once a need has been activated, there is a state of tension that drives the consumer to attempt to reduce or eliminate the need. This need may be utilitarian (i.e., a desire to achieve some functional or practical benefit, e.g., when a person loads up on green vegetables for nutritional reasons), or it may be hedonic (i.e., an experiential need involving emotional responses or fantasies, e.g., when a consumer thinks longingly about a juicy steak). Marketers strive to create products and services that will provide the desired benefits and permit the consumer to reduce this tension. This reduction is reinforcing, making it more likely that the consumer will seek the same path the next time the need is recognized.

Maslow's hierarchy of biogenic and psychogenic needs specifies certain levels of motives. This hierarchical approach, shown in Fig. 2, implies that one level must be attained before the next higher one is

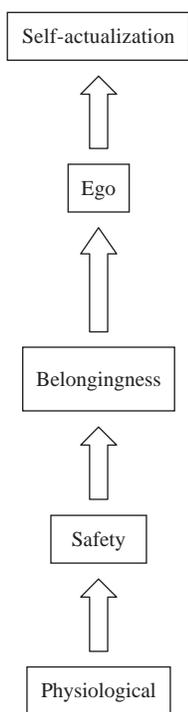


FIGURE 2 Maslow's hierarchy of needs. Adapted from Maslow, A. H. (1970). *Motivation and personality* (2nd ed.). New York: Harper & Row.

activated. Marketers have embraced this perspective because it (indirectly) specifies certain types of product benefits that people might be looking for, depending on the various stages in their development and/or their environmental conditions.

**2.1.2. Information Search**

Need recognition prompts information search, that is, a scan of the environment to identify the options available to satisfy the need. As a rule, purchase decisions that involve extensive search also entail perceived risk, that is, the belief that a poor choice will produce potentially negative consequences. As shown in Table I, perceived risk may be a factor if the product is expensive, complex, and hard to understand or if the consumer believes that the product will not work as promised and/or could pose a safety risk. Alternatively, perceived risk can be present when a product choice is visible to others and the consumer runs the risk of social embarrassment if the wrong choice is made.

**2.1.3. Evaluation of Alternatives**

Information search yields a set of alternative solutions to satisfy the need. Those identified constitute the consumer's evoked set. How does a consumer decide which criteria are important, and how does he or she narrow down product alternatives to an acceptable number and eventually choose one instead of others? The answer varies depending on the decision-making process used. A consumer engaged in extended problem solving may carefully evaluate several brands, whereas someone making a habitual decision might not consider any

TABLE I  
Types of Perceived Risk

Monetary risk	Expensive products
Functional risk	Products requiring time or effort to understand
Physical risk	Mechanical products, medical products, food
Social risk	Socially visible products or those that can influence the impressions made on others (e.g., deodorant)
Psychological risk	Expensive luxuries that may involve guilt or addiction

Source: Solomon, M. R. (2004). *Consumer behavior: Buying, having, and being* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

alternatives to his or her normal brand. Variety seeking, or the desire to choose new alternatives over more familiar ones, can also play a role; consumers at times are willing to trade enjoyment for variety because the unpredictability itself is rewarding.

Evaluative criteria are the dimensions used to judge the merits of competing options. If all brands being considered rate equally well on one attribute (e.g., if all televisions come with remote control), consumers will have to find other reasons to choose one over the others. Determinant attributes are the characteristics actually used to differentiate among choices. For example, consumer research by Church & Dwight Company indicated that many consumers view the use of natural ingredients as a determinant attribute when selecting personal care products. This prompted the firm to develop toothpaste made from baking soda, an ingredient that the company already manufactured for its Arm & Hammer brand.

#### **2.1.4. Purchase**

Once a need has been recognized, a set of feasible options (often competing brands) that will satisfy the need have been identified, and each of these options has been evaluated, the “moment of truth” arrives: The consumer must make a choice and actually procure the product or service. However, other factors at the time of purchase may influence this decision. A consumption situation is defined by factors beyond characteristics of the product that influence a purchase decision. These factors can be behavioral (e.g., entertaining friends) or perceptual (e.g., being depressed, feeling pressed for time).

A consumer’s mood can have a big impact on purchase decisions. For example, stress can impair information-processing and problem-solving abilities. The two dimensions of pleasure and arousal determine whether a shopper will react positively or negatively to a consumption situation. In addition, the act of shopping itself often produces psychological outcomes ranging from frustration to gratification or even exhilaration.

Despite all of their efforts to “pre-sell” consumers through advertising, marketers increasingly recognize that many purchases are strongly influenced by the purchasing environment. Indeed, researchers estimate that shoppers decide on approximately two of every three products while wheeling their carts through supermarket aisles. Dimensions of the physical environment, such as decor, ambient sounds or music, and even

temperature, can influence consumption significantly. One study even found that pumping in certain odors in a Las Vegas casino actually increased the amount of money that patrons fed into slot machines. Time is another important situational variable. Common sense dictates that more careful information search and deliberation occurs when consumers have the luxury of taking their time.

#### **2.1.5. Postpurchase Evaluation and Satisfaction**

Consumer satisfaction/dissatisfaction (CS/D) refers to the attitude that a person has about a product after it has been purchased. This attitude, in turn, is an important determinant of whether the item will be bought again in the future. Despite evidence that customer satisfaction is steadily declining in many industries, marketers are constantly on the lookout for sources of dissatisfaction. For example, United Airlines’ advertising agency set out to identify specific aspects of air travel that contributed to discontent during the travel experience. The agency gave frequent fliers crayons and a map showing different stages in a long-distance trip and asked passengers to fill in colors using hot hues to symbolize areas causing stress and anger and using cool colors for parts of the trip associated with satisfaction and calm feelings. Although jet cabins tended to be filled in with a serene aqua color, ticket counters were colored orange and terminal waiting areas were colored fire red. This research led the airline to focus more on overall operations instead of just on in-flight experiences, and the “United Rising” advertising campaign was born.

Satisfaction is not determined solely by the actual performance quality of a product or service. It is also influenced by prior expectations regarding the level of quality. According to the expectancy disconfirmation model, consumers form beliefs about product performance based on prior experience with the product and/or communications about the product that imply a certain level of quality. When something performs the way in which consumers thought it would, they might not think much about it. If, on the other hand, the product fails to live up to expectations (even if those expectations are unrealistic), negative affect may result. If performance happens to exceed their expectations, consumers are satisfied and pleased. This explains why companies sometimes try to “under-promise” what they can actually deliver.

## 2.2. Biases in the Decision-Making Process

Although the rational model of decision making is compelling, many researchers now recognize that decision makers actually possess a repertoire of strategies—and not all of these strategies are necessarily rational. The constructive processing perspective argues that a consumer evaluates the effort required to make a particular choice and then chooses a strategy best suited to the level of effort required.

As shown in Table II, some purchases are made under conditions of low involvement, where the consumer is not willing to invest a lot of cognitive effort. Instead, the consumer's decision is a learned response to environmental cues, for example, when he or she impulsively decides to buy something that is promoted as a "surprise special" in a store. In other cases, the consumer is highly involved in a decision, and again the stages of rational information processing might not capture the process. For example, the traditional approach is hard-pressed to explain a person's choice of art, music, or even a spouse. In these cases, no single quality may be the determining factor. Instead, an experiential perspective stresses the Gestalt, or totality, of the product or service.

Consumption at the low end of involvement typically is characterized by inertia, where decisions are made out of habit because the consumer lacks the motivation to consider alternatives. Many people tend to buy the same brand nearly every time they go to the store. A competitor who is trying to change a buying pattern based on inertia often can do so rather easily because little resistance to brand switching will be encountered if the right incentive is offered.

At the high end of involvement, one can expect to find the type of passionate intensity that is reserved for people and objects that carry great meaning for the individual. When consumers are truly involved with a product, an ad, or a Web site, they enter a flow state. Flow is an optimal experience characterized by a sense of playfulness, a feeling of being in control, highly focused attention, and a distorted sense of time.

Especially when limited problem solving occurs prior to making a choice, consumers often fall back on heuristics, that is, mental rules-of-thumb that lead to a speedy decision. These rules range from the very general (e.g., "Higher priced products are higher quality products," "Buy the same brand I bought last time") to the very specific (e.g., "Buy Domino, the brand of sugar my mother always bought").

One frequently used shortcut is the tendency to infer hidden dimensions of products from observable attributes. These are known as product signals. Country of origin is an example of a commonly used product signal. In some cases, people may assume that a product made overseas is of better quality (e.g., cameras, cars), whereas in other cases, the knowledge that a product has been imported tends to lower perceptions of product quality (e.g., apparel). Price is also a heuristic; all things equal, people often assume that "Brand A" is of higher quality simply because it costs more than "Brand B."

A well-known brand also frequently functions as a heuristic. People form preferences for a favorite brand and then literally might never change their minds in the course of a lifetime. In contrast to inertia, brand loyalty is a form of repeat purchasing behavior reflecting a conscious decision to continue buying the same brand. Purchase decisions based on brand loyalty also become habitual over time, although in these cases the underlying commitment to the product is much more firm. Because of the emotional bonds that can come about between brand-loyal consumers and products, "true blue" users react more vehemently when these products are altered, redesigned, or eliminated. For example, when Coca-Cola replaced its tried-and-true

TABLE II  
Limited Versus Extended Problem Solving

Motivation	Low risk	High risk
Information search	Low search	High search
Evaluation of alternatives	Weak beliefs Few differences perceived among alternatives	Strong beliefs Significant differences perceived among alternatives
	Most prominent criteria or heuristics used	Many criteria used
Purchase	Limited time spent	Extensive time spent
	Few stores shopped	Many stores shopped
	Store displays influential	Advertising and store personnel influential

Source: Solomon, M. R. (2004). *Consumer behavior: Buying, having, and being* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

formula with New Coke during the 1980s, the company encountered a firestorm of national call-in campaigns, boycotts, and other protests.

Finally, many of people's reactions to products are based on aesthetic responses to colors, shapes, and objects. Many of these preferences are deep-seated or culturally determined. Package designs often incorporate extensive research regarding consumers' interpretations of the meanings accorded to symbols on the box or can. These meanings may be subtle or blatant, but they can exert a powerful effect on expectations about the product within the box or can. In one study, respondents rated the taste of a beer as heavier and more robust when it was served in a brown glass bottle than when the same product was dispensed in a clear bottle.

### 2.3. "Irrational" Decision Making

Psychologists have identified many cognitive mechanisms that interfere with "objective" information processing and decision making. Other researchers have gone a step further in their focus on domains of consumer behavior that cannot be readily explained by a cognitive perspective or where an individual's actions are actually irrational and perhaps even dysfunctional. Indeed, some of consumers' buying behaviors do not seem rational at all because they do not serve a logical purpose (e.g., collectors who pay large sums of money for paraphernalia formerly owned by rock stars). Other actions occur with virtually no advance planning at all (e.g., impulsively grabbing a tempting candy bar from the rack while waiting to pay for groceries).

Still other consumer behaviors, such as excessive eating, excessive drinking, and cigarette smoking, are actually harmful to the individual. These actions may be facilitated by many psychological factors, including the desire to conform to the expectations of others, learned responses to environmental cues, and observational learning prompted by exposure to media. The cultural emphasis on wealth as an indicator of self-worth may encourage activities such as shoplifting and insurance fraud. Exposure to unattainable ideals of beauty and success in advertising can create dissatisfaction with the self, sometimes resulting in eating disorders or self-mutilation.

Consumer addiction is a physiological and/or psychological dependence on products or services. Although most people equate addiction with drugs, virtually any product or service can be the focus of psychological dependence; for example, there is even a support group for Chap Stick addicts. Some psychologists are

now voicing strong concerns about "Internet addiction," a condition whereby Web surfers become obsessed by online chat rooms to the point where their "virtual" lives take priority over their real ones. Compulsive consumption refers to repetitive excessive shopping that serves as an antidote to tension, anxiety, depression, or boredom.

## 3. PSYCHOLOGICAL INFLUENCES ON CONSUMPTION

The business process of market segmentation identifies groups of consumers who are similar to one another in one or more ways and then devises marketing strategies that appeal to the needs of one or more of these groups. One very common way in which to segment consumers is along demographic dimensions such as the following:

- *Age.* People who belong to the same age group tend to share a set of values and common cultural experiences that they carry throughout life.
- *Gender.* Many products are sex typed, and consumers often associate them with one gender or the other. Marketers typically develop a product to appeal to one gender or the other. For example, Crest's Rejuvenating Effects toothpaste, made specifically for women, is packaged in a teal tube nestled inside a glimmering "pearlescent" box.
- *Social class and income.* Working-class consumers tend to evaluate products in more utilitarian terms such as sturdiness and comfort. They are less likely to experiment with new products or styles such as modern furniture and colored appliances. Higher classes tend to focus on more long-term goals such as saving for college tuition and retirement.
- *Family structure.* People's family and marital status influences their spending priorities. For example, young bachelors and newlyweds are the most likely to exercise, consume alcohol, and go to bars, concerts, and movies.
- *Race and ethnicity.* Ethnic minorities in the United States spend more than \$600 billion per year on products and services, so firms must devise products and communications strategies tailored to the needs of these subcultures. Sometimes these differences are subtle yet important. When Coffee-Mate discovered that African Americans tend to drink their coffee with sugar and cream more so than do Caucasians, the company mounted a promotional blitz in black-oriented media that resulted in double-digit increases in sales volume.

- *Geography.* Place of residence influences preferences within many product categories, from entertainment to favorite cars, decorating styles, or leisure activities. For example, BMW found that drivers in France prized its cars for their road-handling abilities and the self-confidence this gave them, whereas drivers in Austria were more interested in the status aspect of the BMW brand.

### 3.1. Psychographics and Lifestyles

Although these segmentation variables are very important, consumers can share the same demographic characteristics and still be very different people. Psychographics are data about people's attitudes, interests, and opinions (AIOs) that allow marketers to cluster consumers into similar groups based on lifestyles and shared personality traits.

Lifestyle refers to a pattern of consumption reflecting a person's choices of how he or she spends time and money. In an economic sense, a person's lifestyle represents the way in which he or she has elected to allocate income in terms of relative allocations to various products and services and to specific alternatives within these categories. Lifestyle, however, is more than the allocation of discretionary income; it is a statement about who a person is in society and who the person is not. Group identities, whether of hobbyists, athletes, or drug users, gel around forms of expressive symbolism.

### 3.2. Personality Theory Applications to Consumer Behavior

A consumer's personality, or unique psychological makeup, may influence the products and marketing messages that he or she prefers. Consumer psychologists have adapted insights from major personality theorists to explain people's consumption choices.

#### 3.2.1. Freudian Theory

Freudian psychology exerted a significant influence on applied consumer research, especially during the early days of the discipline. Freud's writings highlight the potential importance of unconscious motives underlying purchases. This perspective also hints at the possibility that the ego relies on the symbolism in products to compromise between the demands of the id and the prohibitions of the superego. During the 1950s, motivational researchers attempted to apply Freudian ideas

to understand the deeper meanings of products and advertisements. For example, for many years, Esso (now Exxon) reminded consumers to "Put a Tiger in Your Tank" after researchers found that people responded well to this powerful animal symbolism containing vaguely sexual undertones.

#### 3.2.2. Jungian Theory

Freud's disciple, Jung, introduced the concept of the collective unconscious, that is, a storehouse of memories inherited from a person's ancestral past. These shared memories create archetypes, that is, universally shared ideas and behavior patterns involving themes such as birth, death, and the devil, that frequently appear in myths, stories, and dreams. For example, some of the archetypes identified by Jung and his followers include the "old wise man" and the "earth mother," and these images appear frequently in marketing messages that use characters such as wizards, revered teachers, and even Mother Nature.

#### 3.2.3. Trait Theory

Trait theory focuses on the quantitative measurement of personality traits, that is, identifiable characteristics that define a person. Some specific traits relevant to consumer behavior include innovativeness (i.e., the degree to which a person likes to try new things), materialism (i.e., the amount of emphasis placed on acquiring and owning products), self-consciousness (i.e., the degree to which a person deliberately monitors and controls the image of the self that is projected to others), need for cognition (i.e., the degree to which a person likes to think about things and, by extension, expend the necessary effort to process brand information), and self-monitoring (i.e., the degree to which a person is concerned with the impression that his or her behaviors make on others).

## 4. CULTURAL AND INTERPERSONAL INFLUENCES ON CONSUMPTION

Values are very general ideas about good and bad goals, and these priorities typically are culturally determined. From these flow norms, that is, rules dictating what is right or wrong and what is acceptable or unacceptable. Consumers purchase many products and services because they believe that these products will help them

to attain value-related goals. For example, an emphasis on personal hygiene in Japan has created a demand for products such as automated teller machines (ATMs) that literally “launder” money by sanitizing yen before dispensing them to bank customers.

#### 4.1. Subcultures and Reference Groups

Members of a subculture share beliefs and common experiences that set them apart from others in the larger culture. Whether “Dead Heads” or “Skinheads,” each group exhibits its own unique set of norms, vocabulary, and product insignias (e.g., the skull and roses that signifies the Grateful Dead subculture).

A reference group is an actual or imaginary individual or group that influences an individual’s evaluations, aspirations, or behavior. As a rule, reference group effects are more robust for purchases that are (a) luxuries (e.g., sailboats) rather than necessities and (b) socially conspicuous or visible to others (e.g., living room furniture, clothing).

#### 4.2. Opinion Leaders

An opinion leader is a person who is frequently able to influence others’ attitudes or behaviors. Opinion leaders are extremely valuable information sources because (a) they are technically competent and, thus, more credible; (b) they have prescreened, evaluated, and synthesized product information in an unbiased way; and (c) they tend to be socially active and highly interconnected in their communities.

#### 4.3. Diffusion of Innovations

Diffusion of innovations refers to the process whereby a new product, service, or idea spreads through a population. If an innovation is successful (most are not), it typically is initially bought or used by only a few people. Then, more and more consumers decide to adopt it until it might seem (sometimes) that nearly everyone has bought or tried the innovation.

A consumer’s adoption of an innovation resembles the standard decision-making sequence whereby he or she moves through the stages of awareness, information search, evaluation, trial, and adoption. The relative importance of each stage may differ depending on how much is already known about the innovation as well as on cultural factors that may affect people’s willingness

to try new things. However, even within the same culture, not all people adopt an innovation at the same rate. Some do so quite rapidly, whereas others never do at all.

Consumers can be placed into approximate categories based on their likelihood of adopting an innovation. Roughly one-sixth of the people are very quick to adopt new products (i.e., innovators and early adopters), and one-sixth of the people are very slow (i.e., laggards). The other two-thirds are somewhere in the middle (i.e., late adopters). These latter consumers are the mainstream public; they are interested in new things, but they do not want them to be too new.

Even though innovators represent only approximately 2.5% of the population, marketers are always interested in identifying them. Innovators are the brave souls who are always on the lookout for novel developments and who will be the first to try new offerings. They tend to have more favorable attitudes toward taking risks, have higher educational and income levels, and be socially active. In addition, many innovators also are opinion leaders, so their acceptance of an innovation may be a crucial factor in persuading others to try it as well.

As a rule, consumers are less likely to adapt innovations that demand radical behavior changes—unless they are convinced that the effort will be worthwhile. As a result, evolutionary changes (e.g., a cinnamon version of Quaker oatmeal) are more likely to be rapidly adapted than are revolutionary changes (e.g., ready-to-eat microwaveable Quaker oatmeal). The following factors make it more likely that consumers will accept an innovation:

- Compatibility with current lifestyle
- Ability to try the product before buying
- Simplicity of use
- Ease of observing others using the innovation
- Relative advantage over benefits offered by other alternatives

#### See Also the Following Articles

Advertising and Culture ■ Advertising Psychology  
 ■ Economic Behavior ■ Values and Culture

#### Further Reading

- Alba, J. W., & Hutchinson, J. W. (1988). Dimensions of consumer expertise. *Journal of Consumer Research*, 13, 411–454.
- Baumgartner, H. (2002). Toward a personology of the consumer. *Journal of Consumer Research*, 29, 286–292.
- Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research*, 15, 139–168.

- Belk, R. W., Wallendorf, M., & Sherry, J. F., Jr. (1989). The sacred and the profane in consumer behavior: Theodicy on the odyssey. *Journal of Consumer Research*, 16, 1–38.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1988). Constructive consumer choice processes. *Journal of Consumer Research*, 25, 187–217.
- Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, 7, 234–248.
- Cialdini, R. B. (1993). *Influence: The psychology of persuasion*. New York: William Morrow.
- Dichter, E. (1964). *The handbook of consumer motivations*. New York: McGraw-Hill.
- Fournier, S. (1998). Consumers and their brands: Developing relationship theory in consumer research. *Journal of Consumer Research*, 24, 343–373.
- Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing*, 60, 50–68.
- Hofstede, G. (1980). *Culture's consequences*. Beverly Hills, CA: Sage.
- Holbrook, M. B., & Hirschman, E. C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of Consumer Research*, 9, 132–140.
- Jacoby, J., & Chestnut, R. (1978). *Brand loyalty: Measurement and management*. New York: John Wiley.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Maslow, A. H. (1970). *Motivation and personality* (2nd ed.). New York: Harper & Row.
- Midgley, D. F. (1983). Patterns of interpersonal information seeking for the purchase of a symbolic product. *Journal of Marketing Research*, 20, 74–83.
- Oliver, R. L. (1996). *Satisfaction*. New York: McGraw-Hill.
- Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research*, 10, 135–146.
- Ratner, R. K., Kahn, B. E., & Kahneman, D. (1999). Choosing less-preferred experiences for the sake of variety. *Journal of Consumer Research*, 26, 1–15.
- Robertson, T. S., & Kassirjian, H. H. (Eds.). (1991). *Handbook of consumer behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). New York: Free Press.
- Rokeach, M. (1973). *The nature of human values*. New York: Free Press.
- Sirgy, M. J. (1982). Self-concept in consumer behavior: A critical review. *Journal of Consumer Research*, 9, 287–300.
- Solomon, M. R. (1983). The role of products as social stimuli: A symbolic interactionism perspective. *Journal of Consumer Research*, 10, 319–329.
- Solomon, M. R. (2004). *Consumer behavior: Buying, having, and being* (6th ed.). Upper Saddle River, NJ: Prentice Hall.
- Snyder, M. (1979). Self-monitoring processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 85–128). New York: Academic Press.
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4, 199–214.
- Wells, W. D., & Tigert, D. J. (1971). Activities, interests, and opinions. *Journal of Advertising Research*, 11, 27.
- Wicklund, R. A., & Gollwitzer, P. M. (1982). *Symbolic self-completion*. Hillsdale, NJ: Lawrence Erlbaum.
- Wright, P. (2002). Marketplace metacognition and social intelligence. *Journal of Consumer Research*, 28, 677–682.
- Zablocki, B. D., & Kanter, R. M. (1976). The differentiation of life-styles. *Annual Review of Sociology*, 2, 269–297.



# Contracts

René Schalk

Tilburg University, Tilburg, The Netherlands

1. Introduction
  2. What Is a Contract?
  3. Different Types of Contracts
  4. Contracts in Employment
  5. Consequences of Violations of Contracts
  6. Cross-National Differences
  7. Implications
- Further Reading

## GLOSSARY

*obligation* A commitment to some future action.

*relational relationship* Focusing on open-ended relationships involving considerable investments of all parties.

*transactional relationship* Focusing on short-term and monetizable exchanges.

*violation* A failure to comply with the terms of the contract.

*zone of negotiability* Room at the level of the individual party for negotiation about contract terms.

A contract is an exchange agreement between parties. In employment, psychological contracts are drivers of employee behavior with a strong influence on organizational outcomes.

## 1. INTRODUCTION

Contracts are the cornerstones of our society. Contracts about exchanges are a common feature of

societies all over the world. We seem not to be able to do without contracts. In our private lives, we have marriage contracts and other forms of contractual arrangements that people make when they decide to live together. In our social lives, we have contracts in the form of sports club memberships, political party memberships, and the like. We make contractual exchanges when buying houses, booking holidays, paying for insurance, and so on. At work, we have labor contracts. Contracts play a very prominent role in our daily lives. One could say that, in a way, we are nearly permanently occupied with fulfilling contract terms. Therefore, it is inevitable that we will sometimes be confronted with unfulfilled or broken contracts by ourselves or by other parties.

## 2. WHAT IS A CONTRACT?

The belief of obligations existing between two or more parties is the basis for a contract. A contract is an exchange agreement between two (or more) parties. The contract may include written terms, orally communicated terms, or other expressions of commitment and future intent. As generally defined, a contract is based on a mutually agreed on, voluntarily made exchange of promises between two (or more) parties, where each party gains if the exchange agreement is fulfilled and each party loses if it is not.

### 3. DIFFERENT TYPES OF CONTRACTS

Contracts can be viewed from various angles. The economic focus is on transaction or relationship costs of the contractual exchange relationships. The legal focus is on the formalized terms and procedures and on issues related to contract fulfillment. The term “social contract” refers to a pact or an agreement made by all of the individuals who are to compose it.

The term “psychological contract” refers to ideas about mutual obligations in the exchange agreement. One view on the psychological contract is that it includes the perceptions of both parties. Another view limits the psychological contract to individual beliefs about the exchange agreement. It includes and adds to the economic and legal perspectives. The social contract is based on generalized psychological contracts.

Psychological contracts address the important issue of incompleteness of contracts. It is virtually impossible to create complete formal contracts. Unforeseen events may arise over time. Foreseen events cannot always be described unambiguously, and the costs of negotiating all possible events and consequences might be too high. Psychological contracts are generally understood agreements that fill in the gaps of the formal agreement.

A general distinction can be made between transactional contracts, which have the characteristics of an “arm’s length” market transaction (e.g., between buyer and seller), and relational contracts such as complex long-term relationships between employer and employee.

### 4. CONTRACTS IN EMPLOYMENT

A basic feature of employee–employer interaction is the nature of the exchange in the employment relationship. An employment relationship is a contract based on a mutually agreed on, voluntarily made exchange of promises between two (or more) parties, where each party gains if the exchange agreement is fulfilled and each party loses if it is not. Each component can vary in response to the social context in which employment is embedded.

A meeting of the minds of employer and employee is implied in the term “mutuality.” Whether there is actually a moment when parties share a crystal clear understanding regarding the terms of their exchange is open to question. Nevertheless, at the start of the employment relationship, there is at least a perception of agreement if not necessarily agreement in fact.

Both employer and employee have free choice in entering into and accepting the terms of an employment arrangement. However, this is always a matter of degree. Voluntariness in employment is historically rooted in the evolution of institutions such as individual rights, ownership of private property, and collective bargaining. Voluntariness in employment is possible when individuals have the right to control their own time, services, and use of their skills. Employer beliefs regarding worker freedom of choice (e.g., job opportunities elsewhere) influence the inducements that employers offer in exchange for worker contributions.

### 5. CONSEQUENCES OF VIOLATIONS OF CONTRACTS

Fulfilling contract terms by the organization leads to greater employee commitment to the organization and to the relationship. Consequently, readiness to accept work roles and tasks, willingness to engage in extra-role behaviors, and willingness to avoid negative behaviors (e.g., coming to work late, not performing well) will be greater. This will result in more committed, motivated, and trusting organizational citizens.

More favorable psychological contracts are related to a higher degree of organizational commitment and a lower degree of intention to leave the organization. On the other hand, when employees believe that the organization has not fulfilled its side of the bargain, they are less likely to engage in extra-role behaviors (e.g., organizational citizenship). Employees who believe that the organization has violated certain terms of the psychological contract will be more likely to have high rates of tardiness and absenteeism and a higher intention to leave the organization. Anger, disappointment, and strong emotions may also be consequences of not fulfilling contract terms.

### 6. CROSS-NATIONAL DIFFERENCES

Who the parties to the employment agreement are varies considerably across countries, as does the actual level at which the exchange agreement is created. Societies ascribe varying importance to the parties of an employment contract. The status and esteem of government and the state per se differ considerably. Because parties to employment agreements entail a wide range of societal actors, employment agreements can arise at several

levels: between the individual worker and the firm, between groups of workers and the firm, between groups of workers and groups of firms, and in central agreements involving groups of workers, employers, and the state. An industrial relations system involves three sets of actors: employers, workers (typically in groups either within or between firms), and the government (including direct negotiation involving governmental officials, governmental mediation of employee–employer agreements, or creation of laws and statutes specifying conditions of employment).

Three general types of employment and industrial relations systems are direct exchanges between firm and employee; central agreements between employer organizations and unions (or comparable parties), with possible segregation at various levels (e.g., nationwide, type of industry); and employment relationships anchored in society (i.e., the state and institutions), where regulations and statutes predominate in the construction of the employee–employer exchange.

Direct firm–employee exchanges are more prevalent in countries such as Israel, New Zealand, Singapore, and the United States (except in the case of unionized firms). Central agreements play an important role in countries such as Australia, Belgium, India, Mexico, The Netherlands, and Sweden. Regulations play a significant role in countries such as France and Japan.

One striking difference across societies is in the zone of negotiability of the individual employment relationship. Factors shaping it include the individual's personal power in constructing an idiosyncratic arrangement with the employer, the employer's willingness to demand or offer unique or varied employment conditions to individual workers, and the society's regulation of employment and tolerance for unequal outcomes. State intervention or central agreements have the most salient effect on the range and boundaries for what is negotiable between employers and workers. State-level interventions have several effects. Regulations constrain what conditions are negotiable and mandate others. Central agreements, negotiated collectively, limit individual-level variability in contract terms. Both appear to be more prevalent in societies that are relatively intolerant of outcome inequality or where there are institutional concerns regarding disparities in power between labor, on the one hand, and management and owners, on the other. Laws stipulate which conditions are non-negotiable, for example, requiring payment of retirement or unemployment taxes and providing extended family leave benefits following childbirth. Laws also serve to enforce negotiated terms of employment.

Nonlegal factors can also constrain the zone of negotiability. Cultural norms, in particular, can effectively remove certain employment terms from the bargaining table. In societies with strong state influence or central agreements, the focus of negotiation and flexibility shifts to what is possible within or around the rules. There tends to be a relatively clear demarcation of which issues are on the negotiation table and which ones are not. Yet even when the wage system has been standardized with strict rules for salary ranges for certain jobs, there is often room at the level of the individual worker for negotiation about salary within the given range or about the title of a particular job.

## 7. IMPLICATIONS

Contractual agreements play a very prominent role in our society and in the world of work. Contractual exchange agreements at work determine employee behavior and organizational outcomes.

In the contracting process clear communication and careful assessment of the contract terms is necessary; constant monitoring and renegotiating, adjusting to changing needs of the organization and the individual employee.

In the future, we can expect considerable innovation and experimentation in contracting. Critical questions will center on how to balance differentiation and fairness, especially in situations where employees are interdependent, and how to integrate groups and promote cooperation while keeping clear boundaries for individual psychological contracts. A greater variety of contract types, including more “individualized” results of contract negotiations, will exist. Contracts made today are a way in which to both know and shape the future. The management of psychological contracts is a core task for firms that attempt to develop “people-building” rather than “people-using” organizations in an organizational climate characterized by trust. Increasingly, managing the psychological contract is a core task for workers themselves, who seek to meet their own needs in active individual- and group-level negotiations with their employers.

### See Also the Following Articles

Job Analysis, Design, and Evaluation ■ Occupational Choice  
■ Occupational Psychology, Overview ■ Part-Time Work

### Further Reading

- Anderson, N., & Schalk, R. (1998). The psychological contract in retrospect and prospect. *Journal of Organizational Behavior*, *19*, 637–648.
- Herriot, P., & Pemberton, C. (1995). *New deals: The revolution in managerial careers*. Chichester, UK: Wiley.
- Rousseau, D. M. (1995). *Psychological contracts in organizations: Understanding written and unwritten agreements*. Thousand Oaks, CA: Sage.
- Rousseau, D. M., & Schalk, R. (Eds.). (2000). *Psychological contracts in employment: Cross-national perspectives*. Thousand Oaks, CA: Sage.
- Schalk, R., & Rousseau, D. M. (2001). Psychological contracts in employment. In N. Anderson, D. S. Ones, H. Kepir Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology*, Vol. 2: *Organizational psychology*. Thousand Oaks, CA: Sage.



# Cooperation at Work

Eduardo Salas, Dana E. Sims, and Cameron Klein

University of Central Florida, Orlando, Florida, USA

1. Importance of Teamwork
  2. Teamwork Examined
  3. Teamwork Derailers
  4. Facilitators of Teamwork
  5. Promoting Cooperation and Teamwork at Work through Training
  6. Measuring Team Performance and Effectiveness
  7. Conclusion
- Further Reading

## GLOSSARY

**adaptability** The ability to gather information from the task environment and use it to adjust the team's strategy through the use of compensatory behavior and the reallocation of interteam resources.

**cooperation** Includes offering help to only those team members who need it, pacing activities to fit the needs of the team, and behaving in an unambiguous manner so that actions are not misinterpreted.

**cross-training** A training strategy in which teammates develop an understanding for the tasks, mission, objectives, and responsibilities of the team; this strategy targets team members' interpositional knowledge and shared mental models for development with resulting gains in team coordination.

**mutual performance monitoring** Monitoring fellow team members' performance while maintaining individual responsibilities to improve overall group performance.

**shared mental models (SMMs)** Shared understandings or representations of the goals of the team, individual team member tasks, and how team members will coordinate to achieve their common goals; individual team members can

have a varying degree of overlap or "sharedness" among their mental model of the team.

**team** Collective of interdependent individuals who have shared habits of thinking, mental models, and paradigms that guide their perceptions and behaviors toward a common goal.

**team building** Targets role clarification, goal setting, problem solving, and/or interpersonal relations for improvement.

**team coordination training** Focuses on teaching team members about basic processes underlying teamwork; this strategy targets mutual performance monitoring and back-up behavior as well as other critical teamwork competencies and skills for improvement.

**team orientation** The preference for working collectively and the tendency to enhance individual performance by using task inputs from other group members during group tasks.

**team self-correction training** Team members are taught techniques for monitoring and then categorizing their own behaviors; this process generates instructive feedback so that team members can review performance episodes and correct deficiencies.

**team structure** Refers to the power structure (e.g., hierarchical), the number of team members, and the diversity among them.

Teamwork is a set of flexible behaviors, cognitions, and attitudes that interact to achieve desired mutual goals and adaptation to changing internal and external environments. It consists of the knowledge, skills, and attitudes (KSAs) that are displayed in support of one's teammates and the team's objectives. Essentially, teamwork is a set of interrelated thoughts, actions, and feelings that combine to facilitate coordinated adaptive

performance and the completion of taskwork objectives. When teamwork KSAs are mastered and displayed, teams can achieve a level of performance and effectiveness that is greater than the total efforts of all individual team members.

## 1. IMPORTANCE OF TEAMWORK

The nature of work is changing. Recent technological advances, a shift from manufacturing to service-based organizations, increased global competition, and the importance of knowledge workers all have contributed to a dynamic and complex work environment. To survive, organizations must embrace flexibility and adaptability. They must have a systems view that helps them to integrate the meaning of common everyday occurrences into overall interrelationships, interdependencies, and patterns of change to allow them to achieve this needed adaptability. Very often, organizations accomplish these goals through the implementation of teams. It is in direct response to increasingly severe time pressures, technological advancements, and other elements of highly complex environments that teamwork has become an established stronghold and a source of competitive advantage for today's organizations. Teams are also called on when the task demands are such that the capabilities of one individual are exceeded or the consequences of error are high. Therefore, the utility of teams and teamwork has spanned industry to include aviation, nuclear power, and the military.

A team is a collective of interdependent individuals who together have shared objectives, mental models, and procedures that guide their perceptions, thinking, and behaviors toward a common goal. The process in which teams achieve this commonality is called collaboration. Similarly, cooperation has been discussed as a team skill competency that includes offering help to those team members who need it, pacing activities to fit the needs of the team, and behaving so that actions are not misinterpreted. Collective organizations immerse themselves in the virtues of collaboration and cooperation by reducing the hierarchy of social stratification on which traditional bureaucratic organizations rely. Furthermore, they encourage decisions to be made by the collective group to foster a sense of community and shared purpose. Cooperation and teamwork facilitate coordination, communication, adaptability, enhanced employee participation, and empowerment, thereby allowing individuals to achieve collective outputs that are greater than the sum of their parts.

Organizations continue to depend on teams and the synergy they engender to assist in streamlining work processes that promote efficiency, increased innovation, and quality products and services.

This article outlines the nature of cooperation and teamwork as it is manifested in organizations. It examines what derails teamwork and what facilitates it as well as how to promote it through training. The article concludes with a discussion of the need to measure training effectiveness and offers a brief discussion on how this is done. Because teams and teamwork will likely continue to play a vital role in organizations into the future, it is a worthy endeavor to examine their nature.

## 2. TEAMWORK EXAMINED

Despite extensive writing on the topic, a clear definition and description of teamwork continues to be elusive. In general, teamwork is more than just the summated task or task-related accomplishments of the individuals who make up a group. It is the interaction of team-related knowledge, skills, and attitudes (KSAs) of all team members—and the ability to capitalize on this pool of resources—that maximizes the potential effectiveness of the group to outperform even the most competent individual member of the team. The definition of teamwork provided here is a synthesis of previous definitions that have been advanced by established team researchers. Therefore, teamwork is defined as a set of flexible behaviors, cognitions, and attitudes that interact to facilitate taskwork and achieve mutually desired goals and adaptation to the changing internal and external environments. Although many behaviors, cognitions, and attitudes have been presented in the teamwork literature, this article presents and briefly describes only a subset of those “components of teamwork” previously identified as critically important.

Team members may exhibit several behaviors that comprise teamwork: mutual performance monitoring, backup behavior, adaptability, closed-loop communication, and leadership. Mutual performance monitoring involves monitoring fellow team members' performance while still maintaining individual responsibilities. The intent is to improve overall group performance by detecting errors, deficiencies, and overloads that inhibit the team's actions. As team members monitor their teammates' performance and are able to detect deficiencies or overloads, they are able to shift work responsibilities to other team members or provide feedback about errors if and when it becomes necessary. This is referred to as

backup behavior. Taken together, mutual performance monitoring and backup behavior result in a team that is adaptable and flexible. Adaptability has been described as the capacity to recognize deviations from expected action and to readjust those actions accordingly.

Closed-loop communication also facilitates teamwork by ensuring that accurate information is given and understood by team members. It involves the sender initiating a message, the receiver accepting and acknowledging the message, and the sender ensuring that the intended message was received. When communication fails, implicit coordination and the shared understanding of the team's task may deteriorate. Finally, effective team leaders facilitate teamwork by consistently encouraging mutual performance monitoring, backup behavior, adaptability, and communication. The team leader's failure to guide the team in these behaviors may inhibit coordinated and adaptive action and can be a key factor in deficient team performance. However, the behaviors described previously are only a part of the overall picture of teamwork. There are also cognitive and attitudinal components of teamwork that facilitate these behaviors, and a few of them are briefly described in what follows.

Shared mental models have been a frequently discussed cognitive component of teamwork. In 1995, Cannon-Bowers and colleagues suggested that shared mental models are shared understandings or representations of the goals of the team, individual team member tasks, and how the team members will coordinate to achieve their common goals. Similarly, familiarity of task-specific teammate characteristics is also an important component of teamwork. Familiarity includes knowledge of the task-related competencies, preferences, tendencies, and strengths and weaknesses of teammates. Knowledge of teammates and their weaknesses can have a profound influence on team communication and interaction patterns. In fact, research has demonstrated a moderate relationship between high levels of familiarity in teams and both productivity and decision-making effectiveness.

The primary attitudinal component of teamwork is team orientation. Team orientation is frequently described as a preference for working with others. More specifically, however, it is a tendency to enhance individual performance by placing value on group input and using input from other group members while performing a group task. When discrepancies occur among team members' levels of understanding, those with a team orientation have been found to spend more time comparing their comprehension of the situation with that of their teammates to resolve the discrepancy. This provides greater opportunity to reevaluate their shared

understanding of a task and to assist in error detection. Team members who do not have a team orientation are more likely to reject the validity of input from other team members, and this can potentially lead to a decline in team performance. Not all employees see value in working with others, and those who do not are less likely to succeed on teams or to be as cooperative as others. Research has shown that better decisions and higher performance result from teams that contain team-oriented individuals.

Now that the primary cognitive, behavioral, and attitudinal components of teamwork have been examined, the next section presents some concrete explanations for why teams might go off-track or "derail." Although other factors can lead to deficient teamwork (e.g., individual differences, organizational climate, stress), the section focuses specifically on factors that can be influenced through team training and team building.

### 3. TEAMWORK DERAILERS

Teamwork does not "just happen." Often, an organization will place a group of well-qualified individuals together to work on a project and is surprised when the team does not achieve greatness. Despite the many benefits of teamwork and cooperation at work, teamwork does not occur automatically. But why do teams derail, and how does teamwork erode? Team derailment is defined as the process whereby an effective team experiences a significant decline in performance. This section, from the vantage point of a team competency framework, identifies three primary culprits: high levels of stress; lack of knowledge of the team's mission, objectives, norms, and resources; and dissolution of trust. Any of these may lead to a decrement in team performance and cooperation.

Stress can have a significant effect on an individual's or a team's ability to perform a job. For example, teams operating in the military often face a high-stress environment in which multiple tasks must be performed under time pressure and in ambiguous situations. However, stress can result from situations in which resources are stretched or exceeded, there is danger of physical or personal harm, or the capacity to perform is diminished. Outcomes of stress for individuals in teams can be physiological (e.g., increased heartbeat, sweating), emotional (e.g., fear or anxiety), social (e.g., difficulty in coordinating actions), cognitive (e.g., a narrowing of attention), or have a direct influence on performance (e.g., trade-offs in the speed and accuracy of task performance, impaired decision making). None of these

outcomes is particularly pleasant, and all can potentially impair team performance. One solution for teams operating in high-stress environments may be to provide stress exposure training (SET), which is designed to reduce stress through the provision of information, skills training, and practice to trainees.

To avoid team derailment, team members also need to know their team's strategic and tactical goals, what resources are available to help achieve those goals, and under what conditions the team members will be expected to perform. When this knowledge is lacking, coordination of team member actions and provision of backup behavior and support are also deficient. It has been suggested that when a team initially forms, team members should be given the opportunity to gauge each other's abilities and skills and to cooperatively determine the purpose of their team. This is often referred to as creating a shared mental model of the team, the task, and the processes by which the team will complete the task. Having completed this, the team can then determine how it will accomplish its goals and establish norms of behavior and performance. When a team lacks a clear mental model, it fails to recognize and integrate task contingencies as they arise. Similarly, failure to maintain a shared perception of the situation, strategic and tactical goals, and the resources available to accomplish those goals can lead to decreased cooperation and teamwork.

Finally, trust has become a common construct in the teamwork literature due to its positive relationship to organizational citizenship behaviors, successful negotiation, and conflict management. Understandably, all of these behaviors are essential to cooperation as well. In 1998, Rousseau and colleagues defined trust as a willingness to accept vulnerability due to positive expectations of the intentions or behaviors of others. Mutual trust in a team is a deeply held confidence toward the team's climate, internal environment, and members. It is an attitude held by team members regarding the climate or mood of the team's internal environment. When team members work interdependently to achieve group goals, they must be willing to accept a certain amount of risk so as to rely on each other and openly share information. Lack of trust in this situation may result in a reduction of information sharing as well as extensive time spent in monitoring team members, not for the purpose of improving performance but rather for ensuring their trustworthiness. When organizations fail to create policies that foster trust among team members or do not allow sufficient team interaction, cooperation will deteriorate. Now that some ways in which teams and teamwork may erode have been presented, the next

section describes ways in which teamwork may be enhanced in work organizations.

#### **4. FACILITATORS OF TEAMWORK**

A successful team is infused with an energizing spirit that draws the participants together into a cohesive unit and has everyone pulling together to reach a common goal. In general, teamwork can be facilitated, promoted, or encouraged at either the organizational or team level. Important organizational variables include the amount of organizational support offered and the design of the team (e.g., size, diversity, skill composition). At the team level, effective leadership, closed-loop communication, interdependence, and cohesion all contribute to team effectiveness. Although this is clearly not an exhaustive list of all elements that contribute to team functioning, these variables are instrumental in facilitating effective team performance. Each is briefly discussed in turn.

Organizational support is a primary facilitator of teamwork. It is the organization and managers that must instill cultures that encourage common objectives, shared values, mutual trust, frequent and honest communication, and the ability to act in a unitary manner. Sometimes, organizations maintain cultures that do not promote interaction. Instead, these organizations maintain a culture in which blame is constantly placed and responsibility is shifted elsewhere. When this happens, teamwork cannot be facilitated. Furthermore, it is the responsibility of the team leaders to structure team tasks. This includes delegating and assigning the tasks, managing resources, and establishing priorities. Broadly speaking, team leaders set the direction and provide the resources, whereas team members enact organizational objectives and engage in teamwork in service of their organizations.

The design of the team is another organizational level factor that can act as a facilitator of teamwork. Factors such as the size of the team, the diversity within it, and team member skill composition are important. As such, proper selection of team members is important to facilitating team performance. Research suggests that effective teams are composed of between 6 and 10 members. Larger teams provide individual members with the opportunity to "social loaf" as individual inputs may become uninterpretable and indistinguishable from the group's product or service.

Team diversity is another design issue. Research has demonstrated that the amount of cooperation that occurs in teams can be dependent on the degree of heterogeneity of individuals within it, and this heterogeneity can be

related to any differences such as those in skill, qualifications, gender, and/or ethnicity. For example, accomplishing innovation and creativity goals in product teams is more likely when members are very diverse (i.e., heterogeneous). Conversely, teams whose goals include quick production and decision making exhibit increased levels of teamwork when members are very similar (i.e., homogenous). Typically, however, teams composed of heterogeneous members will take longer to begin working together effectively. Over time, and as team members become more familiar with each other, team performance generally increases for teams composed of diverse members.

Team members must display critical knowledge, skill, and attitude competencies while performing in dynamic environments. Team researchers have suggested that general team skills should be used when selecting team members to ensure team effectiveness. Ideally, individuals possessing both task-specific skills (i.e., those relating directly to the tasks to be performed) and team-contingent skills (e.g., those with a team or collective orientation) would be selected to participate in teams. Unfortunately, this luxury is not always available. So, how do the team-level variables of leadership, closed-loop communication, and shared awareness influence team performance and effectiveness?

Team leadership may be the most important element in creating a cooperative work environment. Leadership affects team effectiveness through many processes (e.g., cognitive, motivational, affective). Team leaders set the tone by articulating a clear and motivating vision and creating a supportive climate that promotes effective team processes and behaviors, which may include specific advanced planning, communication of critical process information, and promotion of functional coordination among team members. Information from all team members must be pooled to develop plans and evaluate the consequences of team decisions. However, team leaders must remain flexible and must adapt plans and team strategies to meet task demands so as to avoid a deterioration of teamwork in dynamic and stressful situations. Team leaders' failure to guide and structure team experiences that would otherwise facilitate the development and maintenance of cooperative interactions can be a key factor in ineffective performance. An important point not frequently discussed in teamwork literature is that team members cooperate and commit to teams for different reasons. Some may be motivated to cooperate due to a desire for results, others may be motivated by the satisfaction of working with others, and still others may have less admirable aims.

Therefore, leaders should not assume that motivations are the same for all team members. However, if handled appropriately, even individualistic team member motivations can contribute to common overall team goals.

Communication is also essential for teamwork. Communication creates a shared understanding of the task at hand and ensures that all members possess the required and precise information needed. This is especially important for teams operating in very complex and dynamic environments (e.g., flight deck crews, surgical teams) because it facilitates teamwork behaviors (e.g., backup behaviors) and shared mental models that allow teams to adapt quickly and implicitly. As mentioned previously, closed-loop communication is a particularly important enabler of effective teamwork. Communiqués presented in the prescribed form, tempo, and cadence are examples of proceduralized closed-loop communication used in the military today.

Finally, teamwork behaviors are facilitated by a sense of cohesion and interdependence among team members. Teamwork involves group members' collectively viewing themselves as a group whose success depends on their interaction. Cohesion can be understood as a shared value, whereby the group interests, responsibilities, and success all are taken into account, even at the expense of individual goals. Similarly, interdependence is also typically seen as an essential characteristic of team performance. Effective teams not only have an awareness of the importance of their role in team functioning but also put this into action by fostering within-team interdependence. Effective teams contain individuals who know one another's jobs, offer to help fellow members when it is appropriate to help, and depend on one another for this help.

Discussions of both the possible derailers and potential facilitators of teamwork have now been provided. However, very little has been said about how to promote these behaviors, cognitions, and attitudes. The next section describes ways in which cooperation can be promoted at work through training.

## **5. PROMOTING COOPERATION AND TEAMWORK AT WORK THROUGH TRAINING**

The primary purpose of team training is to provide skills that enhance teamwork. For instance, team training often emphasizes interpersonal and feedback skills in addition to clarifying the need for mutual performance

monitoring, backup behavior, and goal setting. Without this training, efficient and effective taskwork might be handicapped or hindered altogether. Team training is not a program, a place, a simulator, or a collection of individuals being trained together; rather, it is a set of instructional strategies designed to enhance teamwork competencies. Nor is team training the same as team building. Team building primarily targets role clarification and has a greater influence on team member attitudes than on performance.

Although many types of instructional strategies have been developed to assist individuals in teams to cooperate at work, only four of the most prominent and relevant ones are briefly discussed in this section: cross-training, team coordination training, team self-correction training, and team leader training. A discussion of team building is also provided, with a specific focus on the distinction between team building and team training.

The first strategy, cross-training, generally involves training each individual member on the tasks of all other team members. The goal is to create shared task models and knowledge of task-specific role responsibilities to facilitate enhanced performance monitoring and backup behavior. Knowledge of the team's mission and objectives is also targeted in cross-training. This strategy is primarily information based but may also involve simulation exercises.

Team coordination training is similar to cross-training in that it creates a general understanding of teamwork skills, but it does so with a more specific focus on promoting mutual performance monitoring and backup behavior. Coordination training uses practice and behavioral modeling to teach members about teamwork skills and help them understand the underlying processes of teamwork. Team members may be given opportunities to practice and demonstrate the skills taught in low- to moderate-fidelity simulations. These are similar to the ideas that are presented in discussions of crew resource management (CRM) and are frequently used in aviation and military settings.

Another type of training that improves mutual performance monitoring, as well as initiative, leadership, and communication, is team self-guided correction training. The team directly drives this instructional strategy in that team members are taught techniques for monitoring their own behavior and assessing the degree of training effectiveness. The team members then use feedback regarding their performance and effectiveness to correct problems that may have surfaced. The primary focus of this strategy is improved communication and increased levels of performance

feedback to facilitate future performance. Feedback enables team members to perform their subtasks competently and demonstrates the contribution of that performance to the performance of others and of the team as a whole.

Finally, team leader training aims to create leaders who are better able to motivate others, structure tasks, and facilitate shared task models of teamwork behaviors within the team. The emphasis is often on teaching team leaders to effectively debrief their members using input from their whole teams regarding specific behaviors rather than general results or proxies of effectiveness. Lecture, behavioral modeling, and practice are the typical mechanisms used to deliver this type of training.

Another technique that is often used to improve team functioning and effectiveness is team building. Team building should not, however, be confused with team training. Although team training and team building are seemingly similar, they approach teamwork from different perspectives. For instance, team training is generally intended to facilitate a shared understanding of job-related KSAs, whereas team building focuses on the process of teamwork and is intended to assist individuals and groups in examining their own behavior and relationships. It is important to note, however, that team building has been found to have a greater impact on the attitudes of individual team members than on overall team performance. Researchers have suggested that this is due to the fact that team building typically focuses on interpersonal relations, communication, problem solving, goal setting, and/or decision making rather than focusing directly on affecting performance. Because team building is generally one-dimensional in its aim (e.g., problem solving, interpersonal skills), its impact on overall team performance is limited. Finally, it should also be noted that team building is not a one-shot deal; observable improvements in cooperation may take time to appear and often require ongoing attention to attain long-term effectiveness.

Taken together, the importance of team training (and, to a lesser extent, team building) is that it focuses on deficiencies in teamwork KSAs to bolster and coordinate team member actions. Both team and organizational performance are likely to improve as a result of team training interventions. Studies conducted in military and aviation settings have found that enhanced training has resulted in significant team performance improvements. For example, researchers have seen as much as a 45% improvement in mission performance following team training. This includes a 33% improvement in tactical decision-making performance and a 10 to 34%

improvement in team coordination. Regardless of the training strategy used, it is important to measure the impact of training on team performance to determine its utility to the organization. The next section discusses issues related to measuring team effectiveness.

## **6. MEASURING TEAM PERFORMANCE AND EFFECTIVENESS**

The underlying assumption of team training is that teams can be led to perform more effectively. However, to determine whether the enhanced instruction produced its desired effect, the effectiveness of training must be assessed. The effectiveness of training refers to the degree to which desired training outcomes are achieved, that is, the degree to which a team accomplishes its assigned tasks and mission objectives. The most important of these outcomes is the degree to which trainees transfer learned material to the job. When training transfers, the organization can be certain that learned material is used by trainees when they return to their normal work roles. Team performance measures can focus on any section of the team effectiveness input, process, or outcome system and can indicate how the team carried out assigned tasks and objectives.

Extant literature on teams and team effectiveness has identified a number of issues that should be addressed during measurement. Specifically, Salas and colleagues identified five ingredients for any team performance measurement system. First, it must be theoretically based. Second, it must consider multiple levels of measurement. Third, it must be able to describe, evaluate, and diagnose performance. Fourth, it must provide a basis for remediation. Fifth, it must support ease of use. These requirements are described in turn.

The measurement of team performance must be theoretically based, preferably on previously identified models of teamwork. After all, theory building requires that established models be empirically tested to advance the current state of knowledge. Capturing, defining, and measuring the interactions that characterize teamwork are basic to our ability to understand its meaning. Furthermore, understanding the KSAs that define teamwork is critical to establishing a nomological net of the interrelationships of these variables that should be used as the basis for structuring measurement tools. Without grounding the measurement of team performance in theoretical models of

teamwork, team theory cannot move beyond the conceptual stage.

The level of measurement must also be considered. Multiple levels of measurement should be used in the measurement system because teamwork occurs at multiple levels (e.g., individual vs team, process vs outcome). To begin, team performance needs to be evaluated at the individual level, as well as at the team level, to get a complete picture. Effective team performance can be enabled by the actions of team members operating both individually and as a collective.

Often it is impossible to determine what part of performance is due to an individual's abilities and what part is due to the actions of the team as a whole. For teams with a large number of members, individual input to the overall task may be obscured, causing individuals to lose sight of team goals. When this happens, team members will often put their needs in front of helping other team members to achieve the collective goal. Similarly, research has suggested that many individuals prefer individual assessment and find that method to be fairer. This is especially the case when individuals perceive that their individual performance is higher than what is observed at the team level. The consensus seems to indicate that individual performance should be evaluated throughout, whereas team-level outcomes should be assessed following completion of the training task or objective.

Another issue concerning the level of measurement relates to whether a greater emphasis should be placed on evaluating teamwork processes (e.g., compensatory behavior, communication, information exchange, shared task models) or team outcomes (e.g., accuracy, timeliness, product quality, goal accomplishment). Although the outcome of performance is nearly always significant, the measurement instrument should clearly differentiate between the outcome and how the outcome was achieved. In team performance measurement, the decision to evaluate team effectiveness as a team process or a team outcome is particularly important when the purpose of performance evaluation is to provide feedback for performance improvement.

Although many managers focus on team outcomes (e.g., factors that directly affect the bottom line), this evaluation provides no information to the team as to how to make performance improvements. For instance, telling sales team members they did not make the goal of selling 300 widgets provides less information to guide performance improvement than does telling the same team that there had been a noticeable decrease in information exchange among the team members during the past month that resulted in

decreased sales. Measurement must provide information that indicates why processes occurred and how those processes are linked to performance outcomes.

To be of any value, the measurement mechanism must be able to describe, evaluate, and diagnose performance. Measures must be sensitive enough to document the moment-to-moment interactions and changes in performance. Measures must also distinguish between effective and ineffective processes, strategies, and teamwork behaviors. The level of evaluation, type of performance to be evaluated (i.e., process or outcome), and environmental constraints within the organization collectively determine the measurement tools needed to measure team effectiveness. However, if the performance evaluation is conducted for developmental or remediation purposes, it is especially important that the tool be able to describe, evaluate, diagnose, and provide feedback regarding performance. Cannon-Bowers and colleagues provided a concise framework of team performance measurement tools. A few of the more flexible tools (i.e., those that can be used in most situations) are observational scales and expert ratings. More situation-specific tools include decision analysis, policy capturing, and archival records. In a broad sense, performance is typically evaluated through direct observation (e.g., employee behavior) or indirect observation (e.g., employee records, customer communication). The decision to use direct or indirect observation generally depends on what information is available to the evaluator. Both forms have their pros and cons. Similarly, the choice of measurement tool may be guided by the particular organizational and situational constraints imposed on the researcher.

Next, team performance measurement systems must outline a path to improvement. From the performance diagnosis generated, a logical basis for remediation that includes feedback, knowledge of results, and direction for subsequent instruction must be provided. Practitioners must be able to provide teams with feedback necessary to improve future performance. Measurement tools must provide information that indicates why processes occurred as they did and how particular processes are linked to certain outcomes.

A final ingredient of team performance measurement systems is that they must be easy to use. The focus of a particular measure should always be operationalized from a construct that is relatively free of contamination and deficiency, and it should not be tedious or cost-prohibitive to use. One challenge of team research has been to develop powerful, reliable, and valid measurement techniques for measuring team performance.

A final topic to consider within team measurement should be noted. That is the idea of team mental models or the “shared cognition” of a team. The idea that shared mental models allow team members to coordinate by anticipating and predicting each other’s needs through common expectations of performance was discussed previously. However, the ability to measure these expectations as an external observer has proven to be more difficult than have other team processes. Currently, indirect measures of shared cognitions measure individual mental models of the team and task and aggregate the individuals’ mental models to assess the degree of overlap between the cognitions used. Further research is needed to develop tools that can directly address team shared cognition. The diversification of techniques such as automated measures, or measures that are embedded in the task (e.g., simulations), also provides avenues for future research endeavors.

## 7. CONCLUSION

Experience has shown that teamwork and cooperation can substantially affect organizational effectiveness. However, organizations, managers, executives, and leaders need to know what develops and fosters teamwork as well as what derails it. This article has discussed why teamwork is important.

The article also discussed the nature of teamwork, indicating that it consists of the KSAs that are displayed in support of one’s teammates, objectives, and mission. Ways in which cooperation and teamwork can be derailed were examined as well. Similarly, a presentation of teamwork facilitators was provided. For example, very often it is the responsibility of the leader to create a shared understanding of the vision and goals of those who report to him or her. Furthermore, it is the leader who is usually able to obtain resources and distribute them appropriately, whether through reward systems, work tools, personnel, or training opportunities. At the group level, many factors contribute to the level of cooperation that occurs among individuals. Not only do shared vision, the task itself, and the organizational culture play a part, but factors such as the level of diversity, familiarity, and the number of individuals working together also affect cooperation.

The article also presented a discussion of how training is used to promote cooperation and teamwork. Training opportunities are especially important from a cooperation perspective in that training teamwork skills is imperative to effective cooperation. Training

allows members to gain knowledge about their teammates' skills, abilities, roles, and responsibilities, and this allows them to better anticipate each other's needs and to adapt to the changing organizational environment and tasks. Training can also assist in the creation of a shared vision through goal setting and interpersonal relations.

The article then examined the measurement requirements for team effectiveness research. The fact that teamwork and training effectiveness must be measured should not be lost on the reader. Without effective description, evaluation, and diagnosis of team effectiveness, we would not know where to begin to correct performance deficiencies.

From an organizational perspective, teams and the social system in which they are embedded are interdependent, such that a change in one will affect the other. Thus, organizations that encourage cooperative behaviors through their leadership and organizational systems (i.e., reward systems) will facilitate and encourage cooperation among employees. There is clear evidence that cooperative work arrangements improve overall organizational performance. Individuals also have a role in the effectiveness of cooperation and teamwork as an organizational tool. Evidence indicates that attitudinal factors such as team identity, orientation, commitment, and trust are important. Commitment to the group will increase the willingness of the individual to expend effort and have higher involvement in the group task that is required for cooperation.

Understanding the nature of teamwork and how its interrelated components manifest themselves in terms of cognitions, behaviors, and feelings is critical to promoting coordinated adaptive team performance in increasingly complex and dynamic environments.

Thus, actions taken to improve cooperation and teamwork can have a substantial impact on the organizational bottom line.

### See Also the Following Articles

Organizational Structure ■ Work Teams

### Further Reading

- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo, & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333–380). San Francisco: Jossey-Bass.
- Ilgen, D. R. (1999). Teams embedded in organizations: Some implications. *American Psychologist*, *54*, 129–139.
- Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, *23*, 531–546.
- McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. Guzzo, & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 149–203). San Francisco: Jossey-Bass.
- Rothschild, J., & Whitt, J. A. (1986). *The cooperative workplace: Potentials and dilemmas of organizational democracy and participation*. New York: Cambridge University Press.
- Salas, E., & Cannon-Bowers, J. A. (2000). The anatomy of team training. In S. Tobias, & J. D. Fletcher (Eds.), *Training and retraining: A handbook for businesses, industry, government, and military* (pp. 312–335). New York: Macmillan.
- Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990's and beyond. *Group Dynamics*, *4*, 44–67.



# Coping

Carolyn M. Aldwin and Loriena A. Yancura

University of California, Davis, California, USA

1. Types of Coping
2. Coping and Health Outcomes  
Further Reading

## GLOSSARY

**adaptive defense mechanisms** Coping behaviors, including humor and altruism, that help to alleviate anxiety; also called mature defense mechanisms.

**coping strategies** Consciously chosen, intentional behaviors and cognitions that can flexibly respond to environmental demands.

**coping styles** Consistent methods of dealing with stress that focus on information accessing as well as emotions.

**defense mechanisms** Unconscious dispositional behaviors (e.g., defensive styles) associated with pathology.

**depersonalization** The act of separating one's identity or sense of reality from a situation.

**emotion-focused coping** A wide range of strategies (e.g. avoidance, withdrawal, expressing emotion, and the use of substances such as alcohol or food) that are directed toward managing one's emotional response to a problem.

**meaning making** A coping strategy that involves trying to see the positive or meaningful aspects of the stressful situation, especially with severe or chronic stressors. It is sometimes referred to as cognitive reappraisal.

**problem-focused coping** Behaviors and cognitions aimed at solving a problem that is generating anxiety. It is also called instrumental action.

**psychotic defense mechanisms** Individual distortions of a situation that may involve outright denial of reality, neurotic defense mechanisms, or may consist of denial of the

significance of an event, such as intellectualization. They are sometimes called immature defense mechanisms.

**religious coping** A multifaceted coping strategy including elements of social support coping, problem-focused coping, and emotion-focused coping, that aims to conserve or transform meaning in the face of adversity.

**sequelae** Pathological condition and/or a secondary consequence, in this case resulting from trauma.

**social support coping** Seeking advice, emotional aid, and/or concrete aid from others.

In the past 25 years, more than 30,000 articles have been published on coping. This is not only a testament to the popularity of this construct but also reflects the enormous interest in the psychological and physical effects of stress. Theoretically, at least part of the reason for differential vulnerability and resilience to stress lies in how individuals cope with stress. Testing this hypothesis has led to the growing understanding of the complexity of the adaptation process. There is simply no "silver bullet" that will magically solve all problems but, rather, the effects of coping vary by how coping strategies are assessed, the type of problem, and a variety of situational factors, as well as the skill and resources of the individual facing the problem.

## 1. TYPES OF COPING

There are three major approaches to the study of coping: psychoanalytic approaches, coping styles, and coping

processes. These approaches to coping differ as to whether they focus on unconscious mechanisms, personality or information processing styles, or environmental influences.

### 1.1. Psychoanalytic Approaches

Psychoanalytic approaches focus on the use of defense mechanisms, which are unconscious ways of fending off anxiety. Defense mechanisms differ from coping strategies in that they are unconscious, dispositional (e.g., defensive styles), and are associated with pathology, whereas coping strategies are consciously chosen, intentional, and can flexibly respond to environmental demands. However, defense mechanisms vary in the degree to which they distort reality and may be hierarchically arranged from psychotic to adaptive or “mature.” Psychotic defense mechanisms may involve outright denial of reality, whereas “immature” or neurotic mechanisms may consist of denial of the significance of an event, such as intellectualization. The most adaptive or mature defense mechanisms include humor and altruism.

Because they are believed to act on an unconscious level, defense mechanisms are difficult to study and they are usually assessed using in-depth interviews and case studies. Thus, there have been few large-scale studies of their effectiveness. However, they did provide the initial impetus for the study of how individuals cope with stress.

### 1.2. Coping Styles

Coping styles differ from defense mechanisms in that they focus on how people deal with information as well as emotions, and are thought to be consistent across time and situation. As such, they typically take the form of dichotomies such as repression–sensitization, blunting–monitoring, and approach–avoidance. Research studies have typically found that approach styles are associated with better outcomes than avoidance styles. However, this does not hold true across all circumstances, such as with uncontrollable stressors. Furthermore, studies that empirically test consistency in coping across time or situations are rare, and the ones that do include such tests generally find that individuals are not very consistent in how they cope with different types of problems.

### 1.3. Coping Processes

In contrast to the two previous views, the coping process approach argues that coping is responsive to both environmental demands and personal factors such as values

and beliefs. As such, coping strategies are flexible and unfold over time, either in response to changing appraisals or as a function of developmental processes. Coping strategies are proactive and are not simply responses to environmental contingencies. One can appraise a situation as benign, in which case no coping is needed, or as involving threat, harm, or loss or as a challenge, all of which may evoke various coping strategies. Because this approach emphasizes the flexible nature of coping, the focus is on how individuals cope with particular situations and, as mentioned earlier, several studies have shown that coping strategies do vary across situations.

There are many different conceptualizations of coping strategies, but the five general types of coping strategies are problem-focused coping, emotion-focused coping, social support, religious coping, and meaning making. Problem-focused coping, also called instrumental action, encompasses behaviors and cognitions aimed at solving the problem, such as seeking information, taking direct action, or breaking the problem down into more manageable pieces, a strategy referred to as “chunking.” Sometimes, delaying or suppressing action can be a useful problem-focused strategy. For example, purposefully delaying a direct confrontation with someone may lead to a more rapid solution to a problem than acting in anger.

Emotion-focused coping includes a wide range of strategies that are directed toward managing one’s emotional response to the problem. Some examples are avoidance, withdrawal, expressing emotion, and the use of substances such as alcohol or food. As might be expected, avoidance strategies are often associated with poor outcomes, but other emotion-focused strategies, such as expressing emotion through journals or writing, may be associated with positive outcomes.

Social support involves seeking both emotional and concrete aid from others or advice. The outcome of these types of coping strategies often depends on the social context. For example, confiding in others after a trauma is generally associated with better outcomes, but if the confidant responds negatively, emotional distress may be increased.

The study of religious coping is relatively new. It can contain elements of social support or problem-focused and emotion-focused coping, and it seeks to conserve or transform meaning in the face of adversity. In general, religious coping is associated with positive outcomes, but it does have negative guises. Belief in a punitive God, or feeling that one has been treated unfairly or been deserted, may be associated with much poorer outcomes.

The final category, meaning making, is the least well understood. It is sometimes referred to as “cognitive reappraisal” and involves trying to see the positive or meaningful aspects of the situation, especially with severe or chronic stressors. As with religious coping, how one goes about making meaning may affect its association with outcomes. Simply asking “Why me?” may be associated with poorer outcomes, but coming to realize how a problem fits into the larger pattern of one’s life may be a painful process but in the end may be one way in which individuals grow from stressful or traumatic experiences.

#### **1.4. Measurement Issues**

Coping style measures ask individuals how they typically cope with problems. The dichotomized nature of coping styles lends itself well to psychometric assessment. The factor structures of coping style inventories are generally stable, and the factors correlate well with psychological outcomes. Given that coping may not be very stable, however, their ecological validity is somewhat questionable.

Process measures ask individuals to select a recent problem and indicate how they coped with that type of problem. Unfortunately, process measures typically have somewhat “messy” psychometrics, and their factor structures may be unstable. This makes comparison across situations or over time problematic.

Sometimes a more accurate picture can be obtained using multiple assessments, either through daily diaries or through event sampling throughout the day. Although some scholars have criticized the psychometric properties of these measures (often consisting of only one item per strategy), these types of procedures may be especially useful in the context of coping with chronic stressors because they permit within-subject analyses that may yield strong relationships between coping and outcomes.

## **2. COPING AND HEALTH OUTCOMES**

### **2.1. Mental Health Outcomes**

In many ways, being able to successfully cope with problems can be considered a hallmark of psychological health. Difficulties in coping may be associated with emotional distress, including depression and anxiety. However, the relationship is neither simple nor straightforward. Because there may be differences

between how people cope with trauma and with more everyday stressors, these two different types of problems are addressed separately.

#### **2.1.1. Coping with Stressors**

The relationship between coping and psychological symptoms is characterized by complexity along a number of different dimensions. First, coping strategies have situation-specific effects. Many studies have demonstrated that the effectiveness of coping strategies rests on the controllability of the stressor. Controllable stressors, such as problems at work, are more effectively managed with problem-focused coping, whereas uncontrollable stressors, such as the death of a loved one, are more effectively managed with emotion-focused coping.

Second, there may also be a confound between stress, distress, and coping. Higher levels of stress may evoke more coping strategies and also create more distress; thus, all coping strategies will correlate with higher levels of distress in particularly difficult circumstances. Only by controlling for the stressfulness of the problem can the true relationship between coping and mental health be revealed.

Third, there is a difference between short- and long-term outcomes. Avoidance coping strategies, for example, may decrease psychological distress in the short term but cause great distress in the long term if the underlying problem remains unresolved. Brief periods of respite, however, can allow individuals to “recharge their batteries” and promote better problem-focused coping in the long term.

Fourth, the causal directionality of coping may not be clear. Rather than coping resulting in depressive symptoms, individuals who are depressed may cope in different ways than nondepressed people.

Fifth, people may employ the same coping strategies with different levels of skill. Thus, it makes sense to determine the efficacy of the strategy—that is, whether it “worked” for that person in that particular situation.

Finally, most research has focused only on coping as a means of reducing symptoms. However, effective coping may lead to more positive outcomes, such as increases in mastery, self-knowledge, and better social relationships. Thus, it is not surprising that the literature on coping and mental health is rich but somewhat contradictory.

#### **2.1.2. Coping with Trauma**

Traumatic stressors are ones that either threaten or expose individuals to loss of life and major physical damage to

themselves, loved ones, or large numbers of individuals. Surprisingly, the trauma literature is much clearer in demonstrating the efficacy of coping strategies. In general, how individuals cope with trauma and its sequelae are better predictors of the development of posttraumatic stress disorder (PTSD) than simply exposure to trauma.

There are four ways in which coping with trauma differs from coping with less severe stressors. First, traumatic circumstances are often uncontrollable and many people fall back on the use of unconscious defense mechanisms, such as denial and repression. For example, depersonalization may be the only possible response to being tortured. Second, disclosure and seeking support from others appear to be especially important in traumatic situations. However, if the disclosure evokes negative responses from others, PTSD symptoms may become worse. Some individuals are relatively successful at “partitioning off” their trauma experience and rebuilding their lives.

Third, the process of coping with traumatic events and their sequelae is much longer and may take years. Finally, the coping strategy of meaning making is particularly important in dealing with trauma. Whether an individual is able to make some sense out of traumatic events can determine whether those events lead to positive or negative mental health outcomes.

## 2.2. Physical Health Outcomes

The evidence for the effect of coping on physical health outcomes such as pain, blood pressure, catecholamines, lipids, and immune system function is much weaker, in part because studies are relatively rare and may focus on special populations, such as heart disease patients or AIDS victims.

We identified four models by which coping may affect health outcomes. The direct effects model examines a simple correlation between coping and health outcomes, regardless of the situation or type of stressor. For example, instrumental action may be associated with better lipid or immune system profiles whether the stressor is major or minor. Second, coping may moderate or buffer the effects of stress on health outcomes. That is, there are no direct effects, but individuals under high levels of stress who use instrumental action may have lower blood pressure levels. The mediated effects models suggest that the effects of coping depend on a variable or variables intervening between coping and physical health outcomes. For example, many of the physiological effects

of coping on health might be explained through its effect on psychological distress. If an individual's coping strategies decrease negative affect, then it may be associated with better immune system profiles. The fourth model is the contextual effects model, which states that the effect of coping on health might depend on environmental context. For example, Zautra and colleagues showed that seeking social support may be effective in reducing pain for diseases such as osteoarthritis but less effective for rheumatoid arthritis, in part because the symptoms in the latter illness are less clear and thus may be more difficult for others to understand.

There is evidence for each of these models; however, none of them has received overwhelming empirical support. Clearly, much research is needed to illuminate the pathways by which coping influences physical health outcomes.

## See Also the Following Articles

Health and Culture ■ Posttraumatic Disorders ■ Stress

## Further Reading

- Aldwin, C. M. (1999). *Stress, coping, and development: An integrative approach*. New York: Guilford.
- Cramer, P. (2000). Defense mechanisms in psychology today: Further processes for adaptation. *American Psychologist*, 55, 637–646.
- de Ridder, D., & Kerssens, J. (2003). Owing to the force of circumstances? The impact of situational features and personal characteristics on coping patterns across situations. *Psychology & Health*, 18(2), 217–236.
- Folkman, S., & Moskowitz (1994). Coping: Pitfalls and promise. *Annual Review of Psychology*, 55, 745–774.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal & coping*. New York: Springer-Verlag.
- Mikulincer, M., & Florian, V. (1996). Coping and adaptation to trauma and loss. In M. Zeidner, & N. S. Endler (Eds.), *Handbook of coping: Theory, research, applications* (pp. 554–572). New York: Wiley.
- Pargament, K. (1997). *The psychology of religion and coping: Theory, research and practice*. New York: Guilford.
- Skinner, E. A., Altman, J., & Sherwood, H. (2003). Searching for the structure of coping: A review and critique of category systems for classifying ways of coping. *Psychological Bulletin*, 129, 216–269.
- Tennen, H., Affleck, G., Armeli, S., & Carney, M. A. (2000). A daily process approach to coping: Linking theory, research, and practice. *American Psychologist*, 55, 626–636.
- Valliant, G. E. (1993). *The wisdom of the ego*. Cambridge, MA: Harvard University Press.



# Counseling and Culture

Paul B. Pedersen

University of Hawaii, Honolulu, Hawaii, USA

1. Introduction
  2. Cultural Bias in Counseling
  3. Cultural Encapsulation of Counselors
  4. Culturally Learned Assumptions in Counselor Education
  5. The Cultural Grid Linking Counseling and Culture
  6. Multicultural Theories of Counseling
  7. Multiculturalism as a Fourth Force
  8. Resistance to Multiculturalism as a Fourth Force
  9. Conclusion
- Further Reading

global context, whether that perspective is held in the Western or Eastern geographic hemisphere.

Counseling is a product of Western cultures whose values have continued to dominate the counseling profession through the imposition of cultural assumptions as counseling services have spread to other cultures, demonstrating the generic centrality of culture to achieve competence in the counseling process.

## GLOSSARY

**cultural encapsulation** The tendency of people to build walls or capsules around themselves by simplifying or minimizing the importance of other people's culture's.

**cultural grid** A framework combining personal factors (e.g., behavior, expectation, values) on one dimension and the cultural contexts in which those personal factors were learned on the other dimension.

**culture-centered perspective** A perspective that makes the cultural context central to an understanding of every situation.

**fourth force** The view of multiculturalism as a "fourth dimension" of counseling to supplement the psychodynamic, humanistic, and behavioral dimensions.

**multicultural competencies** A list of 34 specific competencies describing how an understanding of culture is basic to the counseling process; this list has been evolving since 1962.

**Western culture** A dominant culture sociopolitical perspective favoring individualism and linear thinking in the

## 1. INTRODUCTION

Because all behaviors are learned and displayed in a cultural context, accurate assessment, meaningful understanding, and appropriate intervention require consideration of the cultural context when providing counseling services. Making the consumer's cultural context central rather than marginal when providing psychodynamic, behavioral, and humanistic services strengthens those traditional counseling perspectives. However, because counseling first developed as a field in a Euro-American cultural context, counseling typically reflects Euro-American assumptions in textbooks, theories, tests, ethical guidelines, methods, and other services. The purpose of this article is to discuss the generic importance of making cultural context central rather than marginal to the counseling process.

## 2. CULTURAL BIAS IN COUNSELING

In 1996, a report by the Basic Behavioral Science Task Force of the National Advisory Mental Health Council (NAMHC) documented the extent of cultural encapsulation of mental health services. First, anthropological and cross-cultural research has demonstrated that cultural beliefs influence the diagnosis and treatment of mental illness. Second, the diagnosis of mental illness differs across cultures. Third, research has revealed differences in how individuals express symptoms in different cultural contexts. Fourth, culturally biased variations in diagnosis vary according to the diagnostic categories relevant to the majority population. Fifth, most providers come from a majority culture, whereas most clients are members of minority cultures. If the standard practices of mental health services are themselves encapsulated, as suggested by the NAMHC report, these cultural biases will certainly influence the practical applications of psychology through counseling and therapy.

Counseling has typically assumed that there is a fixed state of mind that is obscured by cultural distortions and that relate behaviors across cultures to some universal definition of normative behavior described in textbooks. A contrasting anthropological perspective assumes that cultural differences are clues to divergent attitudes, values, or perspectives that differentiate one culture from another based on a culture-specific viewpoint. Anthropologists have tended to take a relativistic position when classifying and interpreting behavior across cultures. Psychologists, in contrast, have linked social characteristics and psychological phenomena with minimal attention to cultural differences. When counselors have assumed the same interpretation of similar behaviors regardless of the cultural context, cultural bias has been the consequence. The influence of cultural bias has led to the encapsulation of counselors by dominant culture perspectives regardless of the cultural context.

## 3. CULTURAL ENCAPSULATION OF COUNSELORS

In 2001, Ponterotto and colleagues described how counseling has a reputation among many minorities as protecting the status quo of the dominant culture by requiring minority clients to fit the dominant culture system even when that system is acknowledged to be unjust. In 1999, Sue and Sue described how these

attitudes have resulted in documented examples of scientific racism.

In 1962, Wrenn first introduced the concept of cultural encapsulation. This perspective assumes five basic identifying features. First, reality is defined according to one set of cultural assumptions. Second, people become insensitive to cultural variations among individuals and assume that their own view is the only right one. Third, assumptions are not dependent on reasonable proof or rational consistency but are believed to be true regardless of evidence to the contrary. Fourth, solutions are sought in technique-oriented strategies and quick or simple remedies. Fifth, everyone is judged from the viewpoint of one's self-reference criteria without regard for the other person's separate cultural context. The encapsulation has not diminished over time. There is evidence that the profession of counseling is even more encapsulated now than it was when Wrenn wrote his original article.

In 1994, Albee described how completely psychology—and counseling as based on a psychological foundation—in the United States has been encapsulated during the past 100 years:

Most of the early leaders in psychology embraced ideological views that stressed the natural superiority of a white male patriarchy, the acceptance of social Darwinism, [and] the inferiority of women and of the brunette races. Calvinism stressed economic success as the hallmark of salvation, and psychology concurred. Anti-semitism and homophobia were standard. Eugenics spokesmen urged the elimination of the unfit and inferior and opposed welfare programs, decent wages, and safe working conditions.

These views continue to be held today, although sometimes in more subtle forms. The encapsulation of counselors is most evident in the way in which counselors are trained and educated.

## 4. CULTURALLY LEARNED ASSUMPTIONS IN COUNSELOR EDUCATION

The training of counselors has been characterized by conventional assumptions that reflect the values and priorities of the Euro-American cultural context. [Table 1](#) describes 10 examples of these assumptions that Pedersen used in 2002 to demonstrate their pervasive influence in applications of counseling.

**TABLE I**  
**Western Cultural Assumptions in Counseling**

<i>Euro-American cultural assumption</i>	<i>Comments</i>
Counselors all share the same unimodal measure of what is normal behavior.	There is a frequent assumption that describing a person's behavior as "normal" reflects a judgment that is both meaningful and representative of a desired pattern of culture-specific behaviors across social, economic, political, and cultural contexts, that is, represents the dominant culture perspective of culture-specific behavior.
Individuals are the basic building blocks of society.	This individualistic presumption is reflected in psychological terms such as personality and self-awareness. Conventional counseling is focused primarily on the development of individuals rather than collectivities or groups such as the family, the organization, or society itself.
Only problems defined within the framework of the counselor's specialized expertise or academic discipline are of proper professional concern to the counselor.	There is a tendency to isolate the professional identity of the counselor from that of other professionals even though multicultural problems wander across these boundaries freely.
There is a superior quality judgment attached to psychological abstractions.	In the counselor's use of professional jargon, he or she assumes that the same words have the same meaning across cultural contexts. Although this assumption is typical of a "low-context" culture where the context is less relevant, it would not apply to "high-context" cultures where all meaning is contextually mediated.
Independence is usually desirable, and dependence is usually not desirable.	As part of the Western emphasis on individualism, there is a belief that individuals should not be dependent on others or allow others to become dependent on them. This is not the case in a more collectivist culture.
Clients are helped more by formal and professional counseling than by their natural support systems.	Family and peer support are the primary resources in many cultures where professional counseling is less available or is inappropriate. In some non-Western cultures, counseling is a last resort only when everything else has failed.
Everyone thinks the same way, moving linearly from cause to effect.	It is not just the content of one's thinking that is culturally mediated but also the very process of thinking itself. Nonlinear thinking, typical of many non-Western cultures, will seem illogical to linear thinkers.
Counselors need to change clients to fit the system and not change the system to fit the clients.	Advocacy by counselors is frequently considered unethical and unprofessional. Much of the counseling literature relates to client adjustment, sometimes even when the system is wrong and the client is right.
History is not relevant for a proper understanding of here-and-now problems.	Counselors are more likely to focus on the immediate here-and-now events that created a crisis and to consider historical background as a distraction at best and as a defensive evasion at worst.
We already know all of our culturally learned assumptions.	Each time counselors discover one of their cultural biases, they disprove this assumption. As they increase their contact with persons and groups from other cultures, this process of self-discovery is accelerated.

One way in which to challenge the universal application of dominant culture perspectives, regardless of the cultural context, is to separate behaviors from expectations in counseling. This will demonstrate how shared positive expectations (e.g., safety, respect, fairness) might be expressed by very different behaviors and how similar behaviors (e.g., direct/indirect, loud/quiet) across cultures might be linked to very different expectations in a two-dimensional “cultural grid.”

## 5. THE CULTURAL GRID LINKING COUNSELING AND CULTURE

The connection between counseling and culture can be described visually in a cultural grid that includes individual behavior, expectations, and values on one dimension and the “culture teachers” of those individual characteristics on the other dimension. The cultural grid, as described by Pedersen in 2000, is an attempt to demonstrate the personal–cultural orientation of the individual in relationship to the many different cultures to which that individual belongs. The cultural grid shows an “inside-the-person” or intrapersonal cultural framework for demonstrating how cultural and personal factors interact. Through the cultural grid, each individual behavior can be linked to the culturally learned expectation on which that behavior is based, and each expectation can be linked to the culture teachers from whom those expectations were learned.

The cultural grid provides a systematic framework for the counselor to interpret each client behavior in the context where that behavior was learned and is displayed. The cultural grid demonstrates that there are many different culture teachers present in the counseling interview, with some belonging to the counselor and some belonging to the client. The cultural grid further provides a framework for comprehending the client’s culturally learned behavior from his or her point of view.

Pedersen pointed out that each cultural context is complicated and dynamic, influenced by many culture teachers who take turns being salient according to every time and place. An awareness of one’s cultural identity requires the ability to identify how each behavior is an expression of specific culturally learned expectations and how each expectation was learned from specific culture teachers. The culture teachers might come from family relationships, power relationships, or other memberships. Judging other people’s behavior out of

context or without regard to why they engaged in the behavior or where they learned the behavior is likely to be misleading at best and totally wrong at worst. It is not easy to discover why people do what they do, but that does not excuse the counselor from the obligation to try. Unless each client behavior is understood in context, that behavior is very likely to be misunderstood by the counselor (Fig. 1).

The culturally learned assumptions on which current theories of counseling are based, such as those favoring individualism over collectivism, need to be reassessed so that the theories of counseling can be applied across cultures more accurately and meaningfully.

## 6. MULTICULTURAL THEORIES OF COUNSELING

One culture-centered perspective that applies cultural theories to the counseling process is found in a 1996 book on multicultural counseling theory (MCT) by Sue and colleagues. This approach is based on six propositions that demonstrate the fundamental importance of a culture-centered perspective:

1. Each Western or non-Western theory represents a different worldview.
2. The complex totality of interrelationships in the client–counselor experiences and the dynamic changing context must be the focus of counseling, however inconvenient that might become.
3. A counselor’s or a client’s racial/cultural identity will influence how problems are defined and will dictate or define appropriate counseling goals or processes.
4. The ultimate goal of a culture-centered approach is to expand the repertoire of helping responses available to counselors.
5. Conventional roles of counseling are only some of the many alternative helping roles available from a variety of cultural contexts.
6. MCT emphasizes the importance of expanding personal, family, group, and organizational consciousness in a contextual orientation.

As these MCT propositions are tested in practice, they will raise new questions about competencies of multicultural awareness, knowledge, and skill in combining cultural factors with psychological process. How does one know that a particular psychological test or theory provides valid explanations for behavior in a particular cultural context? What are the cultural

Personal Variables			
	Where you learned to do it	Why you did it	What you did
<b>Cultural Teachers</b>			
1. Family relations Relatives Fellow countrypersons Ancestors Shared beliefs			
2. Power relationships Social friends Sponsors and mentors Subordinates Supervisors and superiors			
3. Memberships Coworkers Organizations Gender and age groups Workplace colleagues			
4. Nonfamily relationships Friendships Classmates Neighbors People like me			

FIGURE 1 Within-person cultural grid.

boundaries that prevent generalization of psychological theories and methods? Which psychological theories, tests, and methods can best be used across cultures? Which psychological theories, tests, and methods require specific cultural conditions? MCT assumes multicultural competencies.

Multicultural competencies have been developed by Sue and colleagues over the past 40 years or so. In a comprehensive description of these 34 competencies in 1998, Sue reviewed these multicultural competencies. A summary of those competencies in three dimensions, emphasizing multicultural awareness, knowledge, and skill, can be described in three developmental dimensions:

- *Dimension 1.* Counselors' awareness of their own assumptions, values, and biases includes (a) awareness of cultural heritage comfortable with differences but aware of limits; (b) knowledge about oppression,

racism, and discrimination; and (c) skills in self-improvement toward a nonracist identity.

- *Dimension 2.* Understanding the worldview of the culturally different client includes (a) awareness of emotional reactions toward other racial/ethnic groups, (b) knowledge about the culture of the client population and its influence on counseling and society, and (c) skills in mental health issues of other cultures and active involvement with ethnic minority groups.

- *Dimension 3.* Developing appropriate intervention strategies and techniques includes (a) awareness of religious and spiritual indigenous mental health resources; (b) knowledge about how counseling fits with other cultures, institutions, and assessments; and (c) skills in culturally appropriate counseling of indigenous people.

These three dimensions of multicultural competency have tended to be described as abstractions, and much

more work is needed to link these abstract competencies to individual cases and situations. Ponterotto and colleagues did an excellent job of reviewing the literature already in print to apply the multicultural competencies in practice. As of this writing, Constantine and Sue are working on an even more comprehensive demonstration of how the multicultural competencies can be applied in practice. Applying these competencies in practice will reduce the number of racist incidents in the counseling profession generally.

The lack of trust in people who provide counseling services and the belief that the status quo was being held are documented in “scientific racism” and “Euro-American ethnocentrism.” According to Sue and Sue’s 1999 book, cultural differences were explained by some through a “genetic deficiency” model that promoted the superiority of dominant White cultures. The genetic deficiency approach was matched to a “cultural deficit” model that described minorities as deprived or disadvantaged by their culture. Minorities in the United States were underrepresented among professional counselors and therapists, the topic of culture was trivialized at professional meetings, and minority views were underrepresented populations in the research literature. Consequently, the counseling profession was discredited among minority populations because they viewed counseling as a tool to maintain the differences between those who had power and access to resources and those who did not. Making the cultural context central to counseling is in no way intended to diminish the importance of established counseling theory. In fact, making culture central as a “fourth dimension” will serve to increase the accuracy and importance of established theory.

### **7. MULTICULTURALISM AS A FOURTH FORCE**

In 1998, Pedersen examined some of the issues involved in declaring multiculturalism as a fourth force in psychology:

1. Significant changes are taking place, even though there is disagreement about the nature of those changes.
2. Multiculturalism has become a significant domestic force in most cultures, even though it may not yet provide a coordinated global perspective.
3. Multiculturalism has sometimes been used to rationalize oppression in countries such as South Africa; consequently, it has a bad reputation in those countries.

4. It may be premature to describe multiculturalism as a fourth force, although it has had a more powerful impact in the more applied areas of psychology such as counseling.

5. The U.S. version of multiculturalism is grounded in the individualist values of that cultural context more than in a non-Western collectivist perspective.

6. Within-group differences of ethnocultural groups, such as demographics, status, and affiliation, function “like” cultures regarding issues of age, disability, gender, and other special interests.

7. Cultural similarities (e.g., among youth-age strata) probably exceed similarities across generations in each separate ethnocultural group.

8. Multiculturalism will change not only the content of our thinking but also the very process of thinking itself.

9. Making culture central enhances the meaningful usefulness of traditional psychological theories in ways that might lead us to call this new perspective a fourth dimension rather than a fourth force.

Adding a fourth dimension of multiculturalism to counseling theory is not an easy task but remains a very important aspect of making the cultural context central to counseling.

### **8. RESISTANCE TO MULTICULTURALISM AS A FOURTH FORCE**

In 1998, Sue identified sources of resistance to the term “multiculturalism as a fourth force”:

1. Some view multiculturalism as competing with already established theories of psychological explanation in ways that threaten the professions of counseling and psychology.

2. The terms “multiculturalism” and “diversity” are loosely associated with affirmative action, quotas, civil rights, discrimination, reverse discrimination, racism, sexism, political correctness, and other highly emotional terms.

3. To the extent that multiculturalism is connected with postmodernism, the arguments against postmodernism as a valid theory are also applied to multiculturalism.

4. Those favoring a universalistic perspective contend that the same practice of counseling and therapy applies equally to all populations without regard to cultural differences.

5. Others contend that there are no accepted standards for describing multiculturalism as a theory in practice and that it is too loosely defined to be taken seriously.

6. There are no measurable competencies for multicultural applications of counseling or adequate standards of practice.

7. Multiculturalism is too complicated, and it would be unrealistic to expect counselors to attend to such a range of factors simultaneously.

8. More research is needed on multicultural competencies, standards, methods, and approaches.

9. Multicultural standards cannot be incorporated into the counseling profession until all groups have been included.

10. Multiculturalism represents reverse racism, quotas, and is anti-White.

In discussing these sources of resistance, Sue pointed out the tendency to misrepresent or misunderstand the notion of multiculturalism and the dangers of that misunderstanding.

Although the task of reducing cultural bias in counseling is extremely difficult, it remains an “aspirational goal” of the profession.

## 9. CONCLUSION

In 1999, Sue and colleagues summarized what needs to happen for psychology programs to facilitate positive multicultural initiatives so as to include (a) faculty and student preparation in the development of cultural competence; (b) a multicultural curriculum in all aspects of education and training; (c) minority representation among students, staff, faculty, and administration; (d) an inclusive and positive campus climate; (e) recognition of culturally biased teaching and learning styles; (f) people providing a social support network and services that understand the minority experience; and (g) recognition that current programs, policies, and practices negating multicultural development must be changed.

Multiculturalism is not just a trend or fad; rather, it is a permanent phenomenon of social functioning that reflects the advanced technology of a world where international boundaries have been replaced by close interactions among people around the globe. As the primary support systems of the village and family are diminished in authority around the world, more and more societies are looking for alternatives to provide the necessary social support in their communities. The field of counseling is a promising alternative for many of the world's

cultures provided that counseling can be modified to fit with a variety of cultures and not merely impose Western or dominant culture values.

## See Also the Following Articles

Cross-Cultural Psychology, Overview ■ Cultural Psychology ■ Racial and Ethnic Minorities, Counseling of ■ Training, Cross-Cultural

## Further Reading

- Albee, G. W. (1994). The sins of the fathers: Sexism, racism, and ethnocentrism in psychology. *International Psychologist, 35*, 42–78.
- Basic Behavioral Science Task Force of the National Advisory Mental Health Council. (1996). Basic behavioral science research for mental health: Sociocultural and environmental processes. *American Psychologist, 51*, 722–731.
- Berry, J. W., Poortinga, Y. H., Segall, H. N., & Dasen, P. J. (1992). *Cross-cultural psychology: Research and applications*. Cambridge, UK: Cambridge University Press.
- Marsella, A. J. (1998). Toward a “global–community psychology”: Meeting the needs of a changing world. *American Psychologist, 53*, 1282–1291.
- Pedersen, P. (1998). *Multiculturalism as a fourth force*. Philadelphia: Brunner/Mazel.
- Pedersen, P. (2000a). *Handbook for developing multicultural awareness* (3rd ed.). Alexandria, VA: American Counseling Association.
- Pedersen, P. (2000b). *Hidden messages in culture-centered counseling: A triad training model*. Thousand Oaks, CA: Sage.
- Pedersen, P. (2002). Ethics, competence, and other professional issues in culture-centered counseling. In P. Pedersen, J. Draguns, W. Lonner, & J. Trimble (Eds.), *Counseling across cultures* (5th ed., pp. 3–28). Thousand Oaks, CA: Sage.
- Ponterotto, J. G., Casas, J. M., Suzuki, L. A., & Alexander, C. M. (2001). *Handbook of multicultural counseling* (2nd ed.). Thousand Oaks, CA: Sage.
- Sue, D. W., Bingham, R. P., Porche-Burke, L., & Vasquez, M. (1999). The diversification of psychology: A multicultural revolution. *American Psychologist, 54*, 1061–1069.
- Sue, D. W., Careter, R. T., Casas, J. M., Fouad, N. A., Ivey, A. E., Jensen, M., LaFromboise, T., Manese, J. E., Ponterotto, J. G., & Vazquez-Nuttall, E. (1998). *Multicultural counseling competencies: Individual and organizational development*. Thousand Oaks, CA: Sage.
- Sue, D. W., Ivey, A. E., & Pedersen, P. B. (1996). *Multicultural counseling theory*. Pacific Grove, CA: Brooks/Cole.
- Sue, D. W., & Sue, D. (1999). *Counseling the culturally different: Theory and practice* (3rd ed.). New York: John Wiley.
- Sue, S. (1998). In search of cultural competencies in psychology and counseling. *American Psychologist, 53*, 440–448.

Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist*, 54, 1070–1077.

Wrenn, G. (1962). The culturally encapsulated counselor. *Harvard Educational Review*, 32, 444–449.

Wrenn, G. C. (1985). Afterward: The culturally encapsulated counselor revisited. In P. Pedersen (Ed.), *Handbook of cross cultural counseling and therapy* (pp. 323–329). Westport, CT: Greenwood.



# Counseling Interview

Paul H. Glasser

Rutgers, The State University of New Jersey,  
New Brunswick, New Jersey, USA

Sara F. Fine

Bar-Ilan University, Ramat Gan, Israel

1. Purpose and Goals
  2. Dynamics of Counselor-Client Interaction
  3. Issues, Dilemmas, and Contradictions in the Counseling Process
  4. Conclusion
- Further Reading

## GLOSSARY

**catharsis** The expression of feelings and emotions by a client or patient during a clinical interview.

**clinical interview** A relationship between a psychiatrist, a psychologist, a social worker, or other types of professionals with a client or patient, the focus of which is the facilitation of the growth and development of the client or patient.

**confidentiality** A client's or patient's trust that the counselor will not reveal to others the information given to him or her.

**contract** An agreement between a counselor and client or patient about the goals of the treatment process and the means that will be used to achieve them.

**counseling** A generic term that refers to a relationship between two or more people in which one person facilitates the growth and development of others in order to help them deal with their problems more effectively.

**counseling interview** A dyadic relationship between a counselor and a client or patient in which the former facilitates the growth and development of the latter in order to help him or her deal more effectively with his or her problems.

**intake interview** The first interview by a counselor in which an assessment of the client or patient is made and an agreement is reached by both parties about what the goals of the counseling process will be and how these goals will be achieved.

**presenting problem** A problem the client or patient is experiencing or, when dealing with an involuntary client, the problem the referring agency defines.

**privileged communication** Laws that give the counselor the right to hear almost anything a client or patient says with the responsibility of not repeating it to others.

**psychiatric interview** A dyadic relationship between a psychiatrist and a patient for the purpose of diagnosis and treatment of a patient who has problems in interpersonal functioning and/or internal psychological and emotional difficulties and that includes the assessment and treatment of any physiological sources of the patient's distress.

**trust** The client or patient's belief that whatever the counselor does is in his or her best interest.

The purpose of the counseling interview is to help a client or patient with a problem he or she is experiencing or a problem the referring agency defines. This is known as the presenting problem. The counselor must be in control of the interview and respond to the client in a professional way that is different from that of friends and neighbors. The rhythm and cadence of the interview requires that the counselor allow the client to lead, focus on his or her feelings, be open and honest in response to the client, and be reflective so that the client learns that this is his or her problem. Basic to the

counseling process is that the client must learn to trust the counselor so that he or she will be open to seeing the problem differently and dealing with it more effectively. The variety of methods and techniques the counselor uses to achieve this end are described in detail in this article.

## 1. PURPOSE AND GOALS

The purpose of the counseling interview is to help the client with a problem he or she is experiencing or, when dealing with an involuntary client (e.g., a criminal or delinquent or a parent accused of abuse or neglect), the problem the referring agency defines. This is often called the presenting problem. Although it is imperative that the counselor deal with the problem as the client or referring agency sees it, this does not mean that the counselor always perceives the problem in the same way. For example, a wife may come to the counselor because she believes her husband does not love her anymore, but it soon becomes apparent that her interpersonal response to her husband is at least partially responsible for his negative reactions to her. However, it may take a number of sessions before the counselor believes it is possible for the client to understand the problem in a way that is different from his initial presentation. [Although both counselors and clients may be men or women, for purposes of style we refer to counselors in the feminine and to clients in the masculine.]

Early in the interviewing process, it is important to set goals for counseling and to discuss the means that will be used to achieve these ends. Agreement is reached with the client that must reflect both an understanding of the problem by the client and counselor and a way of dealing with the problem that is likely to be effective and agreeable to the client. This agreement about means and ends is called the contract. In some counseling situations, the contract is implicit in the setting. For example, the employment counselor is there to help the client find a job, or the academic counselor sees her purpose is to help the student find the best college for him. In other situations, the contract may be much more complicated; for example, the father accused of abusing his child who believes “spare the rod, spoil the child” may need to agree to consider alternative approaches to child rearing. In the latter situation, considerable discussion between the client and the counselor, sometimes over a series of interviews, may be necessary before a contract can be established.

## 2. DYNAMICS OF COUNSELOR-CLIENT INTERACTION

There is a rhythm and cadence in the counseling process that often needs to be established in the early interviews. First, the counselor allows the client to lead. It is the client who is asking for help and he must do so in the way that makes the most sense to him. This does not mean that the client will present material in the most logical way. Rather, he will present the “story” in his own way, which in itself should have meaning to the counselor who is listening. The organization of the material presented may have as much meaning as the content. The counselor will make note of what is not said as well as what is said. She will notice any changes in speech patterns and times of silence as the narrative flows. By allowing the client to lead, the counselor is beginning to establish the trust that the client needs to continue.

Second, the focus, particularly in the early interviews, should be on feelings. Clients come to each new counseling session, even those that seem to deal with relatively unemotional situations, with feelings— anxiety or fear, grief or despair, rage or betrayal. The student may be afraid that he will not get into the college of his choice or that he may disappoint his parents, the teenager may be furious at her parents for limiting her activities, or the husband may be overwhelmed with grief after the death of his wife. In each case, before the client can move on he must have the opportunity to express these feelings. The expression of such feelings is called catharsis. The counselor can help the client by putting an accurate name to these feelings and describing their effect, intensity, and magnitude, which helps the client to control them and then move on to deal with the problem that causes them.

A third issue is the quality of the counselor’s communication. She must be open to what is heard and open and honest in her response. It means accepting the client’s statements, not arguing or debating with the client about what has been said or what the client feels and not using the word “but” to rebut what the client is feeling or saying. It means being nonjudgmental even when the client expresses values incompatible with those of the counselor or that seem morally reprehensible. This does not mean that the counselor will never disagree with the client or that she approves of the client’s antisocial behavior. It does mean the counselor will not do so until she has begun to establish a relationship of trust with the client, and that the

counselor will not confront the client with his unacceptable behavior until the client has sufficient positive feelings about the counselor that he can accept such confrontation with understanding and insight.

There are three therapeutic-type responses that are very useful in allowing the counseling interpersonal relationship to develop. The first we call minimal encourager. This is nothing more than nodding, murmuring “uh huh,” or using simple statements such as “Tell me more” or “Please go on.” Such statements are minimal and do not intrude on the client’s flow of material.

The second type of response is summation of content. By summarizing or paraphrasing the client’s content in her own words, the client can confirm or modify the counselor’s perceptions. This allows the client to continue to control the interview situation while at the same time the counselor confirms an understanding of what is being said.

The third technique is to reflect the client’s feelings about the situation or person he is describing. This is a powerful response because this teaches the client at least three things: (i) that it is okay to have feelings, no matter what they are; (ii) that the counselor is not afraid of any of his feelings, no matter how intense they are; and (iii) that naming these feelings is the first step in getting them under control. This will sometimes be followed by the client expressing a great sense of relief: “Someone finally understands.” The counselor’s silence after that will allow the client to think about what he said and this will often be followed by some of the most significant statements that will be made in the interview.

Finally it should be noted that the counselor’s reflective responses should be such that the client learns that he owns the problem. This is particularly important if the client blames others for the problem. “My child does such terrible things I have no alternative but to physically punish him.” The counselor’s appropriate response is not to suggest alternative punishments at this point but rather to reflect on the client’s feelings: “He must make you very angry.” Thus, the mother may begin to see that it is her anger that led to her being accused of being abusive and not the child’s behavior.

### **3. ISSUES, DILEMMAS, AND CONTRADICTIONS IN THE COUNSELING PROCESS**

The counseling process is a complex one. What makes it different from other forms of interpersonal interaction is

that the counselor responds differently from the typical relative, friend, or neighbor to the person who comes for help. This means she must be particularly alert to her own responses to what the client tells her. This leads to dilemmas and contradictions in the helping process that must be carefully thought through in order to maintain the integrity of the process. Some of these issues, dilemmas, and contradictions are discussed here.

#### **3.1. Values and Ethics**

Each of us has our own sense of what is right and wrong and our own preferences about what is important to us. Our ethics and values to a great extent determine who we are and how others see us. Furthermore, we tend to choose friends whose values and ethics are similar to our own. Some of us believe it is legitimate to use the loopholes in the law to reduce our taxes; others think this is cheating. Some of us prefer to spend our money on a house and car; others prefer to use our funds to give to charity or for political action. However, in most circumstances we are not in a position to influence others about what they think and what they value. Counselors, though, may be confronted with clients whose thinking about these and many other issues is quite different from their own.

What do they do about this? In most situations, if these values and ethics are unrelated to the problems the client brings to them, they do nothing. They have no right or obligation to attempt to change a client’s values or ethics, even if they find them repulsive. If this makes it difficult for the counselor to work with the client, she must refer him to another counselor.

Sometimes these values are related to the client’s problem. Then what? The use of argument or logic hardly ever works to change values, beliefs, or attitudes. Often, the only alternative for the counselor is to help the client see how these values are dysfunctional—how they may get him into trouble. Only when the client can make this connection is he likely to change. However, this can occur only if the counselor has established a relationship of trust with the client. One of the fundamentals of such trust is confidentiality.

We have noted that referral to another counselor may be necessary if the client’s values and ethics are so different from those of the counselor that she finds it difficult to work with him. This may occur when there are major differences in culture and/or socioeconomic class between the client and the counselor. There are two other circumstances in which such a referral may be wise: when the counselor feels hostile to the client

and cannot figure out why and when the counselor believes that she does not have the expertise to be helpful to the client.

### 3.2. Confidentiality

Basic to a client's trust in the counselor is that she will not reveal to others the information given to her. Many helping professionals have what is known as privileged communication, which by law gives the counselor the right to hear almost anything the client says with the responsibility of not repeating it to others. A violation of confidentiality is likely to destroy the helping relationship and may even make it difficult for the client to seek help from any other counselor. Yet, if the client trusts the counselor, he may say things that seem outrageous to relatives and friends. He may tell the counselor that he hates his mother, that he wishes a child was dead, or that he is having an extramarital affair. Again, what makes the counselor special and helpful is that she reacts differently from friends and relatives. Above all, the information remains with the counselor except in the circumstances described next.

There are times when it is not only both ethical and legal to disclose information but also may be required. When a child or elderly person has been abused or neglected, the counselor is required to report this to the local protective service agency. When a client threatens another person with violence, the law requires that the counselor protect the potential victim by hospitalizing the possible perpetrator or informing the potential victim of the danger. Sometimes, it is a good idea for the counselor to discuss the case with a supervisor or to consult with another professional in deciding the legal, moral, and professional dimensions of her actions. Finally, the counselor may be required by the client's insurance company to provide certain information in order for the client to be reimbursed for her fees. However, in each of these instances, before confidentiality is violated, it is essential that the counselor get the client's permission, preferably in writing. This also provides her with the opportunity to discuss the issues related to the disclosure of information with the client and may open up new avenues for helping the client.

### 3.3. Trust

As implied previously, central to a successful or helpful counselor–client relationship is trust. However, this in

itself is a dilemma. Often underlying the problem with which the client comes to the counselor is a lack of trust in those people with whom he has contact. Thus, one of the first and most important tasks of the counselor is to enable the client to learn to trust her.

What is the meaning of trust? In this context, trust means that the client believes that whatever the counselor does is in the client's best interest. Trust means that the client feels that no matter what he says it will not be used against him. Trust means that when the client is angry at the counselor, disagrees with the counselor's attitudes and values, or is ambivalent or indecisive, the counselor will not respond with anger or rejection, belittle, or badger him. Trust means that when the counselor confronts the client about his thoughts, feelings, and intentions, the client knows that the intent is to help him face up to the problems he is experiencing and that the counselor will provide support and encouragement as well. Trust means that the client understands that change is painful but that any pain the counselor inflicts on the client has as its purpose the reduction of the problem for which the client came into counseling.

However, if trust is so important, how does the counselor build it, especially since clients are unlikely to trust the counselor at the beginning of the counseling process? First, the counselor must be open and honest. The counselor must not evade the client's questions or respond with platitudes or try to give false hope. In her own way, she must convey to the client that although her attitudes and values may be different from his, each has a right to what each thinks and believes. The counselor respects the client for whom he is despite any differences there may be between them.

Second, the counselor encourages the client to lead and she is willing to follow, especially at the beginning of the counseling process. She allows the client to tell his "story" in his own way.

Third, the counselor encourages catharsis in order to relieve some of the client's anxiety, while at the same time she provides empathy and support. She listens with sensitivity and compassion, demonstrating by eye contact, body position, facial expression, and occasional words of reassurance that she is listening. However, this too can be a dilemma for the counselor, especially at the beginning of the counseling process. If too much anxiety is relieved through catharsis, the client is likely not to be motivated to return. On the other hand, if there is little or no relief of anxiety, the client is likely to think that the counseling process

is not helpful and again may not return. It is through experience that the counselor learns how to balance this issue in work with clients.

Thus, anxiety is a two-way process that has both negative and positive consequences. Most clients come to us because they have so much anxiety that they need to do something to reduce it. This is the motivation for change. On the other hand, many involuntary clients present a different type of dilemma. Many of them are not likely to be anxious. Many psychopaths and sociopaths do not experience anxiety in the same way most of us do. Without anxiety, there is no motivation for change, and without such motivation change is very difficult if not impossible. Thus, our job as counselors is to increase this type of client's anxiety level. One prominent way to achieve this is to point out the consequences of the client's behavior (e.g., loss of freedom by going to jail for dealing drugs or loss of life by overdosing on drugs). Nonetheless, this is a very difficult process.

Fourth, in order to establish trust, the counselor makes use of the client's priorities. That which the client sees as his most important or immediate problem must be dealt with. This is so even if the counselor believes that there are other problems that the client is experiencing that are more acute or may underlie the problem he brings to the interviewing session. The counselor must respect what the client believes is important.

A fifth method of gaining the client's trust is known as joining. This involves identifying with the client by following some of his behavior and speech patterns. Sometimes, it means mirroring the client's movements (leaning forward when the client does) or matching the client's timing (slowing down one's speech to be in harmony with the client's pace).

Sixth, if the counselor can provide immediate relief for one of the client's problems, trust is likely to occur. We have discussed how catharsis is one way of achieving this, but there are other ways as well. Sometimes, information can be very helpful. An explanation of unemployment or Social Security eligibility requirements can be very helpful with some clients. Explaining what is expected in college application materials to graduating high school students can relieve much anxiety. Referral to another agency or resource, such as a local tenants council or a store that sells reasonable clothing, can be seen as very helpful.

The "a-ha" response is another technique to establish trust. This occurs when the client arrives at a special understanding of his own situation that he did not have

before. It happens when the counselor is able to ask just the right question at the right time or make a crucial observation just when the client is ready to accept this about himself.

Finally, trust can and should be built by giving the client a sense of hope by assuring him that it is possible to change, to give up even the most painful feeling. However, for this to occur, the counselor must have the sense that she can be successful with this client. If the client does not have hope, the counselor must try to understand why and sometimes consult with a supervisor or colleague if this feeling persists. Without this sense of hope, it is highly unlikely that the counseling process can be helpful.

### 3.4. Diversity

Despite the counselor's good will, and her wish to be helpful to others, there are some clients who at first seem to reject the counselor's efforts and seem unwilling to trust her. These are people who are different from the counselor. They may be of a different race, may come from a different ethnic background or culture, or may be of a different gender or belong to a socioeconomic group different from that of the counselor's. They may have a difficult time understanding her, and she may have a difficult time understanding them. Her behavior may seem strange to them, and their behavior may seem strange to her. How can they trust someone who is so different from themselves?

Furthermore, if clients are from an oppressed group they may not only be anxious about the counselor's differences but also angry. Their experiences with "people like us" may have been very negative. The wheelchair-bound professional was denied employment despite his training and experience because of his disability. The black welfare mother was accused of child abuse when the teacher saw a bruise on the child even though it was caused by a fall. The gay or lesbian couple was denied health benefits because they were not legally married, although they have been living together for more than 20 years. The woman earns considerably less than her colleagues despite equal years of experience because of her gender. All of these people have reason to be angry, although the counselor is probably not the source of the anger.

How does the counselor deal with such differences? She must start by being a learner as well as a teacher. When the counselor does not understand the reasons for a client's behavior, it is her responsibility to figure out what is going on. She must not only teach the client

but also be open to learning from him. If the counselor is working primarily with a certain type of client—Puerto Ricans, the blind, or low-income women—it is useful for the counselor to read as much as she can about this group. However, this is not a substitute for asking the client about his behavior because not all clients fit the stereotypes described in the literature. By asking, the counselor not only learns about the client and his special differences but also shows respect for these differences.

However, what about the client who is angry? He may show that anger in a variety of ways; he may have a temper tantrum in the counselor's office, miss appointments, be unresponsive, give superficial answers to questions, or even lie. First, the counselor must remember that his anger is unlikely to be personally against her as an individual but rather who she represents. If she returns the anger, she will lose the client's trust. Rather, she needs to focus on the client's feelings, trying to determine the source of his rage. The counselor must let the client know that she understands that it may take him considerable time to trust her but that she is patient. She must try not to be defensive, such as by reassuring the client that she is not like other people. Although none of this may work for awhile, client anger has a way of wearing down, leaving the counselor a way of reaching the client.

### 3.5. Listening

The heart of the counseling process is the ability of the counselor to listen. She not only listens to the content of what the client is saying but also listens to how it is said. Is the client's speech a monotone or emotional, fast or slow, loud or quiet? Does it change depending on the topic being discussed? Does the tone of voice convey the image of a child, a seductive woman, an angry parent, or a sense of hopelessness or helplessness? Listening includes observation of the client as he speaks. Is there eye contact? Are there changes in facial expression or body position? All of these are clues about what is going on in the client's mind, and the client's thought processes provide the counselor with some understanding of the sources of the client's problem. The counselor must put all of this information together to assess the client's strengths and weaknesses and to plan with the client how she can best be of help to him. As she listens, the counselor must decide what response may be helpful to the client.

The counselor's responses are often different from those that might be expected from friends, neighbors,

and relatives. The counselor raises questions that the client had not thought about before. The counselor makes statements that challenge the client's way of thinking. The counselor confronts the client's assumptions about himself and others. The counselor remains silent when the client expects a vigorous rebuttal. These are the responses that make the counseling relationship unique, and these are the behaviors that build trust between the client and counselor.

### 3.6. Silence

As unusual as it may seem, silence is an important part of the counseling process. Silence by the client has meaning that the counselor must try to understand. Silence by the counselor can be used to help the client in a variety of ways.

The client may be silent just because something he said or the counselor said has led him to think in new ways, to see new meaning or make connections he did not make before. He may be sorting out thoughts and feelings. This is a time when the counselor should also remain silent, as difficult as this may seem. When the client speaks again, a significant breakthrough in his thinking may occur.

The client may be silent because he is feeling deep emotion. The counselor lets the client experience this but not drown in it. She may ask, "What are you feeling right now?" The client may also be silent because he is confused. The counselor can interrupt the silence by indicating that what she said may be confusing and then try to clarify what has happened in the conversation.

The client may be silent because he is resisting the counselor's probing and/or rejecting the counselor's authority. There is tension in the air. The counselor's response might well be that both of them seem to be uncomfortable and maybe they should talk about what is causing the tension between them.

Thus, silence is not to be feared in the counseling process. Rather, it is to be understood and used by the counselor as a way of moving the helping process forward.

## 4. CONCLUSION

The counseling interview is a complex process that requires considerable training and experience by the counselor to be effective. Although the previous sections discussed the importance of listening and

silence by the counselor, it must not be forgotten that she must be in control of the interview to be effective. It is the counselor's ability to respond—and sometimes not respond—in ways that are not expected by the client that enables her to see the problem differently and deal with it more effectively. When successful, the counseling process can not only be of great aid to the client but also provide much satisfaction to the counselor.

### See Also the Following Articles

Counseling and Culture ■ Ethics and Social Responsibility  
■ Group Counseling ■ Measurement and Counseling  
■ Trust

### Further Reading

- Benjamin, A. (1981). *The helping interview*. Boston: Houghton Mifflin.
- Devore, W., & Schlesinger, E. G. (1991). *Ethnic-sensitive social work practice*. Columbus, OH: Merrill.
- Donner, S., & Sessions, P. (1995). *Garrett's interviewing: Its principles and methods* (4th ed.). Milwaukee, WI: Families International.
- Fine, S. F., & Glasser, P. H. (1996). *The first helping interview: Engaging the client and building trust*. Thousand Oaks, CA: Sage.
- Lum, D. (1996). *Social work practice and people of color: A process stage approach* (4th ed.). Pacific Grove, CA: Brooks/Cole.
- Shulman, L. (1998). *Skills of helping individuals, groups, and communities*. Itasca, IL: Peacock.



# Cross-Cultural Psychology, Overview

J. W. Berry

Queen's University, Kingston, Ontario, Canada

Harry C. Triandis

University of Illinois at Urbana-Champaign,  
Champaign, Illinois, USA

1. Definition of Cross-Cultural Psychology
  2. Definition of Culture
  3. History of Cross-Cultural Psychology
  4. Methodology
  5. Dimensions of Cultural Variation
  6. Evaluation of Findings and Applications
- Further Reading

## GLOSSARY

**cross-cultural psychology** The systematic study of behavior and experience as it occurs in different cultures, is influenced by culture, or results in changes in existing cultures.

**cultural psychology** The intensive study of psychological processes in one culture, which identifies how features of the culture affect these psychological processes.

**indigenous psychology** The systematic study of the culturally unique concepts and processes found in a given culture.

All human behavior is influenced by the culture in which a person develops. Thus, there can be no complete account of psychological phenomena without taking the cultural context into account. This claim applies not only to the findings of psychological researchers, but also to practitioners, who must

consider the cultural settings of their applications as well as the cultural roots of the behavior.

## 1. DEFINITION OF CROSS-CULTURAL PSYCHOLOGY

According to Triandis and Lambert (1980), "Cross-cultural psychology is the systematic study of behavior and experience as it occurs in different cultures, is influenced by culture, or results in changes in existing cultures." And per Berry *et al.* (1997), "The field is diverse: some psychologists work intensively within one culture, some work comparatively across cultures, and some work with ethnic groups within culturally plural societies."

Some think of culture as an independent variable that influences psychological processes, and some emphasize that culture and psychological processes are mutually constituted, i.e., "make each other up." Berry argued that culture is both outside and inside the individual. It is independent of the individuals who arrive into a culture and become incorporated into a culture through the process of enculturation; but individuals also incorporate the culture through the same process, so that it also becomes an organismic variable, ready to be transmitted anew to other arrivals. In 2000, a special issue of the *Asian Journal of Social Psychology* discussed the

similarities and differences between indigenous, cultural, and cross-cultural psychology, which reflect these differences in point of view.

## 2. DEFINITION OF CULTURE

The conceptualization of culture is by no means a simple matter. One possible way to think about culture is that “culture is to society what memory is to individuals,” as per Kluckhohn (1954). It includes what has worked in the experience of a society, so that is what was worth transmitting to future generations. It includes both objective elements, such as tools, bridges, chairs, and tables, and subjective elements such as norms, roles, laws, religions, and values.

Sperber used the analogy of an epidemic. A useful idea (e.g., how to make a tool) is adopted by more and more people and becomes an element of culture. Barkow, Cosmides, and Tooby (1992) distinguished three kinds of culture: metaculture, evoked culture, and epidemiological culture. They argued that “psychology underlies culture and society, and biological evolution underlies psychology” (p. 635). The biology that has been common to all humans, as a species distinguishable from other species, results in a metaculture that corresponds to panhuman mental contents and organization. Biology in different ecologies results in “evoked culture” (e.g., hot climate leads to light clothing), which reflects domain-specific mechanisms that are triggered by local circumstances and leads to within group similarities and between groups differences. What Sperber describes Barkow *et al.* call “epidemiological culture.”

Elements of culture are shared standard operating procedures, unstated assumptions, tools, norms, values, habits about sampling the environment, and the like. Because perception and cognition depend on the information that is sampled from the environment and are fundamental psychological processes, this culturally influenced sampling of information is of particular interest to psychologists. Cultures develop conventions for sampling information from the environment and for how much to weigh the sampled elements. For example, people in hierarchical cultures are more likely to sample clues about hierarchy than clues about aesthetics. Triandis argued that people in individualist cultures, such as those of most cultures of North and Western Europe and North America, sample elements of the personal self (e.g., “I am busy, I am kind”) with high probability. People from collectivist cultures, such as those of most cultures of Asia, Africa, and South America, tend to

sample mostly elements of the collective self (e.g., “my family thinks I am too busy, my co-workers think I am kind”).

Just as individuals are incorporated into their primary culture through the process of enculturation, people may also become members of more than one culture through the process of acculturation, depending on life experiences.

### 2.1. Relationships with Other Perspectives

Cross-cultural psychology derives from both anthropology and psychology, taking various concepts, theories, and methods from each discipline. Because of this, the field is not unitary, but is composed of various perspectives. In our view, cross-cultural psychology combines two broad traditions. First, we use the comparative method (the “cross”), making assumptions about the comparability of behavior (see below) and employing research and statistical methods drawn primarily from individual psychology. Second, the field is concerned with the unique sociocultural contexts in which people develop and act (the “cultural”), employing observational and narrative methods drawn primarily from cultural anthropology. The “cultural psychology” and “indigenous psychology” schools of thought emphasize this second perspective. However, in practice, both perspectives employ individual data, contextual data, and the comparative method in their work.

## 3. HISTORY OF CROSS-CULTURAL PSYCHOLOGY

The first person to write anything relevant about this topic was probably Herodotus, during the 5th century BC. There are three histories of the field that start with Herodotus, who had the insight that all humans are ethnocentric. This is necessarily an aspect of the human condition, because most humans are limited to knowing only their own culture and thus are bound to use it as the standard for comparing their culture with other cultures. It is unfortunate that this is necessarily the human condition, because it makes people think that their norms and values are universally valid, and any deviations from them is seen as not only wrong but also immoral. Much human conflict can be traced to ethnocentrism. It is only when humans have experienced several other cultures that they

become sufficiently sophisticated to see both the strengths and weaknesses of every culture and thus to become less ethnocentric.

Jahoda provided a detailed history of the field in which he discussed the work of Vico in the early 19th century, Wundt in the early 20th century, Tylor, Rivers, Boas, Luria, Bartlett, Benedict, Vygotsky, and others. These writings were not central to the activities of most psychologists until the late 1980s and early 1990s, when many of the findings of cross-cultural psychology began challenging some of the theories of mainstream psychology.

The field became a distinct field of psychology in the early 1970s, when awareness of the need to stop being ethnocentric in research, and the need to develop measurements of psychological constructs that are both culturally sensitive and equivalent across cultures, became acute. Avoiding intellectual colonialism, such as the exploitation of data from other cultures without giving credit to local psychologists, resulted in an effort to develop a code of ethics for cross-cultural research. However, this code was not adopted by any association of psychologists because it was felt that it was overly strict and unrealistic and would inhibit research.

In the 1970s, most psychologists used an absolutist viewpoint. That is, it was assumed that all psychological discoveries were valid everywhere, and culture provided only minor modifications of these discoveries. In the 1980s, however, a number of researchers assumed more relativistic positions, in some cases even denying the psychic unity of humankind. Nowadays, a universalistic perspective is taken by many cross-cultural psychologists.

Journal publications with a cross-cultural psychology focus started in 1966. Leonard Doob, editor of the *Journal of Social Psychology*, and Germaine de Montmollin, editor of the *International Journal of Psychology*, announced that they would favor cross-cultural articles. The Human Relations Area Files started a journal that year that eventually became *Cross-Cultural Research*. It is now the official organ of the Society for Cross-Cultural Research. The *Cross-Cultural Newsletter* (first edited by Triandis) was launched in 1967. It eventually became the *Cross-Cultural Bulletin of the International Association for Cross-Cultural Psychology*. A directory of cross-cultural psychologists was first published by Berry in 1968. The *Journal of Cross-Cultural Psychology* (Walter Lonner, editor) started in 1970 and became the official organ of the International Association for Cross-Cultural Psychology. The Association pour la Recherche Interculturelle was established in 1984, and included

French speaking cross-cultural psychologists. The Annual Review of Psychology introduced chapters reviewing cross-cultural work in 1973. The first edition of the *Handbook of Cross-Cultural Psychology* (Triandis, editor) appeared in six volumes in 1980–1981. The second edition (J.W. Berry, editor) appeared in three volumes in 1997.

### 3.1. Theory

The term absolutist was used to refer to a theoretical position that assumes that psychological findings are valid for all people everywhere; culture is thought to play little or no role in the development or expression of human behavior. Two other theoretical perspectives have been differentiated from absolutism. Relativism is essentially the opposite point of view: culture is so important in understanding human behavior that no psychological research or application can be done without knowing, and taking into account, the cultural context. An intermediate position (universalism) takes the position that human psychological processes are common, species-shared qualities; culture plays important variations on these basic processes during behavioral development and expression. In addition, absolutism tends to make comparisons freely; relativism generally avoids them; and universalism makes judicious comparisons (based on the underlying commonality) while taking care to understand the influence of culture. In a sense, the basic commonality makes comparisons possible (see discussion of equivalence, below), while the cultural variation in expression makes them interesting and worthwhile. In terms of the other perspectives or schools discussed previously, the cultural and indigenous approaches tend to be relativist, while the cross-cultural approach tends to be universalist. (But few psychologists who read this article are likely to be absolutist!)

## 4. METHODOLOGY

### 4.1. Levels of Analysis

Cross-cultural psychologists often work at different levels of analysis. For example, Hofstede has collected data from over 100,000 participants in more than 40 countries. By 2001, the data set was expanded to 50 countries. Responses to a sample of 40 and later 14 goals/value items were collected. He summed the responses to each goal/value item by the participants from each country. Then he factor-analyzed the 40 by

40 (or 14 by 14) matrix, based on 40 or 50 means. This is an ecological analysis and is very different from a within-culture analysis, in which the matrix of correlations is based on the number of participants from that culture.

Readers may need an illustration of the fact that results at the cultural (or group) and individual levels of analysis can be different. Lincoln and Zeitz studied 500 employees divided among 20 social service agencies. They found that supervisory duties were positively correlated with professional qualifications across individuals but negatively correlated across agencies. The latter correlation was due to the fact that the members of the more professional agencies did not need much supervision.

#### 4.2. Comparability and Equivalence

As was noted above, there are different viewpoints regarding whether comparisons are legitimate: relativists tend to object to them, absolutists tend to make them without restriction, and universalists employ them with caution, once certain conditions are met.

All acts of comparison (whether they be of cultures, institutions, or behaviors) require some common dimension along which phenomena can be judged. Apples and oranges can indeed be compared: on sugar content, juiciness, or price. The presence of such commonalities is a prerequisite for making a comparison; however, it is not enough. Cross-cultural psychologists seek to establish the comparability of two (or more) phenomena, using a number of ideas and techniques. One of these is equivalence, in which conceptual meaning and psychometric properties can be demonstrated to be sufficiently similar between two sets of data to be compared. It is no longer accepted practice to simply use an instrument, obtain a score, and apply a test of significance to the difference. While this has been known for a long time, it is still ignored by some writers who submit papers to the editors of psychological journals.

#### 4.3. Emics and Etics

One approach to better controlling cross-cultural comparison is with the use of the concepts of emics and etics. All cultures categorize, but the categories that are used are culture specific. For instance, our eye is capable of discriminating 7,500,000 million colors, but we get along with a couple of dozen color names. Through the use of Munsell color chips, which provide an objective domain to which we can compare the way people in different cultures categorize colors, we can

determine that the categories are not identical. The stimuli at the core of major color categories such as “red” are the same, but the periphery is not. So, even when we deal with something as simple as the identification of colors we must take cultural specificity into account. All categories are culture specific. We call these emic, from Pike’s use of phonemics and phonetics. But while most categories are emic, there are elements that are universal, which we call etic. In short, if we stick only to the core of each color range (the etic part) we can use the word “red” and its translations across cultures with reasonable equivalence. But there are regions of the category that are not equivalent across cultures.

Furthermore, there are languages that have very few color names. Some languages have only two words that are relevant to color that are more or less equivalent to white and black. When a language has three color names, it adds a word for red. If it has four or five names, then it also has words that are more or less equivalent to yellow and green. Languages that have progressively richer color vocabularies have a word for blue, then for brown, finally for purple, pink, orange, and gray.

People who do not distinguish between green and blue have difficulties in discussing differences in color between grass and the sky. If we study this domain and assume that they see color the way we do, we are likely to obtain distorted results. To avoid this problem, when we do a study we should talk to many people from each language to see how they cut the pie of experience. This would identify their emics for dealing with the particular topic. When we do that, some of their categories may also appear in other languages. Then, they may in fact be etics.

When we construct instruments such as tests or interview questions, or scales, we need to incorporate both emics and etics. The emics allows us to measure a phenomenon with cultural sensitivity, “the way the natives see it.” The etics allows us to compare the cultures. For example, a scale of social distance was constructed for Greeks and Americans in 1962. The scale values were standardized separately for each culture, and were made equal to zero for “I would marry this person” and 100 for “As soon as I have a chance I am going to kill this person.” The item “I would accept this person as a roommate” made sense only in the United States and had a scale value of 29.5. The item “I would accept this person in my *parea*” was a Greek emic, and had a value of 31.1. A *parea* is an intimate group of friends that meets almost daily to enjoy things together. “I would rent a room from this person” had a scale value

of 57.5 in the United States and 42.8 in Greece, suggesting that in Greece it was a more intimate activity. In a culture such as India, where touching is a very important issue, one would use emic items such as “I would allow this person to touch my earthenware.” In a culture that has residential segregation, an item such as “I would allow this person to live in my neighborhood” would be a useful item to measure social distance.

We also need to consider the fact that often a construct used in the West has very different meanings in other parts of the world. For example, when people in the West use the concept of “intelligence,” they include the ideas of speed, successful adaptation, and effective completion of tasks. In parts of Africa, people think that to be intelligent means to be slow, sure, and not make mistakes. In other parts it includes the ideas “knows our traditions” and “does what the elders expect.” Among the Cree, a Native American tribe, “lives like a white” is conceptually close to “stupid.”

#### 4.4. Translation

In the construction of instruments that can be used equivalently across cultures, researchers need to use translation. This is a complex process.

### 5. DIMENSIONS OF CULTURAL VARIATION

As has just been noted, comparisons require common dimensions on which they can be made. Beyond this simple dictum is a complex reality: there are many possible dimensions that could be employed in comparing cultures, institutions, and behaviors.

When we think of the relationship between culture and psychology, it is useful to use a metaphor from geography. We can discuss continents by describing mountain ranges, rivers, bays, etc. We can discuss countries by describing similar features, cities, villages, roads, etc. We can describe cities by describing roads, buildings, etc. We can describe villages by specifying the location of every building in relation to every other building. We can describe neighborhoods by discussing the economic, social relationships, religions, and political organizations of the residents, etc. Finally, we can discuss households by giving the biography of each member of the household, etc.

Note that as we move to lower levels of abstraction, we need more dimensions, details, specifications.

Psychologists deal with phenomena that are based on the common biology of humans, as it adapts to different environments (continents). Social psychologists also examine different environments (ecologies), for instance, social relationships in which people influence each other a great deal, as one finds in agricultural societies, or relatively little, as one finds among hunters and gatherers. Then, different behaviors can be identified. For example, there is more conformity in agricultural than in hunting environments. The societies that are agricultural have more inequality than the societies that engage in hunting and gathering.

Different disciplines focus on different levels of analysis. For example, literary productions describe events at the level of households. Anthropologists describe neighborhoods. Sociologists tend to describe cities and villages. We believe that all this work is valuable. All the information is necessary for a balanced understanding of the culture and psychology relationships.

In this section we focus on some of the dimensions of cultural variation that correspond to the level of continents. The data are obtained from different countries, though this is a very rough procedure because each country contains hundreds of cultures. Yet because regions of the world (e.g., Europe versus East Asia) have something in common, we can discuss the ways in which they are similar and different. Other work that we have no space to discuss is done at the other levels of analysis and is also valuable. The dimensions of cultural comparison are also called “cultural syndromes.”

#### 5.1. Complexity

We mentioned previously the contrast between hunters and gatherers and agricultural societies. Agricultural societies are more complex in the sense that they have more types of roles than hunting and gathering societies. As we move to modern information societies we find even more complexity, for instance, many more roles than in more simple cultures. If we consider all forms of specialization, e.g., the difference between medical specializations, in complex societies there are more than a quarter of a million roles.

#### 5.2. Tightness and Looseness

Societies differ in the number of norms concerning correct behavior. Some have many norms, e.g., how to bow and how to smile, and others have few norms. Furthermore, some cultures punish severely those who deviate from norms whereas others are lenient. In tight

cultures, there are many norms, and those who do not follow the norms are criticized or even killed. In loose societies, there are not many norms, and people who do not behave according to the norms are not punished. People are very likely to say “It does not matter” when they see a deviation from a norm. An example of a tight culture was the Taliban in Afghanistan. They killed people in large numbers for relatively minor deviations from arbitrary norms that they imposed on the basis of supposedly religious guidance. Some of these norms, such as not listening to music and not watching Western television, had nothing to do with religion. An example of a loose society, on the other hand, is Thailand, where people tend to smile even when employees do not show up for work as expected. The imposition on norms depends of the domain or the situation. For example, in the United States there is tightness concerning how to deal with banks, your checkbook, loans, mortgages, and the like. But now there is considerable looseness concerning who you can live with. Of course, there are also major individual differences. Some people are compulsive about their own and others’ behavior, while others are very lenient (e.g., any behavior is okay; it’s none of my business to criticize this behavior).

Hofstede identified a dimension he called uncertainty avoidance, which is related to tightness. Countries high on this dimension were Greece, Portugal, and Guatemala; low were Denmark, Jamaica, and Singapore. Subjective well-being was negatively correlated with this dimension, presumably because people in tight cultures are anxious about being criticized or punished for behavior that is not perfectly “correct.”

### 5.3. Collectivism and Individualism

Societies that are both simple and tight tend to be collectivist. Societies that are both complex and loose tend to be individualist. Collectivism is a cultural pattern found especially in East Asia, Latin America, and Africa. It is usually contrasted with individualism found in the West, e.g., in Western and Northern Europe, the United States, Canada, Australia, and New Zealand. Note, however, that any typology is an oversimplification; societies are not purely individualist or collectivist but some mixture of the two. The construct has been widely used in the study of cultural differences. Unfortunately, however, most of the research compared samples from North America versus East Asia. Thus, in the description of the attributes of the construct and its antecedents and consequents we rely too heavily on the differences between the United States and Canada on

the one hand and China, Korea, and Japan on the other hand. We do not know if the information we present below applies equally to Africa and Latin America, though we suspect that much of it does. Much more research is needed on this topic.

Among the most important characteristics of the collectivist cultural pattern are the following:

1. Individuals define themselves as aspects of a collective, interdependent with some ingroup, such as one’s family, tribe, co-workers, or nation, or a religious, political, ideological, economic, or aesthetic group.
2. They give priority to the goals of that collective rather than to their personal ones.
3. Their behavior is determined more often by the norms, roles, and goals of their collective than by their personal attitudes, perceived rights, or likes and dislikes.
4. They stay in relationships even when the costs of staying exceed the advantages of remaining.

These are the defining attributes of collectivism. There are as many kinds of collectivism as there are collectivist cultures. To distinguish among collectivist cultures additional attributes are necessary.

Triandis summarized the literature that used that construct in social psychology, and Kagitcibasi provided a critical evaluation of the construct. Hofstede found that the United States, Australia, and Britain were highest in individualism and Panama, Ecuador, and Guatemala were highest in collectivism. Most Western countries are individualist and most East Asian countries are collectivist. However, gross national product per capita is correlated with individualism; thus, wealthy East Asian countries such as Japan are not as collectivist as one would expect from their geographic location.

Within any society there are individuals who behave like persons in collectivist cultures, called allocentrics. They contrast with those who behave like persons in individualist cultures, who are called idiocentrics. There are both allocentrics and idiocentrics in every society, but their distributions are different, with more allocentrics found in collectivist cultures.

Collectivism is maximal in relatively homogeneous societies, such as theocracies and monasteries, while individualism is maximal in heterogeneous societies that are very affluent. Thus, there will be few idiocentrics in monasteries and few allocentrics among Hollywood stars.

All individuals have access to cognitive systems that include both allocentric and idiocentric cognitions, but they sample them with different probabilities, depending on the situation. For example, if the ingroup is under

attack, most individuals become allocentric. In the company of other allocentrics, the norms for allocentric behavior become salient, and individuals are more likely to sample allocentric cognitions. Some situations provide very clear norms about appropriate behavior (e.g., in a house of worship), while other situations do not (e.g., at a party). Individuals will be more allocentric in the former than in the latter situations. When the ingroup can supervise an individual's behavior, norms are more likely to be observed, and the individual will be more allocentric. In one study, allocentrics and idiocentrics were randomly assigned to simulated organizations that were individualist or collectivist, and the degree of cooperation exhibited by the individuals was studied. It was found that cooperation was high when allocentrics were in a collectivist organization. In the other three cells of the experiment, there was very little cooperation. In short, both the personality and the situation were required to predict the behavior.

Collectivism can appear in all or none of the domains of social life. For example, it can be found in politics, religion, aesthetics, social life, economics, or philosophy, as was the case in China during the Mao period, or in none of these domains, as among Hollywood stars.

### 5.3.1. Antecedents of Collectivism

Some of the antecedents of collectivism are listed in Table I and are discussed below:

*Homogeneity of the collective.* If people disagree about the norms of proper behavior, or the goals that people should have, it is difficult for people to behave according to the norms of the group.

TABLE I  
Some Antecedents of Collectivism

---

<ul style="list-style-type: none"> <li>• Homogeneity of the collective</li> <li>• Low exposure to other cultures</li> <li>• Interdependence is adaptive to ecology</li> <li>• Population density</li> <li>• Social stability</li> <li>• Low economic status</li> <li>• Large families</li> <li>• One large ingroup rather than many ingroups</li> <li>• Salient common fate</li> <li>• Collective mobility is advantageous</li> <li>• Strict religious upbringing</li> <li>• Little exposure to modern mass media</li> </ul>
--

---

*Low exposure to other cultures.* People who know only one culture tend to be maximally ethnocentric, authoritarian, and submissive to ingroup authorities. Those who are more educated, traveled, and have lived with more than one cultural group develop idiocentric tendencies.

*Interdependence is adaptive to ecology.* People are more interdependent in agricultural societies than in information societies. For example, when the goal is to complete large projects such as irrigation canals or defensive walls, collectivism is more likely. In societies where people are financially interdependent, collectivism is high. People who can do their job when they are alone are more likely to be idiocentric.

*Population density.* In dense social environments, many rules that are designed to reduce conflict and ensure the smooth functioning of the group develop.

*Social stability.* When the collective is stable, it is more likely to develop agreements about norms, and to make sure that the norms are observed. There is evidence that the older members of all societies are more allocentric than the younger members.

*Low economic status.* The lower social classes tend to be more conforming to social norms than members of the upper classes. When resources are limited, one often depends on group members for assistance, especially in emergencies. These factors increase collectivism. On the other hand, in all cultures, those in positions of leadership tend to be idiocentric.

*Large families.* In large families it is not practical to allow each child to follow idiosyncratic schedules or to have much privacy. Many rules are enforced, and that creates collectivism.

*One large ingroup rather than many ingroups.* Those who only have one ingroup can channel all their energy into that group. Also, they cannot afford to develop poor relationships with members of that group, so they are more likely to observe its norms.

*Salient common fate.* Common fate with members of the ingroup (e.g., when the ingroup is under attack) increases collectivism. Time pressure for decisions has similar consequences.

*Collective mobility is advantageous.* If individual upward social mobility is not possible, then collective mobility may be used. Thus, individuals invest their energy in promoting the status of their ingroup.

*Strict religious upbringing.* Most religions require observance of a large number of norms and threaten to punish those who ignore these norms. That increases collectivism.

*Little exposure to modern mass media.* U.S.-made television is widely available throughout the world.

Content analyses show that the themes used are highly individualistic (e.g., emphasis on pleasure, doing what the individual wishes to do even if that is inconsistent with the wishes of authorities). Countries where people have little exposure to Western mass media are more collectivist.

### 5.3.2. Consequences Of Collectivism for Cognition, Emotion, and Motivation

Marcus and Kitayama have reviewed much evidence showing the influence of collectivism on cognition, emotion, and motivation (see Table II).

*Thinking.* Nisbett, Peng, Choi, and Norenzayan examined a broad range of literature concerning thought patterns in ancient Greece and China, two cultures that correspond to the individualism–collectivism contrast. They argued that the Greeks thought analytically and the Chinese holistically. The Greek traditions influenced the West, which now uses logic, rules, and categories and sees the world as a collection of discrete objects, while the East sees the world as a collection of overlapping and interpenetrating substances. They reviewed much empirical work that contrasted Western and Eastern cultures, and found that the West focuses on objects and their attributes, while the East focuses on the context of objects and the relationships of objects and events to each other. Western thought

is linear, concerned with abstract analysis, and pays much attention to contradictions. Eastern thought is dialectical, experience-based, and pays little attention to contradiction. The West is more field independent; the East more field dependent. Explanations of events in the West focus on the attributes of the object, e.g., the attitudes, beliefs, and values of individuals as determinants of behavior; in the East, the focus is on the context of the object, e.g., group memberships, norms, and social pressures that influence the behavior of individuals. Numerous other cognitive patterns contrast these two kinds of cultures and show not only quantitative but also qualitative differences in thinking.

*Attributions.* Allocentrics make external attributions in explaining events, whereas idiocentrics show the opposite pattern. Thus, the idiocentrics make the fundamental attribution error (i.e., seeing internal rather than external causes for the behavior of other people, while the actors themselves see external causes as explanations of their own behavior more often than internal causes) more frequently than do allocentrics. Also, allocentrics attribute their successes to the help they have received from others and their failures to personal shortcomings (e.g., I did not try hard enough). Allocentrics tend to think that social environments are stable and individuals are easy to change so that they can fit into the environment; idiocentrics think that individuals are stable and social environments are easy to change.

*Emotions.* Members of collectivist cultures get angry when something unpleasant happens to ingroup members more often than when something unpleasant happens to themselves. They do not express negative emotions as often as do members of individualist cultures.

*Description of self.* When asked to complete 20 statements that begin with “I am . . .” members of collectivist cultures provide sentence completions that have more social content (e.g., I am a brother, I am a member of the Communist party) than do members of individualist cultures.

*Socialization.* In collectivist cultures, children are socialized to be obedient, be reliable, follow traditions, do their duties, sacrifice for the ingroup, and be conforming. An important value is self-control.

*Low well-being.* While members of all cultures are satisfied with their life, members of collectivist cultures are less satisfied than members of individualist cultures. Well-being tends to be correlated with individualism. This relationship persists even when statistically controlling for income.

**TABLE**  
Consequences of Collectivism

---

- Thinking
- Attributions
- Emotions
- Describe the self by using roles, groups
- Consequences of socialization
- Low well-being
- Social behavior
- Receive much social support if they conform
- Have poor skills for entering new groups
- Are easy to mobilize for confrontations with outgroups
- Have the ideal of fitting-in rather than shining
- Show little social loafing
- Recreation occurs in stable groups
- Communication
- Conflict resolution
- Conformity
- Apologize more for transgressions of the ingroup

---

*Social behavior.* Collectivists are extremely concerned with “saving face,” and even attempt to save the face of the persons with whom they are interacting. Preserving harmony within the ingroup and being pleasant in relationships with ingroup members are often important values. On the other hand, collectivists do not care much about their relationships with outgroups. They usually behave quite differently toward ingroup versus outgroup members, showing sacrifice and extreme cooperation with ingroup members and suspicion and hostility toward outgroup members.

*Social support.* Allocentrics report that they receive more social support and a better quality of social support than do idiocentrics. However, those who do not conform to ingroup norms are rejected, and may even be killed.

*Poor skills for entering new groups.* Members of collectivist cultures do not have good skills for entering new groups, because they are not socialized to do this. They are comfortable as members of their ingroups, thus they do not need to enter other groups.

*Easy to mobilize for confrontations with outgroups.* When ingroups are in conflict it is easier to arouse members of collectivist cultures to fight for the ingroup than it is to do so with members of individualist cultures. Ethnic cleansing is more often found in collectivist than in individualist cultures.

*Have the ideal of fitting-in rather than shining.* Allocentrics prefer to fit in and be like most others in their ingroup rather than shining and sticking out. Idiocentrics tend to be high in self-enhancement (i.e., report that they are better than most other people on most desirable attributes). Allocentrics tend to be modest when they compare themselves to others.

*Show little social loafing.* Allocentrics show little evidence of social loafing (i.e., the tendency to produce less when working in a group versus when working alone, if their output is not identifiable). They usually produce as much as possible when they work with ingroup members, but they do show social loafing if their co-workers are outgroup members.

*Recreation.* Members of collectivist cultures seek recreation in groups to a greater extent than members of individualist cultures (e.g., bowling rather than skiing).

*Communication.* Members of collectivist cultures rely on the context (e.g., distance between and position of bodies, eye contact, gestures, tone of voice) more than on the content of communications. For example, rather than saying “no,” they may serve incongruous foods in order to put across their negative message without

causing the other person to lose face. When they describe a person they are likely to use context, e.g., “He is intelligent in the marketplace,” rather than decontextualized statements, such as “He is intelligent.” Silence is used more in collectivist than in individualist cultures. Compliments are not used as frequently as in individualist cultures. The languages of collectivist cultures do not require the use of “I” or “you.” Collectivists use more action verbs (e.g., he asked for help) rather than state verbs (e.g., he is helpful), which suggests that they use communications that make greater use of context, are more concrete, and place less emphasis on the internal attributes of the person.

*Conflict resolution.* Members of collectivist cultures avoid conflict with ingroup members, and in the case of disagreement they prefer silence to argument. They prefer mediation to confrontation.

*Conformity.* Members of collectivist cultures show more conformity than members of individualist cultures in experiments that measure conformity.

*Apologize more for transgressions of the ingroup.* Members of collectivist cultures apologize more when an ingroup member commits a crime than is common in individualist cultures.

### 5.3.3. Evaluation of the Construct

While people in collectivist countries indicate that they are less happy than people in individualist countries, individualism has been found to be associated with poor marital adjustment, thus high divorce rates, and more delinquency, crime, drug abuse, suicide, experimenting with early sex, and stress.

Naroll has reviewed much evidence that indicates that tight social groups are needed for social control and the good life. Diener and Suh suggest that “real happiness” depends on having many small positive experiences rather than having a few very intensive positive experiences. Thus, while people in collectivist countries may feel unable to do “their own thing” and thus report low levels of happiness at the moment they are asked if they are happy by the researcher, they may be happier in the long run. Thus, each cultural pattern has both positive and negative aspects.

## 5.4. Power Distance

Hofstede identified a dimension of cultural variation he called power distance. Cultures high on this dimension have members who see a large distance between those

who have power and those who do not. Thus, for instance, people in such cultures fear disagreement with their superiors. Countries high on this dimension were Malaysia, Guatemala, and Panama. Countries low were Denmark, Israel, and Austria. This dimension was highly correlated with collectivism. It was also correlated with the level of corruption of the country.

### 5.5. Sex Differentiation

Hofstede has identified a dimension he calls masculinity–femininity. In feminine cultures, members of the culture attach more importance to relationships, to helping others, and to the physical environment than people in masculine cultures. In masculine cultures, people emphasize careers and money. The goals of men and women are more differentiated in masculine than in feminine cultures. National samples that emphasized male goals were found in Japan, Austria, and Venezuela, whereas the Scandinavian countries and the Netherlands emphasized female goals. In the countries that were high in masculinity, men and women gave very different responses to Hofstede's value questions, while in the countries that emphasized female goals, the answers obtained from men and women were the same. Similar differences were observed across occupations, with engineers giving masculine and secretaries giving feminine answers. The consequences of this cultural difference included such matters as people in masculine countries wanting brilliant teachers versus people in feminine countries wanting friendly teachers. Differences in politics were also identified. In masculine countries people supported tough policies toward poor people and immigrants. Economic growth was given priority over preservation of the environment in masculine countries. In feminine countries, compassionate policies toward the weak and the environment were given high priorities.

### 5.6. Dealing with Time

There are cultural differences in the way time is viewed and in the speed of life. People in countries that are rich, individualistic, and in cold climates move fast and arrive for appointments at the correct time. Those in poor, collectivist countries located in hot climates tend to move slowly and arrive for appointments late. For example, because in collectivist cultures interpersonal relations are very important, if a person who is on her way to an appointment meets a friend, the chances are that she will spend much time with the friend and be very late for the appointment. But it will not be

necessary for her to apologize for being late, as most likely the other people at the appointment will also arrive late.

In all countries, cities are faster than rural environments. Levine devised a number of methods that objectively measured speed; for example, how fast do people walk in the street, how long does it take to mail a letter, how accurate are the clocks in public places. These measures were intercorrelated, showing concurrent validity. He took several samples on each measure in 31 countries. Switzerland, Ireland, and Germany were the fastest and Brazil, Indonesia, and Mexico the slowest countries.

Hofstede introduced a measure of long- versus short-term orientation. A long-term orientation was found in East Asia, among countries with a Confucian tradition, while a short-term orientation was found in the Philippines, Nigeria, and Pakistan.

### 5.7. Missing Topics

The relationship between culture and psychology is so extensive and complex that in this article we were able to present only an introduction. Cultural differences in trust, although not discussed here, are especially important because in countries that are low in trust it is difficult to have economic development. Low trust is frequently found in collectivist cultures where outgroups are often seen as enemies. Because most people belong to outgroups, people in low-trust countries feel surrounded by enemies. A study by Bond, Leung, and 60 other researchers identified a factor that is related to distrust. Countries such as Pakistan and Thailand were high on this factor whereas most of the European countries were low. It is not clear at this time if economic development results in high trust or if high trust leads to economic development. It is likely that the relationship is reciprocal.

## 6. EVALUATION OF FINDINGS AND APPLICATIONS

Approximately 20 years ago, the World Health Organization (WHO) approached the International Association for Cross-Cultural Psychology with what appeared to be a rather simple question: "What do you know, as cross-cultural psychologists, that can help the WHO to achieve its goal of Health for All by 2000?" A number of meetings were held to discuss, and to try to answer, this question. The result was a volume of topical appraisals and an evaluative overview of them. Dasen

*et al.* concluded that there was considerable research-based knowledge, but little that was readily applicable. Two reasons were advanced to explain this discrepancy. First, most knowledge acquisition is driven by a research orientation more than (even rather than) a view to eventual application: at the end of a project, there is often the question, “What do we do with all this?” Second, most research, even research that is explicitly cross-cultural, is rooted in one dominant psychological perspective (e.g., WASP, Western academic scientific psychology) that smacks of ethnocentrism, even if that is inadvertent.

A later approach was made to a broader range of questions that went beyond health to other domains of human activity. Once again we concluded that there was great potential but little actual achievement: a substantial knowledge base had not yet been applied. More recently, a number of authors addressed this question for a limited number of behavioral domains, and found that while progress had again been made, the gap between knowledge and practice remained wide. This state of affairs is not unique to cross-cultural psychology, but it is made more salient by the urgent need to use psychological knowledge in many parts of the world. Critical projects are needed in societies other than those in which the knowledge was generated.

One way to assist in the process of application is to present the knowledge base for findings that may qualify as “universals,” and hence that may be applicable across cultures. In such cases, the problem of cultural irrelevance or incompatibility may be reduced, even though the science–application gap may remain. Following the sequence in this article, we can first make use of the knowledge provided by our cognate disciplines of cultural anthropology, linguistics, and sociology. These disciplines have established universals in sociocultural systems and institutions, ones that can be found in all societies. Some of these have been portrayed in Section 5. What we take from these is not that the variation hinders application, but that the common dimensions provide a basis for the cross-cultural use of various research findings.

From these other disciplines, we have also adopted theoretical perspectives (especially relativism and universalism) and concepts that provide foundations for our methodological tools (emics, etics, equivalence); other applications include cultural similarities and differences in organizational behavior. Finally, an important application is training people to learn how to interact successfully with those from other cultures.

## See Also the Following Articles

Cognition and Culture ■ Cultural Psychology ■ Cultural Syndromes ■ Gender and Culture ■ Indigenous Psychologies ■ Intelligence and Culture ■ Intergroup Relations and Culture ■ Leadership and Culture ■ Perception and Culture ■ Power, Authority, and Leadership

## Further Reading

- Aberle, D. F., Cohen, A. K., Davis, A., Levy, M., & Sutton, F. X. (1950). Functional prerequisites of society. *Ethics*, 60, 100–111.
- Barkow, G., Cosmides, L., & Tooby, J. (Eds.) (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
- Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology*, 4, 119–128.
- Berry, J. W. (1975). *Applied cross-cultural psychology*. Lise: Sets & Zeitlinger.
- Berry, J. W. (1976). *Human ecology and cognitive style: Comparative studies in cultural and psychological adaptation*. Beverly Hills, CA: Sage.
- Berry, J. W. (2000). Cross-cultural psychology: A symbiosis of cultural and comparative approaches. *Asian Journal of Social Psychology*, 3, 197–205.
- Berry, J. W., Poortinga, Y. H., & Pandey, J. (1997). *Handbook of cross-cultural psychology* (Vol. 1, 3rd ed.) Boston, MA: Allyn & Bacon.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications*. New York: Cambridge University.
- Brislin, R. (Ed.) (1990). *Applied cross-cultural psychology*. Newbury Park: Sage.
- Dasen, P. R., Berry, J. W., & Sartorius, N. (Eds.) (1988). *Health and cross-culture psychology: Towards applications*. Newbury Park: Sage.
- Diener, E., & Suh, E. M. (Eds.) (2000). *Subjective well-being across cultures*. Cambridge, MA: MIT Press.
- Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology*, 69, 851–864.
- Frijda, N., & Jahoda, G. (1966). On the score and methods of cross-cultural research. *International Journal of Psychology*, 1, 110–127.
- Greenfield, P. M. (1997). You can't take it with you. *American Psychologist*, 52, 1115–1124.
- Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Thousand Oaks, CA: Sage.
- Inglehart, R., & Baker, W. (2000). Modernization cultural change and the persistence of traditional values. *American Sociological Review*, 65, 19–51.
- Jahoda, G. (1992). *Crossroads between culture and mind*. Cambridge, MA: Harvard University Press.

- Jahoda, G., & Krewer, B. (1997). History of cross cultural and cultural psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross cultural psychology* (2nd ed., pp. 1–42). Boston, MA: Allyn & Bacon.
- Kagitcibasi, C. (1997). Individualism and collectivism. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology* (2nd ed., pp. 1–50). Boston, MA: Allyn & Bacon.
- Klineberg, O. (1980). Historical perspectives: Cross-cultural psychology before 1960. In H. C. Triandis, & W. W. Lambert (Eds.), *Handbook of cross-cultural psychology* (Vol. 1, pp. 1–14). Boston, MA: Allyn & Bacon.
- Kluckhohn, C. (1954). Culture and behavior. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 921–976). Cambridge, MA: Addison-Wesley.
- Lenski, G. E. (1966). *Power and privilege: A theory of social stratification*. New York: McGraw-Hill.
- Levine, R. (1997). *A geography of time*. New York: Harper Collins.
- Lincoln, J. R., & Zeitz, G. (1980). Organizational properties from aggregate data: Separating individual and structural effects. *American Sociological Review*, 45, 391–408.
- Markus, H., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253.
- Naroll, R. (1983). *The moral order*. Beverly Hills, CA: Sage.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108, 291–310.
- Murdock, G. (1967). *Ethnographic atlas*. Pittsburgh, PA: University of Pittsburgh Press.
- Shweder, R. A. (1990). Cultural psychology: What is it? In J. W. Stigler, R. A. Shweder, & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human development* (pp. 1–43). Cambridge, MA: Cambridge University Press.
- Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Oxford, UK: Blackwell.
- Tapp, J. L., Kelman, H. C., Triandis, H. C., Wrightsman, L., & Coelho, G. (1974). Continuing concerns in cross-cultural ethics: A report. *International Journal of Psychology*, 9, 231–249.
- Triandis, H. C. (1964). Cultural influences upon cognitive processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1–48). New York: Academic Press.
- Triandis, H. C. (1989). Self and social behavior in differing cultural contexts. *Psychological Review*, 96, 506–520.
- Triandis, H. C. (1994). *Culture and social behavior*. New York: McGraw-Hill.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
- Triandis, H. C., & Lambert, W. W. (Eds.) (1980). *Handbook of cross-cultural psychology: Perspectives* (Vol. 1). Rockleigh, NJ: Allyn and Bacon.
- Witkin, H. A., & Berry, J. W. (1975). Psychological differentiation in cross-cultural perspective. *Journal of Cross-Cultural Psychology*, 6, 4–87.
- van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.



# Cultural Complexity

Mark W. Allen

California State Polytechnic University, Pomona,  
California, USA

James H. Liu

Victoria University of Wellington, Wellington,  
New Zealand

1. Cultural Complexity: A Recent Phenomenon
  2. Cultural Complexity Defined
  3. Theories on the Origins of Cultural Complexity
  4. Measuring Cultural Complexity
  5. Cultural Complexity and Dimensions of Cultural Variation
  6. Globalization and Cultural Complexity
- Further Reading

## GLOSSARY

**individualism–collectivism** Continuous dimension of cultural variation emphasizing an orientation toward personal goals, independent self-construal, and individual autonomy, at the one end, and group goals, interdependent self-construal, and emphasis on long-standing close relationships within in-groups, at the other.

**looseness–tightness** Continuous dimension of cultural variation describing the degree to which a culture endorses multiple and pluralistic standards of behavior, at the one end, and uniform, socially prescribed behaviors, at the other; loose cultures tolerate deviance from norms, whereas tight cultures emphasize sanctions for violating norms.

**multilineal cultural evolution** Dominant perspective on the evolution of cultural complexity; although it emphasizes some universal patterns in culture change from family level to state societies, it acknowledges multiple causal variables and trajectories in cultural evolution.

**power distance** Continuous dimension of cultural variation describing how relationships between high-ranking and low-ranking persons are conducted, with a high degree

of deference to superiors and social distance between superiors and subordinates, at the one end, and egalitarianism, at the other.

**social power** The means by which a social actor can influence, motivate, or control other individuals; it can be developed through interwoven networks of economic, military, ideological, and political sources of power and control.

**standard cross-cultural sample (SCCS)** A coded sample of worldwide cultures compiled by anthropologists for the purposes of cross-cultural studies.

**unilineal cultural evolution** Popular theoretical perspective during the late 19th century that cultures evolve along a single lineal dimension through set stages, usually based on technological or economic development; it has been long discredited by cultural anthropology and archaeology as simplistic and ethnocentric.

Cultural complexity entails the presence of a multitude of social networks, social groups, hierarchies, power structures, roles, positions, and/or divisions. Whereas kinship is the primary organizational concept in simpler societies, complex societies have multiple levels of crosscutting ties, affiliations, responsibilities, membership, values, norms, and rights. Individuals pursue their own particular goals through social networks, and no two people have the exact same network or objectives. Perhaps the key difference between simple societies and complex societies is the presence of social power, that is, the means by which human actors can manipulate others in the pursuit of social goals.

## 1. CULTURAL COMPLEXITY: A RECENT PHENOMENON

Cultural complexity is a recent phenomenon of human society. Paleoanthropology and the fossil record reveal a record of at least 5 million to 6 million years of hominid evolution. Hominids with large brain cases, the genus *Homo*, date back to circa 2 million years ago. Archaic *Homo sapiens*, including the Neanderthals, stretch back to perhaps 400,000 years ago. Our own species, *Homo sapiens*, can trace its origin to Africa some 150,000 to 200,000 years ago.

Yet even after the appearance of physically modern humans (with brains apparently as large and complex as our own), human culture and society remained relatively simple for tens of thousands of years. Despite having the same mental “hardware” as we do today, humans showed no flashes of art, complex organization, ethnicity, or other indications of significant change until the Upper Paleolithic period in the Old World circa 40,000 years ago. Even so, this early florescence of art, religion, technology, ethnicity, identity, and other human cultural hallmarks was still characterized by small and mobile populations, the lack of institutionalized leadership, extremely limited social power, and fairly simple social networks.

It is not until the first appearance of dense population, food surpluses, and settled villages roughly 10,000 years ago (the Neolithic period in the Old World and somewhat later during the Formative period of the New World) that cultural complexity can first be discerned in the human record. Since that point, there has been a drastic trend toward fewer and more complex societies. Eventually, chiefdoms, archaic states, and early empires in turn gave rise to modern nation-states during the 19th century. The 20th century witnessed the rise of economic, political, and cultural globalization. Clearly, there has been a breakneck rush toward ever higher levels of cultural complexity over the past few centuries. Perhaps as a consequence of this pattern of historical evolution, cultural complexity has been used primarily to characterize premodern societies rather than to differentiate between modern nation-states.

Social scientists interested in the relationship between human culture and biology should keep this point clearly in focus: The modern complex human brain was in place far earlier than were the complex cultural and social manifestations that it permits. Evolutionary psychologists should be cognizant of this when inferring connections between current human social behavior and the biological apparatus of the mind.

## 2. CULTURAL COMPLEXITY DEFINED

Given that it is a central theoretical concept in the social sciences, especially anthropology, an explicit definition of cultural complexity is surprisingly hard to pin down. Carneiro provided one influential perspective, arguing that complexity focuses on society as a structural end product, whereas evolution focuses on the processes that gave rise to this. Following Spencer, Carneiro defined evolution as “a change from a state of relatively indefinite, incoherent homogeneity to a state of relatively definite, coherent heterogeneity, through continuous differentiations and integrations.” Hence, cultural complexity would be characterized by coherent heterogeneity—a state of organized differentiation. Such a definition is necessarily vague in that cultural complexity is used primarily as a relative term to rank order various cultures and societies (and this can be done reliably) rather than standing alone in terms of essential meaning according to necessary traits or attributes.

Cultural complexity entails the presence of a multitude of social networks, social groups, hierarchies, power structures, roles, positions, and/or divisions. Whereas kinship is the primary organizational concept in simpler societies, complex societies have multiple levels of cross-cutting ties, affiliations, responsibilities, membership, values, norms, and rights. Individuals pursue their own particular goals through social networks, and no two people have the exact same network or objectives. Perhaps the key difference between simple societies and complex societies is the presence of social power by which human actors manipulate others in pursuit of social goals. With social power comes the potential for social inequality, ranking, hierarchy, and social stratification. It is also possible, however, to have multiple divisions of horizontal rather than vertical power, that is, heterarchy rather than hierarchy. In such cases, there is a different but equal distribution of power.

## 3. THEORIES ON THE ORIGINS OF CULTURAL COMPLEXITY

A wide range of theoretical perspectives has been employed to understand how and why human cultures become complex. Progressivists such as Locke first opened the door to explaining differences in human societies without recourse to religious or mythological causes. By the mid-19th century, naturalists and social

scientists alike were profoundly influenced by Darwin's theory of natural selection. Social Darwinists such as Spencer and early anthropologists such as Tylor applied evolutionary concepts to the development of human society. Morgan and other unilineal cultural evolutionists believed simplistically that human societies naturally progressed through a series of stages usually tied to economic or technological patterns. Such thinkers, unfortunately, tended to invoke racial arguments to explain differences in cultures and their level of attainment. Nevertheless, such thinking had a strong influence on Marx in his theories of social evolution and economic materialism. Similarly, Freud was heavily influenced by unilineal evolution as he outlined the development of human societies from simple to complex through universal psychological patterns.

The early and mid-20th century saw two major perspectives on cultural complexity struggle for supremacy in both American and European social sciences. "Multilinear evolutionists" favored materialist explanations focused on economics, technology, energy, ecology, and the like. Unlike their 19th-century predecessors, however, these theories recognized numerous causal factors and possible cultural trajectories. Evolutionists rarely gave credence to ideology, symbolism, or the role of the individual. Cultural relativists, on the other hand, avoided generalizations and focused instead on the diffusion of cultural traits or functional descriptions of culture and social structure. The most prominent example in American cultural relativism is Boas. In Europe, social theory was led by schools of sociology (e.g., Durkheim, Mauss, Weber), structuralism (e.g., Levi-Strauss), and functionalism (e.g., Malinowski, Radcliffe-Brown, Evans-Pritchard). These perspectives often downplayed materialism and instead employed ideology, symbolism, and deep cognitive structures as causal factors. Some of these perspectives also attached great importance to the role of the individual as a cultural actor with agency.

The social sciences, however, did develop a synthesis view of cultural evolution and cultural complexity by the end of the 20th century. It recognizes that extreme views are unlikely to derive a parsimonious explanation for all cases. Mann has been a leading figure of the synthesis through his multidimensional social power theory, which stresses that individuals are agents that pursue their goals through overlapping networks of economic, military, political, and economic power. Social power is the means by which individuals acquire and manipulate the flow of information, energy, social status, wealth, and other resources.

Social scientists during the past few decades have had a more sophisticated view of cultural complexity, one that takes account of a variety of overlapping factors.

#### 4. MEASURING CULTURAL COMPLEXITY

A recent review by Chick outlined three major approaches to the measurement of cultural complexity (a broader review of seven specific measures was developed earlier by Levinson and Malone). The three types of measures are highly correlated, and each has specific strengths and weaknesses. The approaches are (a) attribute counting procedures, (b) development scores based on the Standard Cross-Cultural Sample (SCCS), and (c) maximum settlement size.

The first measure is termed "culture by accretion." In this method, cultural complexity is assessed by an attribute checklist; the greater the number of attributes possessed, the greater the complexity score. As few as 50 and as many as 618 attributes have been used. The greater the number used, the more accurate the score and the greater the ability to differentiate between cultures at similar levels of complexity. If only a limited number is chosen (e.g., 50), items conforming optimally to Guttman scaling procedures are required.

Domains for the attributes (or traits) used by Carneiro in 1970 were economics, social organizations and stratification, political organization, law and judicial process, warfare, religion, art, tools–utensils–textiles, metalworking, watercraft and navigation, and special knowledge and practices. Sample items included the presence or absence of markets, military conscription, special religious practitioners, and temples. Most of the items were related to technological and organizational elements of culture rather than expressive ones, but the breadth of the attributes nonetheless allows researchers to empirically assess the level of cultural complexity.

Although comprehensive, reliable, and accurate, such measures have fallen out of fashion due to the difficulty in obtaining scores for such a large number of attributes and because the attributes chosen seem to be specific to cultures at certain stages of development. They measure increases in structure rather than growth in the size of a structure (e.g., a society with 7 judges is of equal complexity to a society with 7000 judges in this system). Because the most complex society studied by Carneiro was the Roman Empire, such domains as information technology or energy sources were not assessed in his scoring system.

The second approach, perhaps the best known among the approaches, is based on 10 indicators of development from the SCSS, with each being rated from 0 to 4 and largely focused on technology: writing and records, fixity of residence, agriculture, urbanization, technological specialization, land transport, money, population density, political integration, and social stratification. In 1973, Murdock and Provost reported a correlation of .95 between their scale and Carneiro's scale using a sample of 45 societies common to both studies.

Because this measure depends on rating the degree of sophistication of a few indicators rather than counting the presence or absence of many attributes, Chick called it "complexity by elaboration." The measure factors into two dimensions: social and technological complexity and the complexity of human ecology. Like Carneiro's measure, the sample of the SCSS has few complex modern societies, and so the top end of the scale might not differentiate well. Chick argued that the scale does not adequately assess expressive elements of culture such as arts and dance. But in 1972, Lomax and Berkowitz provided evidence that cultures synchronize activity in productive efforts and expressive performance.

Finally, and most succinctly, there is a measure of complexity by magnitude. Drawing from previous work showing a .97 correlation between Carneiro's scale and the logarithm of maximum settlement population, Naroll and Divale argued in 1976 that maximum settlement population is the single best measure of cultural complexity. The reason for this can be found in Carneiro's 1987 study of single-community societies (i.e., autonomous bands and villages). He reported that the organizational complexity of a village increases as the two-thirds power of its population. Villages that do not increase in organizational complexity as their population increases tend to fall apart. Extrapolating from this, it follows that in larger societies there should also be a mathematical relationship between maximum settlement size and the amount of organizational and technological complexity required to provide for and hold together these populations.

It would be difficult to apply this measure to non-sedentary populations, and it does not capture the types of cultural complexity afforded by globe-spanning mass communication and information technologies. It would also imply that Japan, with more than 30 million people living in agglomeration around metropolitan Tokyo, is the most complex society in the world today.

To summarize, despite difficulty in providing a precise conceptual definition of cultural complexity, and despite limitations in its operationalization, there is consistent evidence of the convergent validity for this concept. However cultural complexity is operationalized, its disparate measures are highly intercorrelated. This suggests that there is a common theoretical construct underlying these measures, however difficult it may be to articulate succinctly.

## 5. CULTURAL COMPLEXITY AND DIMENSIONS OF CULTURAL VARIATION

In relating cultural complexity to the dimensions of cultural variation typically used by psychologists, it is important to remember that the concept originates in anthropology. Most of the societies and cultures used by Carneiro or Murdock and Provost to establish their measures of cultural complexity either cannot or are unlikely to be represented in the corpus of data associated with psychology. This is not to say that the concept of cultural complexity is not useful to psychologists, but it does sound a warning that there is a severe restriction in range among the peoples studied by psychology, including cross-cultural psychology. Established measures of complexity are not especially sensitive to differences at the high end of complexity, which is the range drawn on by most psychologists.

On the other end, it is unclear how less complex societies would score on the dimensions of variation in subjective culture cherished by psychologists. Would hunters and gatherers tend toward individualism or collectivism? Hunters and gatherers certainly relied on kin groups for survival, but these tended to be loose coalitions that could change with the availability of resources or with social conditions. Although interdependence was a basic condition of survival, there were fewer social norms governing behavior and less enforcement of these norms than in more complex societies. In 1994, Triandis argued that complexity is related to affluence and to the existence of multiple subgroups with different norms (hence promoting individualism), but another view is that it is difficult to establish a clear linear relationship.

Lomax and Berkowitz, using both the Ethnographic Atlas and cantometric measures of song, provided evidence that the relationship is curvilinear, with group cohesiveness, group organization, and division

of labor being highest among cultures at middle levels of complexity. A similar argument could be applied to looseness–tightness. Data from an ecocultural approach found that only in the range from moderate (i.e., irrigated agriculture) to highly complex (i.e., industrial) societies is increased complexity associated with more looseness and less social conformity. In the range from hunter-gathers to the onset of agriculture, Berry found in 2001 that increased complexity, as operationalized by societal size, is actually associated with more social conformity (less looseness). At the high end, both the United States and Japan are incredibly complex but are currently used as instances of individualism, on the one hand, and collectivism, on the other.

There appears to be a more clear-cut relationship between cultural complexity and the second most frequently used dimension of cultural variation in psychology, that is, power distance. Put simply, cultural complexity affords power distance. By definition, the amount of differentiation of statuses, roles, affluence, and power with a society increases with its complexity. However, there is considerable latitude in how this differentiation is managed and perceived. For example, Americans aspire toward informality based on egalitarian ideals, whereas Japanese adopt more role-based formality in dealing with different statuses and roles. Although the real difference in incomes between people in upper and lower classes is greater in the United States, the psychological distance between the upper and lower parts of society is perceived to be lower among Americans than among Japanese. But over the broad range of human societies, given the tendency of individuals to exploit structural differences in power to their advantage, the perception of power distance should increase with increases in cultural complexity.

The measurement of cultural complexity is a reminder to psychologists that subjective culture is not the only viable definition of culture. The technological, material, organizational, and political–economic features of society taken by anthropologists to be indicators of cultural complexity may provide preconditions for the emergence of dimensions of variation in subjective culture found in cross-cultural psychology. From the other end, it may be extremely dangerous to assume that dimensions of cultural variation emerging from the analysis of samples from modern nation-states provide an adequate description of the full range of variability in subjective culture possible in human society.

## 6. GLOBALIZATION AND CULTURAL COMPLEXITY

The historical and archaeological record over the past 10,000 years can be read as a steady movement toward increasing levels of cultural complexity, with a massive acceleration during the past 200 years. Concurrent with increases in complexity has been a decrease in the number of independent polities in the world. There were more than 100,000 world polities circa 500 AD compared with the approximately 200 sovereign states represented in the United Nations today. With the onset of globalization, where information technology, high-speed travel, trade, and mass media have reduced geographic boundaries between societies, representing dimensions of cultural variation as dichotomies (e.g., individualism vs collectivism) is becoming increasingly problematic. It is now possible to speak of a global system and the cultural complexity of this system, with hybridity and contact zones between cultures being just as important as core values and practices within cultures.

In such a global system, it may be interesting to speculate about the ramifications of increased complexity. In the developing world, traditional cultures will continue to struggle with the onset of modernity. Collectivism and tightness in these cultures will give way before or hybridize with individualism and looseness. In the system as a whole, power distance will continue to increase, with immense concentrations of power and wealth being in the hands of global elites. In such an emerging system, words such as “individualism” and “looseness” might be inadequate to describe the values and roles of people at the top end of human society. The influence of economic elites—whatever their nationality—on the evolution of global cultural complexity will be immense and largely unprecedented in human history.

Of course, social scientists are aware that cultural evolution is not a one-way street. Cultural devolution, or the reduction of complexity, has also been documented and is usually caused by political or ecological catastrophes. This is not an unforeseen possibility for observers of the global system. Whereas the conceptualization and measurement of cultural complexity has been focused on simpler societies, future research might undertake to describe cultural evolution at the high end of complexity as social scientists seek to describe the interaction between a global supersystem and its constituent parts.

### See Also the Following Articles

Acculturation ■ Conformity across Cultures

### Further Reading

- Berry, J. W. (2001). Contextual studies of cognitive adaptation. In J. M. Collis, & S. Messick (Eds.), *Intelligence and personality: Bridging the gap in theory and measurement* (pp. 319–333). Mahwah, NJ: Lawrence Erlbaum.
- Carneiro, R. L. (1970). Scale analysis, evolutionary sequences, and the rating of cultures. In R. Carroll, & R. Cohen (Eds.), *A handbook of method in cultural anthropology* (pp. 834–871). Garden City, NY: Natural History Press.
- Carneiro, R. L. (1987). The evolution of complexity in human societies and its mathematical expression. *International Journal of Comparative Sociology*, 28, 111–128.
- Chick, G. (1997). Cultural complexity: The concept and its measurement. *Cross-Cultural Research*, 31, 275–307.
- Harris, M. (1968). *The rise of anthropological theory: A history of theories of culture*. New York: Thomas Y. Crowell.
- Hermans, H. J. M., & Kempen, H. J. G. (1998). Moving cultures. *American Psychologist*, 53, 1111–1120.
- Levinson, D., & Malone, M. J. (1980). *Toward explaining human culture: A critical review of the findings of worldwide cross-cultural research*. New Haven, CT: Human Relations Area Files Press.
- Liu, J. H., & Allen, M. W. (1999). Evolution of political complexity in Maori Hawke's Bay: Archaeological history and its challenge to intergroup theory in psychology. *Group Dynamics*, 3, 64–80.
- Lomax, A., & Berkowitz, N. (1972). The evolutionary taxonomy of culture. *Science*, 177, 228–239.
- Mann, M. (1986). *The sources of social power*. Cambridge, UK: Cambridge University Press.
- Murdock, G. P., & Provost, C. (1973). Measurement of cultural complexity. *Ethnology*, 12, 379–392.
- Murdock, G. P., & White, D. R. (1969). Standard cross-cultural sample. *Ethnology*, 8, 329–369.
- Naroll, R., & Divale, W. T. (1976). Natural selection in cultural evolution: Warfare versus peaceful diffusion. *American Ethnologist*, 4, 97–128.
- Triandis, H. (1994). *Culture and social behavior*. New York: McGraw-Hill.



# Cultural Psychology

Patricia M. Greenfield

University of California, Los Angeles, California, USA

Heidi Keller

University of Osnabrueck, Osnabrueck, Germany

1. Introduction
2. The Conception of Culture in Cultural Psychology: Conceptual and Methodological Issues
3. Sociopolitical Roots of Cultural Psychology
4. Exemplars of Research Programs
5. Application of Cultural Psychology
6. Conclusions: An Example of Cultural Process  
Further Reading

## GLOSSARY

**apprenticeship** The interactive process by which cultural experts help novices acquire knowledge in a particular domain.

**coconstruction** Creation of shared meanings or shared activities through an interactive communication process.

**community of practice** A group that regularly comes together to participate in shared activities.

**cross-cultural psychology** The study of similarities and differences across cultures in human behavior, cognition, and emotion, with culture usually being an independent variable. It differentiates etic and emic approaches. Etic approaches imply the transfer of concepts and methodologies from one culture to other cultures; emic approaches imply the development of culturally sensitive concepts and methodologies that are derived from within each culture to be studied. Cross-cultural psychology was first linked to cognitive psychology and cognitive development and later to social psychology and social development.

**ethnography** An anthropological method of field research through participant observation. The researcher utilizes a participatory role as a position from which he or she can

make observations in a particular cultural context. Ethnography is generally carried out by researchers who do not belong to the ethnic community under study but who spend a considerable amount of time within the ethnic community and usually speak the local language. Ethnography covers observational methods, interviews with cultural informants, and focus group discussions.

**ethnotheory** A folk or lay theory of some part of the everyday world.

**expert** An expert has acquired advanced cultural knowledge in a particular domain.

**independence** The self-construal prevailing in large-scale, urban industrialized cultural communities, such as the United States or Western Europe. It denotes a cultural deep structure of a self-contained and bounded self that is separate from and competes with social others. The deep structure becomes instantiated in locally appropriate phenotypical expressions.

**indigenous psychology** The study of local conceptions of humans, their development, and their psychology. Indigenous approaches express dissatisfaction with the ethnocentric Euro-American view of psychology, with its claims of universality. There are two conceptual approaches, one aimed at describing variability within a universal framework (the unity of mankind) and the other aimed at establishing diverse psychologies.

**interdependence** The self-construal prevailing in small-scale, rural, subsistence cultural communities in Asia, Africa, and Latin America. It denotes a cultural deep structure of an interrelated self that maintains harmonious and hierarchically structured relationships with social others. The deep structure becomes instantiated in locally appropriate phenotypical expressions. Historical transitions and multicultural societies may support expressions of both

independence and interdependence in the definition of the self.

**macrodevelopment** Directed change over a relatively long timescale, such as years.

**microdevelopment** Directed change over a relatively short timescale, such as hours or days.

**novice** A novice is at the beginning stages of acquiring cultural knowledge in a particular domain.

**parental ethnotheories** The social cognitions that humans hold with respect to the concept of a child, the concept of a parent, and the developmental process. They are organized as belief systems that pertain to different domains. Parental ethnotheories are shared among members of cultural communities. They have intuitive, nonconscious components and serve as frameworks for parental behaviors and actions.

Social interaction among beings biologically primed for culture creates culture for the group and for the individuals in it. Culture is viewed as a socially interactive process with two main component processes: the creation of shared activity (cultural practices) and the creation of shared meaning (cultural interpretation). Shared cultural knowledge and modes of cognition are the result of processes of coconstruction. The creation, acquisition, transmission, and use of culture are psychological and interactional processes. It is the study of these processes that is the paradigmatic subject of cultural psychology.

## 1. INTRODUCTION

Shared activities are adapted to survival in a particular environment; they are the material side of culture. A key aspect of shared activity is that it involves goal-directed action; this point is particularly emphasized in the German tradition of cultural psychology. Shared meaning is the symbolic side of culture. A key aspect of shared meaning is interpretation, the notion that culture provides a lens through which we see and interpret (evaluate) the world. Empirical study in cultural psychology focuses on these two components of cultural process—shared meaning and shared activity.

Both components are cumulative in nature because culture is created by processes that occur between, as well as within, generations. Meanings and activities not only accumulate but also transform over time. This cumulative and temporal characteristic of culture creates the necessity for historical methods of studying culture as psychological process. It also creates the

necessity for developmental methods for studying the transmission and transformation of culture from the older to the younger generation. Equally important (albeit understudied) are developmental transitions from being a novice socialized by experts to being an expert socializing novices. Cultural psychology does not conceptualize the adult as a static entity with set cultural characteristics but rather attempts to study the diachronic processes of socialization and development that constitute cultural learning and apprenticeship.

As a species, humans are biologically primed to create, acquire, and transmit culture; culture is the primary mode of human adaptation. Because of the centrality of adaptation as a concept, cultural psychology aims to study meanings and activities in context rather than in the more decontextualized setting of the laboratory experiment. Cultural psychology is an interdisciplinary specialty that integrates psychology and anthropology.

## 2. THE CONCEPTION OF CULTURE IN CULTURAL PSYCHOLOGY: CONCEPTUAL AND METHODOLOGICAL ISSUES

Cultural psychology can be understood by contrast with two other approaches to the relationship between psychology and culture—cross-cultural psychology and indigenous psychology. Although all three are fuzzy concepts with partially overlapping exemplars, the paradigmatic instances of the three approaches are quite distinctive; these paradigms are central to the descriptions that follow.

### 2.1. Cultural Psychology Compared with Cross-Cultural Psychology

The goal of cross-cultural psychology is to separate the universal from the culture specific in human behavior. Cross-cultural psychology has a universalistic bias: Culture can be decomposed, its different layers can be peeled off, until the existential and universal human nature remains. In cross-cultural psychology, culture is generally operationalized as an antecedent or independent variable to behavior. Often, the independent variable is the label of a particular national group (e.g., a comparison of Chinese and U.S. participants). Occasionally, as in Harry Triandis's 1980 definition, culture is also conceived as a dependent variable, with

behavior or experience as the independent variables. Whether an independent or a dependent variable, culture is implicitly viewed as being outside of and apart from the individual.

In cultural psychology (original terminology from Price-Williams), culture is, in contrast, seen as inside the individual. Culture is “a way of knowing, of constructing the world and others” (Bruner, 1993, p. 516). Through processes of interaction and communication, these construals acquire a certain degree of intersubjectivity or shared meaning. Shared knowledge and shared meanings generate a set of everyday practices that also define culture in terms of shared activities. Thus, as Jahoda points out, culture and behavior, culture and mind are viewed as indistinguishable.

Accordingly, cultural psychology derives its problems and procedures from an analysis of the nature of culture, with an important focus on everyday life. Whereas the methodological ideal of the cross-cultural psychologist is to carry a procedure established in one culture, with known psychometric properties, to one or more other cultures, to make a cross-cultural comparison, the methodological ideal of the cultural psychologist is for problems and procedures to flow from the nature of culture, both in general and in specific. For example, in their 1981 book, Scribner and Cole focused on literacy as an important cultural element, a way of interpreting and representing the world (the nature of culture in general). They proceeded to find out what each of three different literacies meant to the Vai people of Liberia (the nature of culture in specific) and to study the cognitive consequences of these literacy practices.

The cultural psychology of Scribner and Cole traces its intellectual roots to the work of Russian psychologists Vygotsky and Luria. In the early 1930s, soon after the Russian revolution, Vygotsky and Luria studied the logical processes of collectivized and noncollectivized central Asian peasants. Their stimuli were two kinds of syllogism: “The contents of some were taken from the concrete, practical experience of the villagers; the contents of others bore no relation to familiar, practical life” (Cole & Scribner, 1974, p. 161). The point of their study was the cognitive difference in the way the two groups of peasants handled the two kinds of syllogism. Again, syllogisms are an aspect of human logic (culture in general), whereas the everyday life of the peasants (culture in specific) was a key source of problems to solve. These examples from Scribner and Cole and Vygotsky and Luria not only illustrate the methods of cultural

psychology but also illustrate the importance of cognition in the origins of cultural psychology.

Thus, cultural psychologists study cultural processes directly—for example, literacy or syllogisms as cultural forms and products. In comparison with cross-cultural psychologists, their research designs rely much less on what the anthropologist Beatrice Whiting called “packaged” variables. A packaged variable is an index of culture as an antecedent or independent variable. An example of a packaged variable is the labels “Japanese” or “German.” In principle, each of these labels “packages” a whole suite of shared cultural characteristics and processes. When cultural psychologists make comparisons, they do not use packaged variables. Instead, they use much narrower groupings, such as collectivized vs uncollectivized peasants in Luria and Vygotsky’s work.

Some research falls on the border between cultural and cross-cultural psychology. In the 1960s, for example, Greenfield brought Piagetian cognitive tasks to the Wolof of Senegal, an approach belonging to cross-cultural psychology. However, she did not compare Senegalese to North Americans, which would have been a cross-cultural approach. Instead, she looked at developmental trajectories within each culture, as well as compared children from different ecologies within the Wolof world: urban vs rural and schooled vs unschooled. She also examined the Whorfian hypothesis by doing a semantic analysis of various domains in Wolof and French. This was another way in which research procedures flowed from the nature of the culture—in this case, the linguistic aspects of culture.

The cultural psychological perspective implies that research is conceptualized as a communication process. Communication with the people of the study community in their own language is a prerequisite for cultural analyses of shared activities and shared meanings. The anthropological notion of ethnography as a methodological concept is a necessary first step in order to bring the importance of lived experience in a cultural place from ground to figure. In order to study cultural processes from the participants’ own points of view, cultural psychology has, as Greenfield pointed out in 1997, developed its own tool kit of methods. Some, like ethnography and discourse analysis, have been adapted from anthropology. With other methods, such as the use of video, cultural psychology has followed the leadership of developmental psychology as a discipline. (Later, exemplars of different research programs are introduced, noting different cultural methodologies for different research problems.)

## 2.2. Cultural Psychology Compared with Indigenous Psychology

Although indigenous psychology and cultural psychology clearly have independent origins, they share the notion that the prime subject of study is the subject's creation of meaning systems, particularly systems that are shared or normative within a defined cultural group. In different ways, both traditions have recognized that psychological theories are important aspects of shared cultural meaning. Millers, writing from the vantage point of a cultural psychologist, has, on a theoretical level, asserted the cultural grounding of all psychological theory. This cultural grounding of theory has been a strong motive, if not an explicit metatheoretical statement, for indigenous psychology.

The unique contribution of indigenous psychology is the notion that psychological concepts and psychological theory, not just data-collection techniques, should be developed within each culture. Unlike indigenous psychology, the empirical research tradition of cultural psychology has not been based on formal psychological theories with culture-specific origins. Nevertheless, cultural psychology has increasingly in recent years made ethnotheories (i.e., folk theories) of psychological functioning and development a subject for empirical investigation—for example, in the collaborations of Shweder and Bourne and Harkness and Super.

Indigenous psychology, however, aims to go one step further: The goal of indigenous psychology is to take informal folk theories of psychological functioning and formalize them into psychological theories. Cultural psychology arrived at the empirical study of folk theories, including folk theories of psychological development. Indigenous psychology has, in turn, moved ethnopsychology from an object of empirical study to a source of formal psychological models. In other words, indigenous psychologists have taken steps to translate ethnotheories of psychology into formal theories of psychology and, from these theories, to conduct empirical psychological studies. This goal is incredibly important for psychology as a whole, on both a metatheoretical and a metamethodological level.

However, cultural and indigenous psychology also differ in some respects. Unlike indigenous psychology, which, as Sinha points out, was born as an attempt to decolonize the mind—in fact, many of its pioneers and proponents inhabit former British colonies—cultural psychology still constitutes a “crossing over” into someone else's culture by the investigator. Also, unlike indigenous psychology, its connection to

cross-cultural psychology has tended to be erratic and ambivalent.

On the other hand, because of its own origins with an emerging scientific elite in developing nations, indigenous psychology, especially in East Asia, tends, on the whole, to privilege elite populations (university students) as subjects of study and culture change as a research topic. Cultural psychology, in contrast, tends to give a great deal of attention to relatively stable subsistence village cultures; for example, well-known studies have been done on subsistence groups in Liberia, Morocco, Guatemala, India, and Mexico.

## 2.3. Comparing Indigenous, Cross-Cultural Psychology, and Cultural Psychology

Although indigenous psychology shares the spirit of cultural psychology, its methods (in practice, although not in principle) tend to resemble those of cross-cultural psychology. That is, indigenous psychology most often utilizes standard psychological methodology such as questionnaire formats and tends to study variables rather than processes. For example, Yang and Bond used rating scales and adjective lists in their 1990 study of indigenous Chinese personality constructs.

There are some notable exceptions. For example, Enriquez, the founder of indigenous psychology in the Philippines, has utilized an indigenous (and group-oriented) method of social interaction as a means for data collection. However, by and large, indigenous psychology, unlike cultural psychology, has not experienced the influence of method and theory from anthropology and developmental psychology.

Why has cultural psychology developed in the way that it has? To answer this question, we turn to a study of its sociopolitical roots.

## 3. SOCIOPOLITICAL ROOTS OF CULTURAL PSYCHOLOGY

Cultural psychology has a long tradition and a rather young history in psychology. In the context of defining the place of psychology in the systematics of sciences, in 1920 Erich Stern presented a view that could not be more modern. For him, psychology is inevitably a cultural science since any attempt to understand any psychological phenomena needs to take into account the social cultural environment, ontogenetic history,

and ancestral heritage. Culture represents the legacy of preceding generations as expressed in the dispositions, the consciousness, and the psychology of each living individual whose plasticity allows change in order to adapt to changing surroundings. His thoughtful and comprehensive treatise also disqualified experiments as a tool to understand psychological processes.

The modern conception of cultural psychology developed without connection to such ancestors. It grew out of liberal policy in a world increasingly dominated by the Cold War after World War II. The notion was to help the Third World develop in order to make sure that, on the political level, Third World countries did not fall to Communism. Thus, John Gay and Michael Cole, funded by the Ford Foundation, began their cultural psychological research in Liberia as part of the African Education Program of Educational Services, Inc. of Watertown, Massachusetts. Daniel Wagner originally went to Morocco with the U.S. Peace Corps. Patricia Greenfield's dissertation research in Senegal in 1963 was funded by a grant (awarded to Jerome Bruner at Harvard University) from the U.S.-based Ford Foundation to develop the Institute of Pedagogical Studies at the University of Dakar in Senegal.

Initially independent from this U.S. branch of cultural psychology, a German school of cultural psychology emerged in Saarbruecken. It was founded by Ernest Boesch and further developed by Lutz Eckensberger. These scholars also went to so-called Third World cultures on the basis of politically motivated developmental aid programs. Ernest Boesch went to Thailand using grants from the UNESCO in order to set up an institute; Lutz Eckensberger went to Afghanistan using grants from the former German Ministry of Developmental Aid in order to train teachers in vocational schools. The initial contact with other and very different cultures thus became a major factor in conceptualizing cultural psychologies.

On the institutional level, the Department of Social Relations at Harvard, with its integration of psychological, sociological, and anthropological methods and theories, provided the educational foundation of modern cultural psychology and was particularly crucial in the merger of psychology and anthropology. The actual term cultural psychology seems to date from 1969, when two anthropologists, DeVos and Hippler, wrote an article titled "Cultural Psychology: Comparative Studies of Human Behavior" for the second edition of Lindzey and Aronson's *Handbook of Social Psychology*. A second use of the term cultural psychology emerged in 1980 when Douglass Price-Williams, a psychologist

serving as a member of an anthropology department, wrote an article titled "Toward the Idea of Cultural Psychology: A Superordinate Theme for Study." Hence, cultural psychology has from the beginning represented the meeting of psychology and anthropology.

On the intellectual level, cultural psychology has grown out of dissatisfaction with the universalism and decontextualized methodology of psychology in general and cross-cultural psychology in particular. On the other hand, it has also grown out of anthropology's wish to deal with the person, not merely the culture as a supraindividual envelope. These motivations, from Cole, Shweder, and Bourne in the United States, contrast with that of German cultural psychology. The latter has, in reaction to the Nazi Terror, been driven by a strong desire for a humanistic psychology or even a more human world. These qualities seemed lacking in behavioristic approaches that became mainstream during that time.

## 4. EXEMPLARS OF RESEARCH PROGRAMS

### 4.1. Parental Ethnotheories: An Example of Shared Meaning

The analysis of parental ethnotheories about development has become a vital area of research. Parental ethnotheories are shared beliefs about the goals of child development and the socialization practices that will achieve these goals. According to Harkness and Super, they specify how to become a competent adult in a particular environment. Parents throughout the world hold specific beliefs about proper care and handling of small children. These parental ethnotheories express conceptions of the nature of children, parenting, and development. In this way, ideas about child care practices are related to developmental goals, as defined by the cultural environment. Parental ethnotheories thus link socialization practices to cultural values. They also link the socialization agendas of one generation to the next. Parental ethnotheories generate child-rearing practices that, in turn, create the cultural adult. This cultural adult, in turn, participates in the socialization processes of the next generation.

Although interindividual differences are prevalent with respect to parenting goals and practices in every culture, members of cultural communities can be viewed as groups of individuals who coconstruct a shared reality in different domains of life. According

to Markus and Kitayama, two basic contrasting realities are the value systems that define independent and interdependent construals of the self. These, in turn, as Greenfield and Cocking have shown, generate corresponding developmental goals. These goals are actualized through different socialization practices and parenting styles during different stages of development. Parental ethnotheories can thus be conceived of as the mediating links between these cultural metamodels and behavioral contexts and practices.

Keller and collaborators analyzed parental ethnotheories in cultural communities found to differ in their orientations toward independent and interdependent construals of the selves: West African Nso farmers (interdependent orientation) and middle-class northern Germans (independent orientation). Parental ethnotheories were assessed as reactions to videotaped prototypical interactional situations from these different cultural communities. (This procedure has also proved revealing of cultural patterns in research on teacher and classroom evaluation.) Particularly because these reactions were assessed in a group setting, in which consensus was generally negotiated, it can be assumed that the respondents use their shared cultural framework with all the cultural standard operating procedures and unstated cultural assumptions. Because of the use of actual interactions as stimuli, this procedure further allows one to address the issue of the belief-behavior dissonance. Comments concerning deviations from normative practices reveal the valued practices and beliefs.

These analyses revealed results that are in line with the expectations of different cultural orientations with respect to independence and interdependence. The Nso women comment on parental behaviors that can be related to interdependent developmental goals such as primary care and body contact and subscribe to a physical stimulation model of infancy, in which their special technique of motor stimulation is supposed to foster growth and development. The German women, through their cultural lenses, value face-to-face communication, exclusive attention toward the baby, object play, and language. They subscribe to a communication model of infancy, in which communication occurs between two bodies that are physically separate and the flow of the interactional exchange is supposed to be shared by (quasi) equal partners, thus supporting independent socialization goals.

The evaluations of the German and the Nso women are independent of the origin of the videotapes. That is, the Nso cultural lens interprets physical stimulation practices positively when participants see them on

videos from either culture. In contrast, the Nso cultural lens provides a negative evaluation when these stimulation practices are missing in the videos of either culture. Similarly, German mothers positively evaluate instances of the communication model and negatively criticize their absence on both the Nso and German tapes. However, physical stimulation behaviors are more prevalent in the Nso videos, whereas communication behaviors are more prevalent in the German videos. Thus, the cultural values manifest in the ethnotheories are reflected in socialization practices favored by each cultural group.

Ethnotheories link to socialization practices, which then link to children's behavioral development (e.g., the timing of mastering developmental tasks). Having experienced distinct and frequent body stimulation, the Nso babies develop the gross motor milestones of sitting, standing, and walking significantly earlier than the German babies. On the other hand, the experience of a rich language environment, face-to-face interactions with nonverbal and verbal contingencies, and object play instigate an earlier onset of language in German babies compared to Nso babies. The fact that the Nso community does not value the early onset of language as much as the German cultural community is expressed in a statement of a Nso woman, commenting on the language use of a German mother interacting with her baby: "For them, every child needs to be intelligent." The cultural timing of developmental milestones organizes subsequent developmental pathways in coherent and systematic ways so as to achieve the developmental goals valued in each cultural environment. The mothers who engage in these child-rearing practices were themselves the object of the corresponding socialization practices when they were children. It is believed that childhood socialization is perhaps the most conservative or persistent part of any culture. On the other hand, with changing sociocultural and especially socioeconomic environments, cultural practices also change. However, research shows that practices change long before attitudes and beliefs.

#### **4.2. Apprenticeship: The Historical Transformation of Shared Practice**

As Barbara Rogoff and Jean Lave note, apprenticeship is the participatory structure by which more expert members of a culture induct less expert members into the normative activities and practices of a culture. Corresponding to the two basic types of ethnotheory,

one idealizing the independent individual and one idealizing the interdependent individual, are two basic styles of apprenticeship. The independent style of apprenticeship utilizes more trial-and-error strategies, learner initiative, and division of labor. The interdependent style utilizes more teacher guidance, observational learning, and interdependent problem solving. The former is well adapted to a traditional agrarian setting in which information and resources flow from the older to the younger generation with relatively little change. The latter is well adapted to a commercial economy in which innovation and novelty are valued and the younger generation is more independent of the older.

However, the styles of apprenticeship are not constant; they are socially constructed in response to environmental conditions. Therefore, changes in economic conditions should induce a shift in mode of cultural transmission. This notion was tested by Greenfield, Maynard, and Childs in a video study of weaving apprenticeship in a Maya community at two points in time; the two waves of data collection were separated by two decades. At the second point of data collection, the weaving learners from the first wave had grown up and become mothers whose daughters were now learning to weave.

In both cohorts, researchers microanalyzed weaving videos to understand the interaction between learner and teacher as apprenticeship progressed. The first generation of weaving teachers, mainly mothers, utilized the interdependent style of apprenticeship as their daughters learned to weave. Two decades later, after a period in which the economy moved away from subsistence and toward commerce, the mode of apprenticeship had indeed shifted. Microanalysis of the videos revealed the appearance of a new, more independent style of apprenticeship. Weaving learners were displaying more initiative in the apprenticeship process, peer teachers (compared with teachers from the older generation) had become more prevalent, and there was less cooperative weaving between learner and teacher. This shift was concentrated, moreover, in just those girls who, with their mothers, were most involved in textile commerce.

This historical transition in the apprenticeship process was reflected in changes in resulting weaving practices. What girls and women wove had also changed. The concept of community creativity, in which a few fairly uniform designs identified members of a cultural community, was joined by a concept of individual creativity. Design innovation and individuation were now apparent throughout the woven textiles. Shared practices had changed in the direction of a new cultural model.

The dimension of shift was that dimension most basic in the cultural differentiation of ethnotheories of development discussed previously. This parallel suggests a close conceptual relationship between the analysis of cultural meanings and the analysis of cultural practices.

### **4.3. Connecting Shared Practice to Shared Meaning**

Although development and the analysis of developmental processes are an integral part of cultural psychology, not all research programs address ontogenetic, intergenerational, or historic development directly. Adult cultural experience is also a focus of cultural psychology. An important theoretical issue is the extent to which shared practice results in shared cultural meaning. In the German tradition of cultural psychology, Lutz Eckensberger and colleagues used different microcommunities of practice to explore this issue with adults. The researchers interviewed different samples of people who had different views on a coal plant that was supposed to be built. The groups included people living in the area, experts on coal plants, environmental activists, and politicians. Each group could be considered to share activities and practices in relation to coal plants in general and this plant in particular.

The investigators used interviews to assess the moral meanings constructed by the different groups concerning the coal plant. Their interest was in the structure (e.g., complexity) of the reasoning rather than in its content. Findings revealed that each group had a typical pattern of moral reasoning about the coal plant. Thus, there was a link between activities common to each group in relationship to the plant and level of moral reasoning. The pattern of results indicates a connection between shared activities and the construction of shared structures of moral meaning.

## **5. APPLICATION OF CULTURAL PSYCHOLOGY**

As these examples show, cultural psychology can be applied to analyzing the everyday realities of members of different cultures and subcultures. However, there is another level of application that is also possible: applying findings to social interventions. We present one example. An implication of the modern-day multicultural reality in countries such as the United States or Germany is seen at

school. Multicultural immigration opens up the possibility of different cultural values among students, between students and teachers, and between home and school. An applied project, "Bridging Cultures," began with basic research by Greenfield, Quiroz, and Raeff documenting cross-cultural value conflict between Latino immigrant families and schools. Immigrant parents were generally much more collectivistic in their orientation to child socialization than were their children's teachers. In collaboration with a group of teachers, Greenfield, Trumbull, Rothstein-Fisch, and Quiroz utilized this research to help teachers and schools understand home culture and school culture in order to create educational "bridges" between them. The result of this intervention was greater home-school harmony in educational practices and classroom management. Shared practices and meanings were created between home and school.

## 6. CONCLUSIONS: AN EXAMPLE OF CULTURAL PROCESS

Both shared activities and shared meanings are intrinsic to the human mode of adaptation for survival. They are two facets of shared cultural knowledge and the primary subject matter of cultural psychology. An example will elucidate how shared activities and shared meanings are coconstructed in communities of practice; such coconstruction is the paradigmatic cultural process. The example illustrates that both material culture and symbolic culture result from processes of cultural coconstruction.

Following the Los Angeles earthquake of 1994, many of the material supports of everyday life, such as water, electricity, and roads, were destroyed. In small groups and through the media, people developed new shared knowledge concerning survival activities, such as where to get water, how to circumvent damaged roads to get from point A to point B, and methods for detecting leaking gas. Expertise was shared with novices, such as when a contractor showed his neighbors how to turn off their gas or a ham radio operator provided news of the location and magnitude of the earthquake in the absence of electricity. The nature of culture as a tool for organizing everyday life, a notion developed by anthropologist Thomas Weisner, was quite apparent.

Symbolic communication, through both language and visual media, is a critical means by which social sharing takes place; communication processes were quite intense during this period of adapting to the physical conditions created by the earthquake. As a result, new

shared activities that enhanced physical survival were created through cultural processes of social interaction.

Simultaneously, shared meanings were also created to rationalize and understand the events that had taken place. Like shared activities, shared meanings arose through communication processes. One shared meaning that developed was the custom of asking people how they fared in the earthquake; the normative reply was, "I was fortunate." The search to create shared meaning for a stunning physical event was particularly apparent when, a few days after the earthquake, a local public affairs radio show host convened clergy from many religions to ask them about the larger meaning of the earthquake. His question was, "Did God send the earthquake to punish Los Angeles?" Clearly, adaptation to the aftermath of the earthquake could not be reduced to a process of adapting to physical conditions; the interpretation of these conditions—that is, processes of creating meaning—were part and parcel for the shared culture that developed in response to the earthquake.

This example provides a model of processes that are assumed to occur whenever a new member of society is born: the creation of shared knowledge, activities, conventions, and meanings through communication and social interaction. This microdevelopmental example of culture re-creation by adults occurs in each generation in the macrodevelopmental processes of children. The example is also a model and metaphor of culture change provoked by new ecological conditions. Finally, this example illustrates the potential for cultural variability as a response to different ecological conditions. This is the stuff of cultural psychology.

### See Also the Following Articles

Child Development and Culture ■ Community Psychology ■ Conformity across Cultures ■ Cross-Cultural Psychology, Overview ■ Cultural Complexity ■ Diverse Cultures, Dealing with Children and Families from ■ Family and Culture ■ Indigenous Psychologies ■ Intergroup Relations and Culture ■ Motivation and Culture ■ Perception and Culture ■ System Safety ■ Values and Culture

### Further Reading

- Berry, J. W., Poortinga, Y. H., Pandey, J., Dasen, P. R., Saraswathi, T. S., Segall, M. H., & Kagitcibasi, C. (1997). *Handbook of cross-cultural psychology* (2nd ed.). Boston: Allyn & Bacon.
- Boesch, E. E. (2001). Symbolic action theory in cultural psychology. *Culture and Psychology*, 7, 479–483.

- Bruner, J. (1993). Do we "acquire" culture or vice versa? *Behavioral and Brain Sciences*, 16, 515–516.
- Cole, M. (1995). Culture and cognitive development: From cross-cultural research to creating systems of cultural mediation. *Culture & Psychology*, 1, 25–54.
- Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.
- Eckensberger, L. H., Döring, T., & Breit, H. (2001). Moral dimensions in risk evaluation. *Research in Social Problems and Public Policy*, 9 (Special issue).
- Enriquez, V. G. (1993). Developing a Filipino psychology. In U. Kim, & J. Berry (Eds.), *Indigenous psychologies: Research and experience in cultural context* (pp. 152–169). Newbury Park, CA: Sage.
- Greenfield, P. M. (1997). Culture as process: Empirical methods for cultural psychology. In J. W. Berry, Y. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology. Vol. 1: Theory and method* (pp. 301–346). Boston: Allyn & Bacon.
- Greenfield, P. M., & Cocking, R. R. (1994). *Cross-cultural roots of minority child development*. Hillsdale, NJ: Erlbaum.
- Greenfield, P. M., Keller, H., Fuligni, A., & Maynard, A. (2003). Cultural pathways through universal development. *Annual Review of Psychology*, 54, 461–490.
- Greenfield, P. M., Maynard, A. E., & Childs, C. P. (2003). Historical change, cultural learning, and cognitive representation in Zinacantec Maya children. *Cognitive Development*, 18, 455–487.
- Harkness, S., & Super, C. M. (1996). *Parents' cultural belief systems: Their origins, expression, and consequences*. London: Guilford.
- Keller, K., Völker, S., & Yovsi, R. D. (In press). Conceptions of good parenting in Cameroonian Nso and Northern Germans. *Social Development*.
- Kim, U., & Berry, W. (Eds.). (1993). *Indigenous psychologies: Research and experience in cultural context*. Newbury Park, CA: Sage.
- Miller, J. G. (2003). The cultural grounding of social psychological theory. In A. Tesser, & N. Schwartz (Eds.), *Blackwell handbook of social psychology. Vol. 1: Intrapersonal processes*. Oxford, UK: Blackwell.
- Romero, M. E. (1994). Identifying giftedness among Keresan Pueblo Indians: The Keres study. *Journal of American Indian Education*, 34, 35–58.
- Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Sinha, D. (1997). Indigenizing psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed., pp. 129–170). Boston: Allyn & Bacon.
- Trumbull, E., Rothstein-Fisch, C., Greenfield, P. M., & Quiroz, B. (2001). *Bridging cultures between home and school: A guide for teachers*. Mahwah, NJ: Erlbaum.



# Cultural Syndromes

Harry C. Triandis

University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

1. Examples of Cultural Syndromes
  2. Relationships among the Cultural Syndromes
  3. Practical Significance
- Further Reading

## GLOSSARY

**collectivism** A cultural pattern where the importance of the in-group is paramount; people think of themselves as aspects of groups, give priority to in-group goals, behave according to the norms of the in-group, and do not leave the in-group even if they are dissatisfied with it.

**cultural complexity** The contrast between simple cultures (e.g., hunters and gatherers) and information societies.

**cultural tightness** Cultures that have a very large number of rules and norms about social behavior and in which people are severely punished when they do not follow these rules and norms.

**cultures high in diffusion** Cultures that tend to think holistically, not differentiating among various aspects of situations.

**cultures high in specificity** Cultures in which people think analytically, distinguishing among various aspects of situations.

**egalitarian cultures** Cultures that emphasize equality.

**hierarchical cultures** Cultures that are organized according to status levels, where the dictates of those at the top of the social structure are given much weight and the opinions of those at the bottom of the social structure are not given much importance.

**individualism** A cultural pattern where the importance of the individual is paramount; people think of themselves as independent of groups, give priority to their personal goals, behave according to their attitudes, and easily leave groups with which they are dissatisfied.

**masculine cultures** Cultures that emphasize the differences between men and women and in which men rarely do the jobs that women do.

**outcome cultures** Cultures that emphasize the outcomes of action.

**particularism** Cultures in which behavior depends on the attributes of the person with whom one is interacting.

**process cultures** Cultures that emphasize how things are done.

**universalism** Cultures in which behavior is based on general principles rather than on the specific attributes of the person with whom one is interacting.

A cultural syndrome is a pattern of shared attitudes, beliefs, categorizations, self-definitions, standard operating procedures, unstated assumptions, norms, roles, and values that is organized around a theme. The theme is shared among those who speak a particular language, during a specific historical period, and in a definable geographical region. Cultural syndromes provide information about cultural differences. The number of existing cultural syndromes is unknown. However, some examples of cultural syndromes that have been identified are provided in this article.

## 1. EXAMPLES OF CULTURAL SYNDROMES

The following subsections describe five cultural syndromes that have been studied better than other cultural syndromes.

### 1.1. Cultural Complexity

The contrast between hunters and gatherers and information societies is vast. Simple cultures have few members, whereas complex cultures have many members. The number of people who constitute a cultural group is only one clue. Hunters and gatherers usually consist of bands of approximately 50 individuals who are related to each other through marriage or some other ceremonies. Information societies consist of millions of individuals who are sometimes related to each other through the Internet. In between, there is a very large number of levels of complexity such as “slash and burn” agriculture, regular agriculture, and industrial cultures.

Complexity is associated with the presence of writing systems, the presence of records, fixity of residence, agriculture, urban settlements, technical specializations, modes of transportation other than walking, money, high population density, many levels of political integration, many levels of social stratification, and the presence of different ways in which to make a living. For example, the number of occupations in simple cultures is small, whereas approximately 250,000 occupations have been identified in complex cultures. The number of choices available to individuals in simple societies is small, whereas it is vast (e.g., dozens of kinds of mustard) in complex societies. The number of standards for judging religious, economic, political, educational, social, and aesthetic phenomena is small in simple societies, whereas it is very large in complex societies. Thus, the organizing theme of this cultural syndrome is complexity.

### 1.2. Tightness

In tight cultures, there are many norms, standards, and rules for behavior. Those who do not do what is specified by these norms are criticized, punished, or even killed. In loose cultures, there are few norms, and people tolerate deviations from these norms. When a person does not do what is expected, people simply smile and say that it does not matter. Again, in between

these two extremes, there are thousands of levels of tightness. The organizing theme is doing what is specified by in-group norms.

Conformity is high in tight cultures. Extreme examples of tight cultures are the Taliban in Afghanistan and North Korea. Tight cultures emerge in societies that are relatively homogenous, so that people can agree concerning what norms are to be used as guides for behavior. They also emerge when there is much interdependence required to make a living. For instance, to get a crop, one might need an irrigation system, but an individual cannot construct such a system without help. In addition, when there is enough density in the population to make it likely that deviations from norms will be observed and punished, there is more cultural tightness. The organizing theme is the importance of norms.

Loose cultures are usually exposed to several normative systems, so that people must develop tolerance for behavior that does not follow a single set of norms. Thailand is a culture that has been influenced by both the Indian and Chinese cultures, so it is understandable that it is loose. The organizing theme is the unimportance of norms.

### 1.3. Collectivism

This cultural pattern occurs in most of the traditional cultures of the world, including East Asia, Latin America, and Africa, as well as among Asian Americans. In collectivist cultures, the self is defined as an aspect of a group such as the family or tribe. In addition, religious, political, occupational, aesthetic, educational, and athletic groups can function as in-groups. In these cultures, the self has much social content (e.g., “I am a member of this family,” “I am a member of this religious group”) and is interdependent with one or more in-groups. When there is a difference between personal and in-group goals, in-group goals have priority. Behavior is a function of both attitudes (what one would like to do) and norms (what one should do). Group-based relationships endure even if the individual would like to get away from them. The organizing theme is the importance of the in-group.

### 1.4. Individualism

This cultural pattern occurs primarily in Northern and Western Europe as well as among European Americans, Australians, New Zealanders, and others influenced by Western culture. In this case, the self is

defined as autonomous, is independent of groups, and has little social context (e.g., “I am busy,” “I am proud”). When there is a difference between personal and in-group goals, personal goals have priority. Behavior depends much more on attitudes than on norms. Group-based relationships are subjected to a cost–benefit analysis (e.g., “Do I get out of this relationship more than I have to put in?” [if so, stay in the relationship; if not, leave it]). The organizing theme is the importance of the individual.

Within any culture there is considerable variability concerning the presence of behaviors that correspond to collectivism or individualism. There are countercultural personalities in all cultures. People who tend toward individualism find collectivist cultures to be oppressive and so try to leave them. People who tend toward collectivism find individualist cultures to be cold and so join many groups (e.g., gangs, unions, communes).

#### **1.4.1. Some Contrasts between Individualism and Collectivism**

People in individualist cultures tend to see the self as stable and the environment as changeable (e.g., “If I do not like my job, I change jobs”). Conversely, people in collectivist cultures tend to see the environment as stable and themselves as changeable or ready to “fit in.”

In individualist cultures, people are most likely to sample cues about events “inside” other people (e.g., beliefs, attitudes, values), whereas in collectivist cultures, they are most likely to sample cues about external events and social entities (e.g., norms, roles, situations, social structures, agreements, intergroup conflict).

In 1999, Norenzayan and colleagues reviewed evidence showing that when East Asians make dispositional attributions, they see traits as quite malleable, whereas when Western individualist samples make dispositional attributions, they see them as fixed. The authors reviewed a wide range of information, from laboratory studies to ethnographies, and concluded that probably all cultures make dispositional attributions but that cultural differences occur because people from East Asia make situational attributions much more frequently, and to a greater extent, than do people from the West.

When people communicate in individualist cultures, they sample the content most heavily, whereas when people communicate in collectivist cultures, they sample the context of the communication (e.g., level of voice, eye contact, gestures, emotional expression) most heavily.

Antecedents of individualism include affluence, migration (leaving the in-group), leadership roles, education, and living in heterogeneous environments (e.g., large cities). Antecedents of collectivism include poverty (survival depends on the help of the in-group), stable residence, low social class roles, and living in homogeneous environments (e.g., rural or simpler societies).

#### **1.4.2. Kinds of Individualism and Collectivism**

These cultural patterns take different forms in different cultures. Specifically, there are horizontal and vertical individualist and collectivist cultures. Horizontal individualist cultures, such as Sweden, emphasize that the individual is independent from groups and is self-reliant but also that people do not wish to “stick out.” Modesty is a virtue. People wanting to be “the best” characterize vertical individualist cultures such as the corporate and academic cultures in the United States. Competition is high and modesty is not a virtue in such cultures.

Horizontal collectivist cultures, such as the Israeli kibbutz, emphasize interdependence of the individual and the group, but there is little hierarchy. In theory, every member does all of the jobs, whether the jobs are prestigious or not. In vertical collectivist cultures, such as China and India, sacrifice for the group is virtuous. The individual is not important, and the group is all-important.

Another variety of collectivism is the one that stresses the relationship of children to all of the women in the extended family, as occurs in India, where there is no particular emphasis on the mother–child relationship. This pattern contrasts with the Japanese pattern, where the mother–child relationship is all-important and the relationship of the child to the women of the extended family is not important.

### **1.5. Hierarchy Versus Equality**

This pattern is the same as the vertical–horizontal pattern just discussed, but it is useful to categorize it as a separate syndrome because it is very important and can be conceived independently from individualism–collectivism. In hierarchical cultures, people pay a lot of attention to the status of individuals when organizing their social behavior. In cultures that stress equality, people espouse many ideas, such as “one person/one vote,” that stress the importance of equality. The organizing themes are the importance or unimportance of status.

The syndromes that follow have not been studied well but are included here to suggest future research.

### **1.6. Masculinity Versus Femininity**

In masculine cultures, men rarely do the jobs that women typically do. In feminine cultures, both men and women tend to do most of the tasks of the society (e.g., child rearing). The organizing theme is the importance or unimportance of being male.

### **1.7. Pragmatic Induction Versus Ideological Authoritarianism**

This syndrome contrasts cultures in which people make most judgments based on information that comes from experience with cultures in which people make most judgments based on information obtained from ideological sources such as leaders, tribal chiefs, and gods. The organizing theme is empiricism versus ideology.

### **1.8. Short Versus Long Time Perspective**

In cultures with a short time perspective, the emphasis is on the “here and now,” whereas in cultures with a long time perspective, the emphasis can be on the past or future. The organizing theme is the present versus the past or future.

### **1.9. Valuing Planning Versus Spontaneity**

In cultures that value planning, much effort is expended in planning, whereas in cultures that value spontaneity, planning is considered to be an imposition on one’s freedom of action. The organizing themes are planning versus spontaneity.

### **1.10. Universalism Versus Particularism**

In universalistic cultures, one uses general principles such as “to each according to his or her contribution,” whereas in particularistic cultures, one takes into account who the other person is, such as in-group membership, previous accomplishments, sex, age, and social class, and one gives what is appropriate for a person who has those particular qualities. For example, in waiting for food at a cafeteria, universalistic

cultures use the principle of “first come/first served,” whereas particularistic cultures unquestionably allow high-status persons to go to the front of the line. The organizing principle is the use of universalistic criteria (one behaves in a certain way no matter who it is) versus particularistic criteria (behavior depends on who the other person is).

### **1.11. Specificity Versus Diffusion**

Specific cultures tend to be analytical, paying attention to each element of information such as who said what, when, where, and why, whereas diffuse cultures tend to be holistic, assuming that if a positive (or negative) element is present in one element, it is also present in all of the other elements. The organizing principle is specificity of the stimuli versus the diffuse reaction to the stimuli.

### **1.12. Process Versus Outcome**

In some cultures, people emphasize the process (e.g., “This is the way we do things”), whereas in other cultures, people pay little attention to the process but instead emphasize the outcome (e.g., “We must win”). Thus, in the former kind of culture, one pays attention to how something is done, whereas in the latter kind of culture, the emphasis is on what happens as a result of the action. The organizing principle is emphasis on process versus emphasis on outcomes.

### **1.13. Additional Syndromes**

In addition to the syndromes just listed, there are many narrow syndromes. There is a literature, for instance, that describes syndromes whose organizing principles are honor, achievement, competition, and so on.

## **2. RELATIONSHIPS AMONG THE CULTURAL SYNDROMES**

There is some evidence that syndromes are not unrelated to each other. For example, cultures that are tight tend to be collectivist. In general, collectivist cultures tend to be simpler and tighter, whereas individualist cultures tend to be more complex and looser. But there are many exceptions to these patterns. One usually finds more masculinity in collectivist cultures and more emphasis on pragmatic induction in individualist

cultures. Planning is more likely in individualist cultures. Particularism is more common in collectivist cultures. It is best to measure each cultural syndrome independently of the measurement of the other syndromes.

### 3. PRACTICAL SIGNIFICANCE

There is a correlation between some of the cultural syndromes and specific patterns of behavior. For example, industrial organizations in collectivist cultures are more likely to hire individuals who are in-group members regardless of their level of competence. In collectivist cultures, people tend to think of individuals as being easy to change and ready to adapt to different groups. Thus, it is understandable that they think that any in-group member can become a good employee. Loyalty to the organization, and especially to its authorities, is valued more in collectivist cultures than in individualist cultures. Thus, hiring loyal in-group members is more likely in collectivist cultures than in individualist cultures.

Collectivist managers are more likely to promote on the basis of tenure in the organization than on the basis of individual achievement. Also, collectivist employees prefer to be paid based on the achievement of the group rather than on the basis of individual achievement. But other things being equal, there is more organizational commitment and more training offered to employees in collectivist cultures than in individualist cultures.

Although on average members of a culture may have a particular attribute, there is very high variability within each culture. Thus, one can find employees who are behaving in highly individualistic ways in collectivist cultures and can find employees who are behaving in collectivist ways (e.g., joining unions) in individualist cultures. There are organizations that are individualist (e.g., academia) and organizations that are collectivist (e.g., the military) in all cultures.

When distributing resources such as bonuses, collectivists are likely to take into account the needs of employees (e.g., they have many children) or distribute equal amounts to each employee. In individualist cultures, bonuses go to those who have contributed more to their organizations.

The ideal leader is a benevolent father figure in collectivist cultures. The ideal leader is warm and supportive, but he or she also emphasizes production. Paternalism is a desirable leadership pattern in many collectivist cultures but not in individualist

cultures. In fact, top management is much more aware of the private lives of employees in collectivist cultures than in individualist cultures. For example, top management in collectivist cultures likely will send condolences to an employee who had a death in the family (e.g., when an employee's parent dies), whereas top management in individualist cultures might not even be aware of the death of an employee's parent. Also, management in collectivist cultures likely will send congratulations when an employee's child graduates from high school or is accepted by a college.

Cultural differences in tightness also have implications. For example, people follow rules and specifications in manufacturing much more rigidly in tight cultures than in loose cultures. Improvisation and exploration of different ways in which to do the job are more common in loose cultures.

Differences in the way in which time is used provide a major difficulty in getting along with people from other cultures. Some cultures use monochronic time (e.g., holding one conversation at a time), whereas other cultures use polychronic time (e.g., holding several conversations with different people at the same time). When people from the former cultures are not aware of this cultural difference, they are easily offended by people from the latter cultures. Another contrast is between work time (i.e., time to work and not socialize) and social time. The definitions of what are work time and social time are culture bound. In some cultures, all of the farmers work together on the plot of one farmer and then go together to work on the farm of another farmer, and so on, and all the time they mix work time and social time. Still another contrast is between doing time (i.e., when one is supposed to do things) and not doing time. The desirability of "doing nothing" is very culture bound. In some cultures, it is the ultimate goal, whereas in other cultures, it is associated with utter boredom. Finally, the contrast between clock time (e.g., meetings end at a certain time) and event time (e.g., meetings end when the purpose of the meeting has been accomplished) can be the source of many misunderstandings.

### See Also the Following Articles

Conformity across Cultures ■ Cultural Complexity ■ Cultural Psychology ■ Industrial/Organizational Psychology across Cultures ■ Intergroup Relations and Culture ■ Leadership and Culture ■ Motivation and Culture ■ Perception and Culture ■ Values and Culture

**Further Reading**

- Carpenter, S. (2000). Effects of cultural tightness and collectivism on self-concept and causal attributions. *Cross-Cultural Research, 34*, 38–56.
- Chick, G. (1997). Cultural complexity: The concept and its measurement. *Cross-Cultural Research, 31*, 275–307.
- Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An “experimental ethnography.” *Journal of Personality and Social Psychology, 70*, 945–960.
- Glenn, E. (1981). *Man and mankind: Conflicts and communication between cultures*. Norwood, NJ: Ablex.
- Levine, R. (1997). *A geography of time*. New York: Basic Books.
- Markus, H., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253.
- McClelland, D. C. (1961). *The achieving society*. Princeton, NJ: Van Nostrand.
- Norenzayan, A., Choi, I., & Nisbett, R. E. (1999). Eastern and Western perceptions of causality for social behavior: Lay theories about personalities and situations. In D. A. Prentice, & D. T. Miller (Eds.), *Cultural divides: Understanding and overcoming group conflict* (pp. 239–272). New York: Russell Sage.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.



# Cyberpsychology

José M. Prieto

Complutense University, Madrid, Spain

1. Cyberspace: A Technocultural Realm
2. Cyberpsychology: An Increasingly Grounded Brand Name
3. Cyberpsychology: Six Perspectives as Operational Tools  
Further Reading

## GLOSSARY

**cybernetics** A discipline devoted to the study of information processing and communication systems as well as the functioning of control systems when applied to machines, animals, and organizations.

**cyberpsychology** A new realm of psychological research and action that uses cyberspace as an instrumental means toward an end, analyzes and identifies the underlying dimensions that structure and make sense in cyberspace, studies the cognitive and social processes that facilitate or impede the involvement of cyberspace, discloses pragmatically designed principles leading to success in cyberspace, pays attention to personality disorders or social conflict arising in cyberspace, and introduces quality filters for searching and evaluating Internet resources.

**cyberpunk** A person who knows how to use computers as a means of self-expression and as a tool for producing virtual worlds in which reality and fantasy are intertwined and the individual person becomes a character in a scenario.

**cyberspace** An imaginary and multidimensional space in which people can absorb sensory experiences, enjoy intellectual creations, and interact cognitively and emotionally. It is also an economic survival kit for the information society.

**cyborgs** Initially, fusions of machine and organism, animate and inanimate at the same time. Later, new-age chimeras

blurring mythic hybrids such as man/machine, male/female, white/black, and homosexual/heterosexual, created for pleasure by fine artists and graphic designers; fabulous creatures amalgamating over-elaborate features rooted in science fiction and social reality.

**digital game-based learning** A type of learning through use of a fusion of computer games and video games in which the game player is an actor exploiting the interactive possibilities, the climate of incertitude, and the curiosity appeal of a given game to attain learning goals.

**e-learning** The use of Internet and digital TV technologies to make available a broad array of information and databases that must be transformed in the learning process to produce sound and reliable knowledge and to facilitate human performance in real-life settings.

**homepage** Opening page of a Web presentation that provides links to other documents and that becomes the starting page to which secondary documents are linked. It is often the basic platform of a given Web-based or Web-linked organization.

**hypertext** Basic unit of associated sections within and outside a text connected to each other nonlinearly. Hypertext was invented by Douglas Engelbert in the mid-1960s, and it has become the underlying podium for the availability of documents in cyberspace.

**knowledge management** Active and systematic process of managing information to transform it into sound knowledge by capturing, creating, codifying, distributing, and filtering among networked users. Sometimes it is a substitute for the concept of learning.

**telematics** An English transposition of the French expression *telematique*, derived from the expression *teleinformatique*, that may be translated as “networked computers” or “wired computers.”

**virtual reality** The computerized simulation of an entire three-dimensional surrounding in which a person can perform actions or manipulate objects by clicking on the mouse. Visually the degree of realism is very high but kinetically leaves a lot to be desired.

**World Wide Web (Web)** A hypertext-based information network rooted in internationally accepted standards. It was developed in 1990 at the European Centre for Particle Physics by Tim Berners Lee.

This article has been divided into three sections. The first analyzes the emergence of cyberspace as a counter culture, emphasizing distinctive backgrounds. The second analyzes how cyberpsychology is a brand new name in the phase of increasing its visibility within the domain of applied psychology. The third pays attention to six different strategies used by applied psychologists when they deal with Internet and multimedia databases and hyperdocuments.

## 1. CYBERSPACE: A TECHNOCULTURAL REALM

There is a hidden nexus between the basic unit of electric current, known as the ampere, and “cyber,” a multifaceted root present in a large variety of domains within the umbrella of new information and communication technologies (NICT). The term ampere expresses homage to André Marie Ampère (1775–1836), a French physicist, mathematician, and philosopher. He coined the term cybernetics in 1834 for a prospective new science devoted to the overseeing and control of governments. In classic Greek, *κυβερνάω* means to lead or to pilot, and so the derived noun, *κυβερνητικός* means the pilot. Nobert Wiener (1894–1964) recaptured the word cybernetics in 1948 as a title for his book on “the science of control and communication in the animal and the machine.” Later, in 1950, he emphasized that cybernetics should ultimately be regarded as a matter of “the human use of human beings,” highlighting the nexus between learning and feedback as both technological and cultural frameworks in the functioning of control and governing systems.

During the 1950s and 1960s, cybernetics became a widely accepted label for studying (1) the processing of information and communication systems, (2) the flow of information within a system, and (3) the use of feedback to get intended impacts on goal-directed

activities cropping up in technological artifacts as well as in living organisms and organizational settings. Soon cybernetics lent its name and the new root “cyber” started to circulate, becoming used by computer-literate university graduates in a large variety of disciplines.

In 1970, Alvin Toffler coined the term “cyborgs” in his book *Future Shock*, which devoted several pages to analyzing the possibilities of human–machine integration and the interaction of human brains and databases through networked communication. Initially, a cyborg was described as a fusion of machine and organism, animate and inanimate at the same time. Mechanical or electronic devices built into the body allowed its physiological functioning to extend some abilities beyond normal or to compensate for some disabilities. Psychological expertise was demanded during the design process to optimize performance levels or to smooth training processes and after the surgical implant of devices to facilitate the psychological adjustment of ongoing interactions.

During the 1980s, the conception of cyborgs transformed classical dichotomies into mythic hybrids by blurring binary pairs such as man/machine, male/female, white/black, and homosexual/heterosexual. At first glance, it was a change of focus: cyborgs started to be created for pleasure. Fine artists and graphic designers crossed the line of gender representations and pushed their art into the awareness of the sexual impact of cyberbodies. So, cyborgs became sexy, and a diversified fauna of new-age chimeras emerged. “A cyborg is a cybernetic organism, a creature of science fiction and a creature of social reality,” highlighted Haraway (1991) in her seminal essay “Cyborg Manifesto” published in 1985. From a psychological perspective, these utopian cyberbodies induced morbid fascination and were ready for scrutiny as the latest taste in projective techniques. A cyborg is a cybernetic organism, a creature of science fiction and a creature of social reality.

The term “cyberpunk” started to be used as an argot among writers in counter-cultural circles in the mid-1970s, for example, by John Brunner in his novel *Shockwave Rider* and by William Gibson in several science fiction stories collected in *Neuromancer*. Soon, some creative writers began to call themselves “cyberpunks,” stressing that they used computers and video games as a means of self-expression, personal pleasure, net profit, and out of a sense of duty. In fact, they ascribed the roles of main or secondary characters to “first-generation cyborgs,” that is, microsurgically reworked characters, showing a disembodied

intelligence, bringing to the fore an anarchist cultural background melting away into a high-tech mood and open-minded attitude. In cyberpunk stories, the narrative usually took place in postindustrial and information-governed settings populated by urban misfits, dead reconstructions of people previously alive or artificially smart beings ready to think for themselves and question authority. Cyberpunks were often depicted as skillful people with expertise in knowing how to take advantage of NICT to attain goals that were often malevolent.

During the 1980s, fiction writers realized that they had developed a set of beliefs about the existence of some kind of actual space behind the screen, “a place that you cannot see but you know is there.” This world of lucid dreaming beyond a television or a computer screen came to be known as cyberspace, an imaginary and fictional universe in which sensory experiences take place, the mind is absorbed, and the person feels as one with the set of stimuli, challenges, and performances elicited. It generates a trancelike experience, actual and genuine, known as “consensual hallucination,” the assertion coined by cyberpunk writers. Cyberspace was a place where individuals were welcome if they knew how to stroll along boundless passageways, entering and exiting large intellectual creations made available online. It was a “World III” structure in the terminology launched by Sir Karl Popper (1902–1994).

Cybersex was a term introduced during the 1990s to identify the existence of sensual and hedonistic waterways irrigating the digital underground of cyberspace. It is the domain of a large variety of incorporeal intercourses, sometimes known as “robotopulation” or “onanism for two,” that make the most of computer-generated hyperrealism as well as mind-to-mind networked but yet dreamed scenarios. As Dery put it, “sex with machines, together with dalliances conducted in virtual worlds, seems a seductive alternative in an age of AIDS, unwanted pregnancies and sexually transmitted diseases” (1996, p. 199). It is a kind of online sex play, X-rated and interactive computer programs adapted for the stage by similarly equipped participants recurring to an array of sensor-effectors ready to generate a realistic sense of tactile presence.

Bell and Kennedy and Spiller have compiled large sets of difficult-to-find articles on cyber-cultural challenges and dilemmas, published over several decades during the second half of the 20th century, that otherwise would be rather inaccessible.

## 2. CYBERPSYCHOLOGY: AN INCREASINGLY GROUNDED BRAND NAME

The term cyberpsychology started to circulate in cyberspace in 1994 when Prof. Leon James at the University of Hawaii wrote a pioneering paper, “Cyberpsychology: Principles of Creating Virtual Presence,” which was expanded over a decade and made available online. As a result, some post-modern views on psychology and Jungian connotations were launched: an ineffable continuity of an online communal mind exists and grows in cyberspace as a self-ruling collective unconscious. The existence of a critical mass in interactions and transactions “in any topical zone and the type of ongoing activity” suggest the emergence of a communal mind in the Internet. “Topics and activities in cyberspace create their own virtual zones that become accessible to others across time and space” and give the impression of omnipresence in such a communal mind. The overlap between cyberspace and the mind was asserted under the umbrella of cognitive psychology. Computers were appraised as “convenient and powerful extensions of the human mind” and the consequence was that “every characteristic of the mind can be expected to show up as a property of the cyberspace.” The basic argument justified a new denomination such as cyberpsychology based on the certitude that “cyberspace, like mind, is not in physical space, but in virtual space, without extension, distance, or mass.”

In 1994, José M. Prieto at the University of Madrid set up the homepage of the International Association of Applied Psychology (<http://www.iaapsy.org>), which included the online proceedings of the 23rd International Congress of Applied Psychology. Nielsen, the leading expert in usability challenges, advances, and dilemmas in the production of multimedia and hypertext documents, mentioned this initiative as an illustration of navigating large information spaces. Soon afterward, Prieto began to focus on cyberspace as a setting where psychological teaching, learning, and research programs may take place and where spatial orientation and time management skills play a very significant role in online performance. The initial focus was outlined in the paper “Psychology and Telematics” (a new term derived from the combination of “tele” and “informatics”), and the next step was the interface between literacy in NICT and the involvement of work and organizational psychologists in knowledge management programs. In 1997, it became an e-learning space

nicknamed *prietolandia* by psychology students: an online learning experience for university graduates in psychology from 1995 to 2003, when the program came to an end.

In 1996, Prof. John Suler at Rider University started to write an electronic book, *The Basic Psychological Qualities of Cyberspace*, made available online, another project in a course of continuous expansion. The point of departure was an intensive case study carried out on the psychological and social interactions held by members of a very vivid and evolving community who used to meet in a quiet notorious virtual space known as "The Palace" (<http://www.thepalace.com>) that closed by the end of the 1990s but was again active by 2004. In 1995–1996 it was a visual, auditory, and chat environment where visitors adopted and role-played a large variety of characters by using graphical representations in the realm of pure fantasy and fiction. Suler used the participant observation technique to study cyberspace as "a psychological space" and as "a dream world." The psychoanalytic as well as field theory perspectives shaped the theoretical and operational background, making sense of what emerged in this research project. Internet users felt comfortable because somehow cyberspace was "an extension of their mind and personality—a 'space' that reflects their tastes, attitudes, and interests." Following a psychoanalytical vein, cyberspace was interpreted as a "transitional space" between the self and the other.

A peer-reviewed journal entitled *Cyberpsychology and Behavior* was launched in 1998 (<http://www.liebertpub.com>) and made public during the 24th International Congress of Applied Psychology and the 106th Annual APA Convention, both held in San Francisco in August 1998. Since the year 2000, the electronic and the printed version have been accessible at the same time; subscribers have access to each online issue via a password.

Gackenbach compiled 13 contributions written by 20 authors from Australia, Canada, Israel, the United Kingdom, and the United States focusing on the psychology of Internet use. Special attention was paid to the normal and deviant aspects of the self when staying online, lively relationships maintained during close and distant interactions, and transpersonal issues such as the development of consciousness and collective unconscious. Through surveys and case studies, the incidence of Internet addiction was examined and its presence was restricted to "an exceedingly tiny minority" (p. 71). This is a typical example of research findings that refute comments in the mass

media occasionally introduced by practitioners expressing opinions without reliable data on the subject.

Wallace approached the Internet within a psychological perspective; the outcome was a book titled *The Psychology of the Internet*. It focused on the "online persona" and the way Internet users generate impressions, use masks as ID during interactions, and become involved in group dynamics through conformity and cooperation as well as conflicts and aggressions. She studied how liking and loving take place on the net, calling attention to how interpersonal attractions arise when people try to find out who is next door on the net.

Fink is the lead author of a book devoted to analyzing how to use computers and cyberspace in the clinical practice and psychotherapy. There are 10 separate contributions that each highlight specific topics. Together, they scrutinize the challenges faced by professional psychologists determined to transmute the conventional couch into an online couch, to explore the self by inviting the patient to produce a personal Web page, to go after the consequences of a disembodied gender in cyberspace, and to go along with the consequences of virtual health-care programs. A catalogue of Internet resources for clinical psychologists was added as a bonus.

Gordo-López and Parker, in a book entitled *Cyberpsychology*, brought together a set of 14 articles written by 16 psychologists and social science scholars from Europe, the United States, and Latin America, bringing into focus the relationships between psychology and cybernetics. The prevailing approach is rather theoretical, rooting psychology into the technological background of cyberspace without endorsing fantasies of liberation that surround NICT in the mass media and the popular culture regarding the Internet. The approach is rather interpretative: "Cyberpsychology needs to embed within itself a self-annihilating device, a certain kind of critical and self-critical narrative"; "We want to insist that it is only in its ephemeral use that the potential of cyberpsychological critique can function" (p. 16).

Birnbaum compiled 12 contributions produced by 23 authors from Canada, Germany, Italy, Switzerland, the United Kingdom, and the United States. Studies carried out in the Internet were compared to similar studies carried out in laboratories or in field studies, stressing methodological aspects to be taken into consideration when deciding which is the appropriate medium to use. The role of individual differences and cross-cultural differences as identified through

online research was also examined. Potential uses of advanced computer techniques or pondering the appropriateness of server-side solutions reveal weak points in design as well as the adequacy of server-client relationships. This is an Achilles heel rarely assessed in papers.

Wolfe compiled 12 contributions written by 19 authors from Europe and the United States that examine cost-effective advances and achievements attained by the use of learning technologies placing the development of Web resources on a consistent psychological foundation. It combines theoretical advances and empirical findings and moves from the frame of reference of the individual learner to that of learning communities in university campuses. One out of two college courses list Web resources in the syllabus, and one out of four college courses have their own Web pages, thus producing a gap between those making an offer of online and offline handouts or practical exercises, for instance.

Prensky has emphasized the importance of computerized games as educational and training tools that meet the needs and learning styles of people wishing to improve their expertise by using simulations and theatrical scenarios. There is in fact a generational gap between educators belonging to a pre-digital cohort and students raised in an audiovisual and digitalized cultural background. Digital game-based strategies allow the acquisition of very specific skills on demand for specialized jobs and risky situations. Findings and developments founded on studies derived from the psychology of play provide the background for the cognitive and emotional involvement of students galvanized as players.

Reips and Bosnjak compiled 20 contributions written by 40 authors from Europe and the United States. These papers obtained expert advice and commentaries from two rounds of double peer review from 32 independent reviewers, evidencing high standards in standard publication procedures. They examine the Internet both as an instrument for, and an object of, scientific research, and approach issues such as psychological Web experiments, Web questionnaire studies, studying perception on the net, net-based surveys, and communication research as well as knowledge acquisition and learning throughout cyberspace.

Stein has written an essential Baedeker for students and scholars, a traveler's guide to the topography of cyberpsychology. His book is a comprehensive and accessible overview of what resources are

available, how to track down subject-specific materials and use them efficiently, and how to compile Internet references for term papers, reports, and dissertations. The author has also provided an online catalogue of appropriate hyperlinks mentioned in the book.

### **3. CYBERPSYCHOLOGY: SIX PERSPECTIVES AS OPERATIONAL TOOLS**

It is possible to distinguish six quite different strategic approaches followed by applied psychologists when they use the Internet and multimedia hyperdocuments in daily professional activities. Private life approaches are not considered here, but they are always embedded as background. The six professional strategies advanced are (1) instrumental, (2) dimensional, (3) process-based, (4) pragmatic, (5) psychopathologically biased, and (6) knowledge management-based.

#### **3.1. A Tool for Psychologists**

The first strategy approaches online computer-based resources and support facilities as a tool for conducting a large number of activities that psychologists carry out regularly in university campuses as well as in professional practice. It is an instrumental way of understanding psychology within the cyberspace, the nexus "means to an end." Cyberspace is viewed as a means to (1) communicate via email or transmit documents, (2) participate in discussions via interactive groups and Web-based conferences, (3) exchange or supply files and programs, (4) carry out transactions with customers, and (5) increase visibility across borders. Thus, cyberspace is valued as a useful territory that facilitates the achievement of certain ad hoc aims, rooted in psychological action and effective performance. Psychology students and graduates typically learn to handle only a few of the tools available, those they become comfortable with and use for certain purposes. Newer tools are often avoided because they demand too much time and imply too much uncertainty: only members of younger cohorts are open minded enough to try them. This attitude contrasts with the hedonistic ways of understanding cyberspace: a stage for playfulness and inventiveness, a joyful scenario for divergent thinking among advance expert

knowledge-based groups ready to fix acceptable standards in 5 or 10 years.

### **3.2. Alternate Dimensions of Experience**

The second strategy deals with psychological dimensions of computer-based settings, online environments, and personal and professional grounds of performance instigated by the involvement in computer networks. It is the realm of magnitudes and measurements present in online phenomena and online users. It brings to the forefront the psychological analysis of personality questionnaires, for instance, producing comparisons between factor structures obtained through offline and online versions of the same instrument. Present evidence suggests equivalence regarding what they measure and how well they serve the intended purpose. Similar approaches have been tried in the study of abilities, skills, interests, and values underlying online tasks and activities examined by researchers to obtain the factors or clusters underlying effective or ineffective performance. It allows the identification of multifaceted features such as experience requirements, cyber-users' characteristics and requirements, and communalities and specificities among members of relevant virtual communities. Follow-up studies of observable response patterns obtained through Web-based surveys favor the reconstruction of response undertakings and the identification of specific typological profiles different from those obtained via offline surveys. For example, controlling the display of questions and answering actions allows the number of questions displayed and the number of answers marked to be taken into consideration. Findings support the distinction between "performers" and "explorers" as basic typology. Performers answer all questions displayed, whereas explorers read initial set of instructions and preliminary items and quit suddenly, read the full set of questions screen by screen but never answer, or answer a few or even many questions but do not conclude the protocol. The psychological study of group structures in cyberspace also follows a dimensional approach, since it allows comparisons within and between communities to identify overt or hidden fraternities in communication networks, leadership styles in online chats, and prevailing norms and roles among users when they are on stage in a Web conference room, for example. Psychological dimensions are

also the basis for the study of human diversity in cyberspace.

### **3.3. Psychological Processes and Human Performance**

The third strategy addresses the issue of the psychological processes associated with human performance in networked computers and hypermedia. This process-based notion of cyberpsychology stresses that analysis may be carried out at the individual or at the group level and that it requires follow-up controls and strategies. The distinction between text-based and hypertext-based handouts used in online seminars has served to analyze operational and cognitive interactions concerning navigation, information scrutiny, and knowledge acquisition. Feelings of getting lost while navigating hypertext documents have been detected, and organization and presentation schemes have been suggested and validated to provide reliable frameworks for interested readers. For instance, the use of frames that include a contents index or a graphic map synchronized to sequential actions seems to facilitate the orientation of individual readers. Generational gaps have been remarked upon, and the presence or absence of dynamic icons accentuates the impression of "cool" versus "puerile" settings and thus acceptance or rebuff, by cohorts, of information presented. Eye movement screenings allow the analysis of how much time readers stay absorbed in a given text paying or not paying attention to navigation bar as compared to readers looking through a hypertext document. The navigation process seems to distract persons who are not experts in the subject from the actual text; they retrieve more information compared to those who are expert in the subject who can skim the text with hypertext advantages. Sensation, perception, attention, memory, knowledge acquisition, conscious experience, tacit learning and so on play an active role in the way people behave in cyberspace and have been examined, capitalizing on models and procedures developed in cognitive psychology. The study of group processes in cyberspace enables the discernment of formation, commitment, and identity within a given virtual community, as well as the understanding of influences pervading each community, of decision-making patterns and the value of minority dissent, of conscious and unconscious phenomena, and of autonomy and effectiveness.

### 3.4. Optimizing the Cyberspace Experience

The fourth strategy tackles the understanding of useful principles and guidelines based on psychological grounds that may improve the success of a home page, a portal, or an online e-business or e-commerce action plan. It is a pragmatic notion of cyberpsychology that looks for psychological theories, models, or findings that may refine communication patterns, learning strategies, and the intended presence and visibility of individuals and groups in cyberspace. The focus of attention is the identification of how cyberspace can be used to isolate and advance a psychological elucidation of behaviors and actions and the way of influencing its appearance or withdrawal. For instance, Gestalt psychology principles, formulated at the turn of the 20th century, are still an adequate framework in which to ameliorate the design of multimedia pages and navigation tools. Tacit learning theories have been implemented to reduce the number of procedural steps in online commercial or stock market transactions. Heuristic techniques have been used to optimize the testing of complex hypermedia documents to identify missing links, errors, intended loops, or unfortunate paths, before producing the final version. The pragmatic approach favors a focus on online behavior as natural, on online inventiveness in problem-solving strategies as natural, and on accepting a continuous flow of new online standards and norms as natural.

### 3.5. Cyber Pathology

The fifth strategy focuses on online psychological disorders; online overuse is the recurring example. In fact, it is the consequence of a parody written in 1995 by Ivan Goldberg, a psychiatrist who is quite active online, that made fun of the pathological gambling criteria described in DSM-IV. This resulted in a psychopathologically biased notion of cyberpsychology pivoting on reported cases of pornography, interpersonal or sexual attraction, disembodied gender, marital problems, neglect of children, and so on, as broadcast by the mass media. The high visibility of this perspective resulted in focus on antisocial and manipulative behavior online. However, it must be stated that cyberspace just mirrors a person's regular social behavior and is simply a new scenario for interrelating with others. There is no evidence suggesting that cyberspace may be the direct cause of new mental or emotional problems. It is more a matter of inappropriate virtual behaviors such as (1) spam, invasive, unsolicited, and

superabundant email, (2) "spoof," a message sent with the sender's identity concealed, and (3) "flaming," communication that is perceived as an insult, a direct attack, abuse, insolence, or rudeness. The large majority of cyberspace dwellers understand the consequence of spamming or spoofing because their email inboxes regularly receive a large number of unwanted or forged email messages. These behaviors have highly detrimental consequences, since many derived costs are borne by the person receiving the messages instead of the sender. Paradoxically, these inappropriate behaviors can shock online users, but usually are not considered scandalous or outrageous by offline users such as investigative journalists or psychologists in the field. This is just one example of double standards regarding criminal behavior in cyberspace versus in "real life."

### 3.6. Information and Knowledge Management

The sixth strategy views cyberspace as a huge virtual library that facilitates access to information in accounts, advertisements, books, brochures, databases, discussion groups, documents, journals, manuals, newsgroups, papers, journals, newsgroups, proceedings, software, surveys, syllabi, and so on. At first glance, cyberspace can be viewed as a storehouse ready to grasp the interest of psychologists interested in keeping an open mind to a large variety of old and new subjects, theoretical, technical, or practical. However, the use of filters is required to separate the wheat from the chaff. The same criteria that are useful in the analysis of scientific papers can be used to identify helpful online resources: (1) intended audience, (2) authority on the subject, (3) amount and type of coverage (e.g., in-depth versus superficial), (4) milieu in which the information is accessible, (5) objectivity versus subjectivity in distinguishing facts from values or opinions, (6) precision according to reliable sources, (7) topicality (i.e., high or low profile subjects) and, (8) the ease of use of the interface.

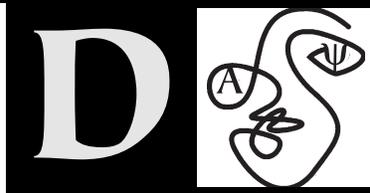
### See Also the Following Articles

Internet Counseling

### Further Reading

- Bell, D., & Kennedy, B. M. (2000). *The cyberculture reader*. London: Routledge.
- Birnhbaum, M. H. (2000). *Psychological experiments on the Internet*. San Diego, CA: Academic Press.

- Dery, M. (1996). *Escape velocity: Cyberculture at the end of the century*. London: Hodder & Stoughton.
- Fink, J. (1999). *How to use computers and cyberspace in the clinical practice of psychotherapy*. Northvale, NJ: Jason Aronson.
- Gackenbach, J. (1998). *Psychology and the Internet: Intrapersonal, interpersonal and transpersonal implications*. San Diego, CA: Academic Press.
- Gordo-López, A. J., & Parker, J. (1999). *Cyberpsychology*. London: Macmillan Press.
- Haraway, D. (1991). *Simians, cyborgs and women: The reinvention of nature*. New York: Routledge.
- Nilsen, J. (1994). *Multimedia and hypertext: The Internet and beyond*. Cambridge, MA: Academic Press Professional.
- Prensky, M. (2001). *Digital game-based learning*. New York: McGraw Hill.
- Prieto, J. M., & Sanz, P. (2002). Cyberpsychology. In *Encyclopedia of life support systems*. Oxford, UK: EOLSS. <http://www.eolss.net>.
- Reips, U. D., & Bosnjak, M. (2001). *Dimensions of Internet science*. Lengerich, Germany: Pabst Science Publishers.
- Spiller, N. (2002). *Cyber-reader: Critical writings for the digital era*. London: Phaidon.
- Stein, S. (2003). *Psychology on the Web: A student guide*. Essex, UK: Pearson Education.
- Wallace, P. (1999). *The psychology of the Internet*. Cambridge, UK: Cambridge University Press.
- Whittle, D. B. (1997). *Cyberspace: The human dimension*. New York: W. H. Freeman.
- Wiener, N. (1948). *Cybernetics: Control and communication in animal and machine*. Cambridge, MA: MIT Press.
- Wolfe, C. R. (2000). *Learning and teaching on the World Wide Web*. San Diego, CA: Academic Press.



# Decision Making

Helmut Jungermann

Technical University of Berlin, Berlin, Germany

1. Introduction
  2. Key Concepts
  3. Features and Distinctions
  4. Value/Utility
  5. Uncertainty/Probability
  6. Models and Modes of Decision Making
  7. Methods
  8. Areas of Application
  9. Conclusion
- Further Reading

## GLOSSARY

**biases** Systematic judgmental tendencies that violate the rules of probability.

**bounded rationality** Decision making within the limits of human cognitive capacities.

**choice** The selection of an option from a set of options.

**decision** The commitment to a course of action.

**heuristics** Mental rules-of-thumb used for the judgment of probabilities.

**image theory** A theory that includes the maximizing strategy as a special case but focuses on a strategy that works on the compatibility of options with the decision maker's goals and plans.

**prospect theory** The currently dominant theory of individual decision making, developed by Kahneman and Tversky, that keeps the core components of the classical rational model but takes the findings of empirical research into account.

**rationality** Decision making consistent with the axioms of the normative model that assumes the use of all available

information, stable preferences, and a maximizing strategy.

**uncertainty/probability** The uncertainty that people associate with the occurrence of the potential outcomes of their decisions.

**value/utility** The value that people attach to outcomes resulting from their decisions.

Decision making is a mental or behavioral commitment to a course of action. In a broader sense, the term "decision making" denotes an information-processing activity of a single decision maker, or of multiple decision makers, that begins with the recognition of a choice situation and ends with the implementation of the choice and the monitoring of its effects.

## 1. INTRODUCTION

More than many other behaviors, individual decision making is a subject not only of psychological research in the laboratory but also of interest in a number of "real-world" domains where significant decisions are actually made and where psychological knowledge has been applied successfully. The study of decision making has its roots in economics and statistics, and this is still the most active field for decision researchers. Topics include how people choose among consumption goods, investment options, and/or retirement plans as well as how managers make production and

marketing decisions and how they make organizational and employment decisions. Another important field involves health-related and medical decisions of both patients and physicians. For patients, topics include how people make their personal decisions about smoking, drinking, taking drugs, and/or sexual behaviors as well as how patients choose among treatment options (e.g., surgery or radiation) and how they can be helped in making such decisions. For physicians, topics include how they choose among diagnostic and treatment alternatives and how their individual or collective decision processes can be supported. A very different field is aviation, where the interest focuses on the decisions that pilots must make—often under extreme time pressure. Closely related are decision situations in high-risk facilities such as nuclear power and chemical plants. Knowledge about people's decision-making behavior has also been used for designing counseling and advice-giving procedures. Examples include when patients are asked for their informed consent, when parents come to a genetic counselor, when people seek investment advice, and when adolescents need behavioral strategies against sexual assault.

## 2. KEY CONCEPTS

A decision is a mental or behavioral commitment to a course of action. An option to change the status quo may be brought to the attention of a person and is accepted or rejected, or one of several given options is preferred over the others. A decision can be expressed by verbal judgment (e.g., "Option X is the best one") or by behavioral choice (e.g., Option X is chosen). In a narrow sense, the term "decision" denotes only the moment in which the commitment is made. In a broader sense, the term denotes an information-processing activity, usually beginning with the recognition of a choice situation and ending with the monitoring of the outcomes of the chosen option. Research has looked at all phases of a decision-making process—how people identify, generate, screen, and modify options; how they mentally simulate courses of action; how they match actions to situations; how they frame, compare, and evaluate all or a subset of options; and how they process information after a choice has been made. Decisions can require different amounts of cognitive effort and be made on different levels of awareness, depending largely on the significance of the problem, the familiarity of the situation,

and the experience of the decision maker. When people drive from their home to their office, or when surgeons perform appendectomies, their decision making is often nearly automatic and is experienced as performed routinely or intuitively. On the other hand, when people in a restaurant choose from a menu, they realize that they must make decisions and often ponder about the options and consider the various pros and cons. Typically, substantial cognitive effort is invested only if the problem is important, complex, or unfamiliar such as in the purchase of a house by a family, the choice of a cancer treatment by a patient, or the development of a new product by a company. Economists tend to consider a decision maker—the homo economicus—as a rational decision maker if he or she uses all available information, if preferences and beliefs are stable and consistent, and if a maximizing strategy is applied. Psychologists (and behavioral economists) favor the assumption that decision makers usually operate with bounded rationality (i.e., within the limits of their cognitive capacity) and that they apply the strategy that best matches their situational assessment and motivation.

## 3. FEATURES AND DISTINCTIONS

The dominant view is that people make their decisions as a result of anticipating and evaluating the potential outcomes and the uncertainties associated with them. This is called a consequentialist perspective. The major part of this article focuses on research that takes this perspective, but nonconsequentialist approaches are also presented later. From the consequentialist perspective, decision problems differ formally in the following ways. First, the consequences of the decision are certain (e.g., the consequences of choosing Apartment A or Apartment B) or uncertain (e.g., the consequences of choosing chemical therapy or surgery). Second, the consequences are unidimensional (e.g., the consequences of gambling in a casino as either a gain or a loss of money) or multidimensional (e.g., the consequences of choosing a job in a company or at a university). Third, the decision process may have one stage (e.g., the decision to hire a person) or several stages (e.g., the decision to develop a new drug). Fourth, the decision may be unique (e.g., decisions among cancer treatments) or may be made repeatedly (e.g., decisions between job applicants). Fifth, another important distinction must be made between individual decision situations and social

ones. In the former case, decisions are made by individuals, and individuals primarily must bear the consequences. This situation is the topic of cognitive psychology. In the latter case, decisions are also made by individuals, but the decisions are made in direct or indirect interaction among these individuals. This situation is the topic of social and organizational psychology.

#### 4. VALUE/UTILITY

People are assumed to make their decisions depending on their preferences for the potential consequences (i.e., states or events) of the given options. Implicitly or explicitly, people evaluate these consequences, and the subjective values, called utilities, determine their decisions. However, the utility of outcomes is not simply a linear function of their quantity. In general, one finds a logarithmic function well known from psychophysics. In economics, this is called the decreasing marginal utility function: The second million causes less pleasure than the first million. The same holds for the negative part of the function: A dessert for \$5 appears to be less expensive after a \$70 dinner than after a \$7 pizza. Advertising and marketing use the knowledge about the psychophysics of spending in many ways.

Evaluations are influenced by a number of factors, including the following prominent ones. First, people value an object more highly when they believe to have achieved it by their performance rather than by luck (i.e., source dependency). Second, people value an object more highly when they consider selling it than when they consider buying the same object (i.e., endowment effect); that is, a person would want a higher price for a painting he or she owns than the person would pay for the same painting. Third, the evaluation of an option is sometimes influenced by investment of money, time, or effort in the past (i.e., sunk cost effect); that is, a person does not leave a boring movie because he or she has paid for the admission ticket. Fourth, the evaluation also depends on the moment in time when the object will be received or when the state will occur (i.e., time preferences); that is, a person prefers to receive \$100 sooner rather than later, and the person probably prefers to see the dentist later rather than sooner (unless he or she has pain).

Multiattribute options (e.g., consumption goods, job applicants) require trade-offs between attributes. Consumer reports provide the relevant attributes of the goods and the values of the goods on these attributes. But consumers must make the trade-offs themselves.

Numerous studies have examined which attributes people use in such situations, how they weigh them, and how they integrate the information about the goods on all attributes. Sometimes, people use noncompensatory rules; that is, they eliminate all options that do not exceed certain cutoff points on one or several attributes or that do not share certain features, independent on their values on other attributes. For instance, a person looking for an apartment may first screen out those apartments that do not have a minimum size, then those whose rents exceed a certain limit, then those whose distance to the beach exceeds a certain distance, and so on until finally one acceptable apartment is left. Alternatively, people may use compensatory rules; that is, they accept that each value on an attribute, however low, can be compensated by high values on other attributes. For the person looking for an apartment, the small size of an apartment might not lead to its elimination because it can be compensated by a low rent. Consumer research has used the theoretical concepts, the research methods, and the empirical findings extensively.

#### 5. UNCERTAINTY/PROBABILITY

Decisions are always made under uncertainty because their consequences occur in the future and, thus, depend not only on the decision makers' choice but also on events and states not under the decision makers' control. In some situations, the consequences can be considered as nearly certain or the fundamental uncertainty is irrelevant for the choice. However, in most decision situations, the uncertainty cannot be ignored; therefore, research on how people deal with uncertainty is an important field of decision research.

For mathematicians, probability is the language of uncertainty; for decision researchers, subjective probability is the language of uncertainty. The subjective probability for an event may be based on the known or assumed relative frequency of that event (e.g., the subjective probability of rain tomorrow, the subjective probability of having a car accident), but in many situations the event is unique and there are no relative frequencies (e.g., the subjective probability of Brazil winning the soccer world championship, the subjective probability of an increase of the gold price by the end of the year). Research has studied the psychological variants of uncertainty, their cognitive characteristics and determinants, the ways in which people express uncertainty (e.g., numerically, verbally), and the ways

in which people understand expressions of uncertainty (e.g., probabilities, percentages, frequencies). Intuitive judgments of probabilities have been shown, under certain conditions, to deviate considerably from our statistical knowledge and to violate the rules of probability theory systematically. These observations led Tversky and Kahneman to develop their heuristics and biases research program. The key assumption is that people intuitively apply mental heuristics rather than statistical algorithms, and although these heuristics are generally useful and efficient, they can produce judgmental errors, fallacies, or biases. Two heuristics are particularly important: representativeness and availability. Representativeness is when events are unique (e.g., the guilt of a defendant) or set in the future (e.g., the outcome of the next election) and people draw on their knowledge about the case and judge the probability by the closeness of the match between the event and the prediction derived from their knowledge. Availability is when events are grouped in categories or by features (e.g., percentages of road accidents in a year) and people judge the probability by the number of examples that come to mind or by the ease with which they come to mind. Many judgmental biases can be explained by the assumption that people use these heuristics, for instance, the base rate fallacy, sample size neglect, regression neglect, the unpacking effect, and the conjunction fallacy. Other prominent biases in subjective probability judgments, not directly related to the use of heuristics, are the hindsight effect, the overconfidence effect, conversion errors, the illusion of control, and unrealistic optimism.

## 6. MODELS AND MODES OF DECISION MAKING

Prescriptive theories of decision making say how decisions should be made if the decision maker wants to follow certain principles of rationality (e.g., the principle of transitivity). These theories are part of management science. They provide formal rules and procedures for structuring and processing the relevant information and, thus, provide support in complex decision situations. Descriptive theories aim to show how people actually make decisions. Because the cognitive capacity is limited, or at least is not always used efficiently, actual decisions are often suboptimal compared with the decisions prescribed by formal theories.

The first descriptive model of decision making under uncertainty was proposed by Edwards in 1954. The

subjective equivalent utility (SEU) model assumes that people try to maximize their subjectively expected utility. The SEU of an option is the sum of the utilities of its consequences, weighted by the subjective probabilities of their occurrence. The decision maker is assumed to choose the option with the highest SEU value. In 1979, Kahneman and Tversky proposed prospect theory (PT) as an updated version of the SEU model. The theory sticks to the assumption that decisions are determined by the values and probabilities of their consequences but takes into account the many observations of decision-making behaviors that do not concur with the SEU model. PT has become very influential, particularly in the economic and medical domain. In 2002, Kahneman was awarded the Nobel Prize in economics, primarily for this theory (Tversky had died in 1996). The essential elements of PT are the value function and the decision weight function. With respect to the value factor, PT assumes that people mentally code the potential outcomes of options in relation to a reference point, that is, the status quo or the aspiration level. Outcomes above the reference point are coded as gains and outcomes below the reference point are coded as losses. Gains and losses are evaluated according to a value function that has two properties. First, the function is concave over gains and convex over losses; that is, additional gains please less, and additional losses hurt less. Second, the function is steeper over losses than over gains; that is, a loss of \$100 hurts more than a gain of \$100 pleases. With respect to the probability factor, PT assumes that subjective probabilities of outcomes are transformed into weights that represent the significance of the occurrence of the outcome. For instance, many people weigh the transition from a 0.99 probability to certainty (1.00) as heavier than a transition from a 0.41 probability to a 0.42 probability. PT provides explanations for a number of empirical phenomena that often are considered as irrationalities or anomalies, particularly in economics. Examples include the finding that preferences are not invariant with respect to their verbal description (i.e., framing effect) and the finding that investors sell winner stocks too fast and hold loser stocks too long (i.e., disposition effect).

An alternative theory, image theory, was proposed in 1990 by Beach. Image theory has become very popular in management and business schools. Beach assumed that people apply an optimizing strategy like the one proposed in PT, only under very special conditions. More often, people examine whether a new option is compatible with their goals and plans (called images)

and accept and implement the option if that is the case. If the option is not compatible and violates important features, other options are searched and explored. When people are confronted with a number of options, they first screen out the ones that are incompatible with goals and plans, and only if more than one option remains do they try to identify the optimal one.

Other researchers have also pointed out that in real life, optimizing decision making is the exception rather than the rule. Features of the problem and the situation are often more important than the potential consequences of a decision. For instance, the behavior of managers in organizations is strongly determined by rules. The decisions of consumers are often determined by affects. In ethical conflicts, decisions are primarily determined by basic values, such as honesty and “do not harm,” irrespective of the consequences.

Still another approach focuses on what is called naturalistic decision making, denoting the process by which people use their experience to make decisions in complex and dynamic environments, often under time pressure and involving high risks (e.g., pilot decisions on the flight deck). For such situations, in 1993, Klein proposed a recognition primed decision model that involves recognition of cue patterns that leads to retrieval of a response option. Serial evaluation of single options is considered typical, and the first option that matches the decision maker’s goals and the situational constraints is chosen.

Models of organizational decision making describe decision making by both single and multiple actors in an organizational context. The normative rational (classical) model assumes that the organization follows a computational value-maximizing decision-making strategy. The model’s usefulness is limited to a small set of situations with specific characteristics (e.g., goals can be described in quantitative terms); consequently, alternative models were developed based more on the actual behavior of decision makers in organizations. For instance, the (behavioral) organizational model focuses on the limited information-processing capacity of decision makers and postulates an outcome-satisfying strategy. In addition, the garbage can model conceives of organizations as “organized anarchies” (e.g., having inconsistent goals). Universities are considered the prototype of such organizations that do not follow one specific decision-making strategy. Outcomes result from the variable participation of their various members and groups in continually changing tasks. All of these models differ in their capacity to cope with different degrees of uncertainty and conflict among interests.

## 7. METHODS

Decision researchers use a variety of quantitative and qualitative methods to examine how people make decisions. Experimental studies in the lab usually ask participants to make judgments about values and importance as well as about uncertainties and risks, to make choices among given options, and to evaluate options, search for information about options, or distribute goods. Reaction times, eye movements, and other behavioral data (including verbal reports) are also used as measures of the information-processing activity of participants. Outside of the lab, the questionnaire is the dominant method, where either decision problems are presented by the researcher, and participants provide the required responses (e.g., students in a lecture) or decision situations arise and the researcher observes and records participants’ behavior (e.g., pilots in a cockpit).

## 8. AREAS OF APPLICATION

The examples provided in this article for illustrative purposes have indicated some of the areas where decision making has been studied and where concepts and findings of decision research have been applied. More systematically, one may distinguish the following major domains: (a) behavioral economics and finance (e.g., investment and savings decisions); (b) business, administration, and management (e.g., employment, organizational, and product decisions; strategic planning with scenarios; communication of decisions; distributed decision making; cultural differences in decision making); (c) marketing and consumer behavior (e.g., product advertising, labeling, and pricing decisions; consumers’ allocation of their scarce means over various commodities and services; consumers’ decisions to acquire, consume, and dispose of products, services, and time); (d) health and medicine (e.g., addictive behaviors such as smoking, drugs, and alcohol; cancer screening campaigns; physicians’ use of probabilistic information; diagnostic and treatment decisions; informed consent; shared decision making); (e) aviation, military command, and firefighting (e.g., pilot crew decisions and decision-aiding systems; battle commander decisions; general aviation safety decisions); (f) high-risk facilities (e.g., operator decisions in nuclear power and chemical plants; emergency and general safety-related organizational decisions); (g) environmental issues (e.g., individuals’ benign attitudes and destructive behaviors;

governmental decisions and their implementation; communication between the public and the experts); (h) ethical and justice issues (e.g., organ transplantation decisions; business ethics; conflicts between self-interest and the other; dilemmas in risk communication); (i) advice, consulting, and counseling (e.g., decision advice in medical and financial problems; decision consulting for companies and agencies; genetic or abortion counseling for potential parents).

## 9. CONCLUSION

Approximately 50 years ago, the study of decision making started with lab experiments examining how people make choices among monetary gambles. Since then, the scope of studies has been extended immensely, reflecting the increasing importance of understanding and improving decision making in our society as well as the progress of methodology and model construction within the research community. However, the extension of the field has also eroded the idea of a general theory of decision making in favor of situation- and domain-specific models and empirical research on the conditions under which the various modes and strategies are being applied.

### See Also the Following Articles

Leadership and Culture ■ Power, Authority, and Leadership

### Further Reading

Beach, L. R. (1990). *Image theory: Decision making in personal and organizational contexts*. Chichester, UK: Wiley.

- Connolly, T., Arkes, H. R., & Hammond, K. R. (Eds.) (2000). *Judgment and decision making: An interdisciplinary reader*. Cambridge, UK: Cambridge University Press.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51, 380–417.
- Flin, R., Salas, E., Strub, M., & Martin, L. (1997). *Decision making under stress: Emerging themes and applications*. Brookfield, VT: Ashgate.
- Hastie, R., & Dawes, R. (2001). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Thousand Oaks, CA: Sage.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, MA: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kahneman, D., & Tversky, A. (Eds.) (2000). *Choices, values, and frames*. New York: Cambridge University Press.
- Koopman, P. L., Broekhuijsen, J. W., & Wierdsma, A. F. M. (1998). Complex decision-making in organizations. In P. J. D. Drenth, H. Thierry, & C. J. de Wolff (Eds.), *Handbook of work and organizational psychology*, Vol. 4: *Organizational psychology* (2nd ed., pp. 357–386). Hove, UK: Psychology Press.
- March, J. G. (1994). *A primer on decision making: How decisions happen*. New York: Free Press.
- Messick, D. M., & Tenbrunsel, A. E. (Eds.) (1996). *Codes of conduct: Behavioral research into business ethics*. New York: Russell Sage.
- Rasmussen, J. (1993). Deciding and doing: Decision making in natural context. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 158–171). Norwood, NJ: Ablex.
- Shapira, Z. (Ed.). (1997). *Organizational decision making*. New York: Cambridge University Press.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.



# Decision Making in Sport

Gershon Tenenbaum

Florida State University, Tallahassee, Florida, USA

1. Introduction
  2. Nonsport Perspectives on DM
  3. DM in Sport: The Dynamics and Mechanisms
  4. DM in Teams
  5. Affective States and DM
  6. Summary: A Holistic View of DM in Sport
- Further Reading

## GLOSSARY

**anticipation** A mental operation that allows the athlete to predict upcoming events based on the elaboration between the current available visual-spatial information and knowledge stored in long-term memory.

**coping strategies and self-regulation** Active or passive emotion-focused or problem-focused methods used by athletes to appraise the environmental pressure that enable appropriate processing of information and decision making.

**decision making** A process by which an individual or a group of athletes select one course of action or actions from several alternative actions to be performed during precompetition or during a competition. It enables the athlete or the group to function under physically, emotionally, and mentally stressful conditions.

**decision-making self-efficacy** The belief that an athlete holds in his or her ability to make the right decision at the right time under competitive pressure.

**perceived stress** The perception of cognitive and somatic stimuli in the competitive environment that evoke feelings of anxiety and other emotions. It is linked to the degree with which one can adjust to the conditions to allow efficient decision-making processing.

**visual-spatial attention** The location in space in which the visual attention is focused, which enables a selection and elimination of environmental cues essential for decision making.

Jake plays in a central-midfield role on his soccer team. He plays at the very heart of the field and all moves, both offensive and defensive, bypass Jake. He has a multitude of decisions to make. Jake is young and prone to overarousal and constant errors. He pays equal attention to all the features of the soccer field and is limited in his capacity to process the most relevant information; thus, he often makes the wrong decisions. Coaches, parents, and teammates are upset with Jake; he is physically talented, but under time and event pressure he consistently chooses the wrong course of action. This article addresses the main issues related to Jack's limited cognitive capacity, which prevents him from making the right moves at the right time. The article addresses the required cognitive skills for proficient decision making and action execution. Several examples from different sports are provided to illustrate the research and theoretical issues.

## 1. INTRODUCTION

Decision making (DM) is a process by which an individual, a group, or an organization selects one preferred action from among two or more possible

actions within a specific situation. This article focuses mainly, but not conclusively, on the procedures involved in an individual athlete's DM processes. In sport, the DM process depends largely on the environmental and temporal conditions and rules under which a decision maker operates. These conditions vary with the nature of the sport. For example, in open-type sports such as basketball and soccer the environment is dynamic and consistently changing. The DM process in such sports requires the athlete to make a response selection (RS) based on past, present, and anticipated actions that may change through the course of time. Under such variable environmental conditions, the decision pertains to (i) which type of response/action to choose and (ii) at which time to execute the selected motor response. In contrast, closed skills, such as shooting and archery, which are performed within predetermined environmental constraints, require the decision maker to attend to his or her own proprioceptive sensory signals and decide when in the course of time to implement the decision.

In all sport environments, the DM process consists of three sequential phases: a preparation phase (i.e., visual search, selective attention, and anticipation); DM and RS through information processing (i.e., working long-term memory elaboration); and an action evaluation phase, in which DM and RS may be altered if new information is available at this stage. These stages are illustrated in Fig. 1. This information-processing model assumes that DM occurs under cognitive control. In cases in which the sport environment triggers a response under severe time constraints (i.e., less than 150 ms, which is the required reaction time under cognitive control), DM may occur under

self-organized rules. In such circumstances, decisions may come "naturally, without thinking." Generally, this self-organized DM process is more likely to occur among highly skilled and experienced athletes.

RS is typical to all sports, although the number of internal and environmental stimuli varies. The RS is more challenging when the number of stimuli increases, the temporal conditions shorten, and the athlete's affective state (i.e., anxiety) becomes less tolerable. The prevailing viewpoint in DM research in sport is that the process occurs within the context of both cognitive and affective systems, which operate simultaneously. The multifaceted perspective of DM in sport situations is presented in this article.

## 2. NONSPORT PERSPECTIVES ON DM

The traditional approach to DM was developed in the economics/statistics domain. The prescriptive approach viewed the person as an "outcome" seeker who pursues a desired goal. Accordingly, decision makers attempt to maximize a certain "expected value" and they process information in a manner that is targeted toward this ultimate end. The process is governed by comparing one's alternatives to an ideal solution that in many cases is unknown and speculative. One application of this approach accounted for gambling behaviors, where the unknown product necessitated a probabilistic estimation of utility. Despite their appeal, the prescriptive theoretical normative models were largely unsuccessful because they had limited ecological validity. Since human beings often make decisions that incorporate more considerations than just the expected value of their choice, these components needed to be identified and taken into account.

Cognitive-oriented researchers have examined the capabilities necessary for DM under certain environmental constraints. Also, cognitive capacity was believed to be a major constraint. Studies by DeGroot, Simon, Chase, and others in the early 1970s demonstrated that repeated exposure, experience, and skill level are associated with task-specific capabilities that enable the decision maker to use strategies and integrate information beyond the previously believed cognitive constraints. For example, expert chess players were able to recognize patterns in situations specific to chess in order to overcome limitations in working

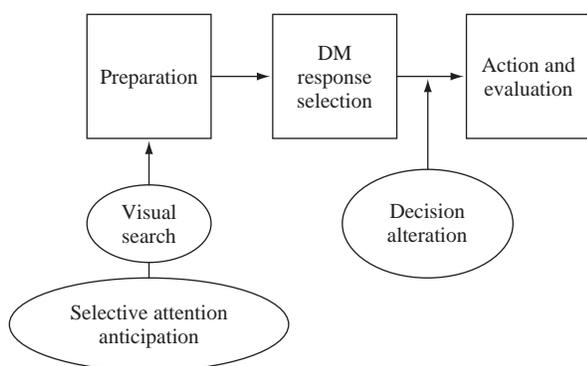


FIGURE 1 Stages of information processing and decision making in sport.

memory, although their overall cognitive capacities were not superior to those of nonexperts.

The naturalistic/descriptive approaches to DM assume that the DM process involves both rational and irrational processes and incorporates personal values, morals, motivation, and emotional level in the DM process. From a naturalistic standpoint, DM consists of defining the problem, seeking alternative solutions (some of which may not be available at a given moment), and then choosing an alternative. According to these approaches, the DM process can be compromised under various conditions, such as those imposing psychological pressure. Several theoretical models, such as image theory, explanation-based theory, recognition-primed DM, and cue retrieval of action, were introduced to account for the human DM process.

Contemporary models within the naturalistic/descriptive approach concentrate on how people make decisions rather than on how they should make decisions. Mechanisms in the form of heuristics are offered to account for the quality of decisions as well as the processes by which decisions are made. These models, which consider the real-world conditions under which decisions are made, have largely replaced the more mechanistic prescriptive models.

The recognition-primed model (RPD) developed by Klein assumed that past experience in similar situations governs the DM process. Klein suggested that DM consists of cue identification, situational goals, alternative action generations, and expectations for possible alterations, all of which can be enhanced with increased experience. The complexity of the situation generates more adjustments or alterations, when time permits. Mental representations, which are established through experience and repetitions, guide the DM process. RPD was found to be valid in chess playing, and this approach is closely related to current concepts and models in sport-related DM processes.

The naturalistic method is a knowledge-driven discourse (i.e., domain-specific declarative and procedural information processes) that according to Lipshitz has not been given its deserved attention. Recent studies emphasize the role that domain- and content-specific features play in utilizing decision strategies under varying environmental conditions. In sport, however, knowledge base in the form of expertise and deliberate practice is well documented, thus advancing the naturalistic conceptualization of DM in simple and complex conditions.

### 3. DM IN SPORT: THE DYNAMICS AND MECHANISMS

DM in sport constitutes a set of adaptive behaviors that enable an athlete to operate effectively within the competitive environment under conditions that vary with regard to emotional, mental, and temporal demands. The efficiency of this capacity is influenced by the richness and the variety of perceptions, as well as one's ability to process information, at a given time. Sport-related tasks are unique in nature and, therefore, encoding relevant information for further processing requires specific perceptual and attentional resources and mechanisms. Depending on the situation at any given moment, a shift between serial and parallel processing of environmental stimuli is required to allow an efficient RS process. Under unconstrained time conditions, information is fed forward for further processing, responses in the form of neural codes are selected and maintained "on alert," and a preferred action is selected at a given time while others remain activated to some degree until decayed. For example, a golfer may consider the weather and the terrain when selecting a particular club while awaiting his next shot. In contrast, under restricted time constraints the DM process operates under direct perception-action rules, which depend on knowledge structure and schemas stored in long-term memory (LTM). A "fast break" in basketball is a common example of this type of condition.

DM in sport can be viewed as a sequence of decisional processes rather than one RS. This sequence starts at an early stage when the athlete observes his or her opponents on a video film or uses his or her visualization skills. While watching the rival athlete's or athletes' style, weaknesses, strengths, and so on, the athlete plans potential actions, which are stored in memory in the form of representations that can be retrieved easily when required. The more the player is aware of the context, the more he or she can anticipate and alter his or her opponent's action even before stepping into the playing court. While competing, however, the first decision is related to the selection of relevant (and elimination of irrelevant) cues in the environment through the use of different visual strategies. A sequence of events in the performing environment is similar to a "mental chronometry" introduced by Posner, Nissen, and Odgen in which warning signals draw attention until an imperative signal triggers a response. During the preparation period, the athlete attends to the visual field and anticipates upcoming

events through the elaboration between working and long-term memory systems. The decisions at the anticipatory stage pertain to what actions the opponent intends to perform and with what probability (e.g., a tennis player anticipates the location of her opponent's serve). The next decision in the sequence regards what response to select from several alternative responses and when to execute it (where and when to move on the court). This decision has both quality and timing components. Last in the sequence is the decision about response alteration when conditions require or impose a change. This decision relies on a quick modification process, which controls the system's adaptation level to the changing and dynamic environment (how the player might recover if the serve goes in an unanticipated direction). The success rate of DM depends to a large extent on different mechanisms operating simultaneously and in mutual interaction. To clarify these processes, the main mechanisms affecting DM in sport are introduced next.

### 3.1. Visual–Spatial Attention

An athlete uses visual–spatial attention to select and discriminate among cues in the playing environment. In many sport settings, a vast array of stimuli, some relevant and some irrelevant, are available to the athlete, and a major aim here is to minimize environmental uncertainty and to enable smooth detection, recognition, recall, and selection of relevant stimuli for higher level processing necessary for RS. The visual system has two main characteristics: fixation, the target toward which visual attention is directed, and duration, the time lag between two shifts in fixation. Scientific evidence indicates that among athletes who play fast ball games, with increases in experience and skill level, the number of eye fixations decreases, whereas fixation duration increases. In other words, more experienced players tend to have fewer shifts in visual–spatial attention.

Visual scanning operates via two possible strategies: the target control strategy, which consists of scanning individual targets in a sequence until one is located that is compatible with a representation stored in LTM, and the context control strategy, which is initiated by memory representations not sensitive to individual objects but rather to a pattern in the visual display. Scanning the environmental stimuli under context control is of greater advantage in open and dynamic settings because it reduces the information-processing load, increases visual attention efficiency,

and simplifies the elaboration between working and long-term memory.

Novice players for whom the environment is not fully familiar adopt a target control strategy, whereas with increased experience and skill level a context control strategy becomes more prominent. With the shift from a target to context visual control strategy comes a reduction in eye fixations and an increase in the duration of fixations. In fast ball games, eye fixation location depends on the visual scan sequence. One first fixates on the large features in the display, such as the entire court in racket sports, then moves to the racket (tennis, badminton, and squash) during the later sequence before a RS is made. Such a synthetic strategy enables the player to integrate information from the core sources of reference at once instead of relying on serially analytic chronological order scan, which is time-consuming and inefficient.

Cueing and priming are also key features in the sport realm that are integral in visual attention. When an athlete anticipates that a move will occur with high certainty, his or her response is faster as a consequence of this “priming a response.” When the environment causes greater uncertainty, response time is slower. While visually scanning the stimuli display, late cuing experiments in sport have shown faster reaction times to cue-primed locations than to unusual and unanticipated ones. However, sport-specific studies have shown that the cueing effect depends on the time interval between the warning and imperative stimuli. Reaction times increased with the time delay between the warning and imperative stimuli. However, in dynamic and fast-changing environments, perceptual-motor integration works differently from that reported for naive subjects in laboratory situations. As Cave and Bichot claim, the visual system is powerful in detecting and selecting objects in parallel to guide a response, and an object's location in the display is another criteria for DM. However, once extensively practiced, the probabilities assigned to upcoming events become the knowledge base of the athletes in the form of neural codes, from which they anticipate that events will occur. They thus become better equipped to prepare an appropriate action regardless of the time and space constraints of the environment. An example of this process is a batter in baseball who, through experience in detecting specific cues, learns which particular pitch (e.g., curve ball or fast ball) is most likely to be thrown in a certain situation.

An extensive knowledge base enables the athlete to minimize the costs associated with shifting attention

from narrow to large areas. When distracters in the space are cued, they interfere with the identification of a designated target and lead to an increase in reaction time. It is argued that when skilled athletes are exposed to several sources of information at one time, each carries with it a “probability value” for upcoming moves. When an anticipated action carries with it a high probability value, it enables the athlete to process the information efficiently and effortlessly and to subsequently make an adequate decision at the time. In contrast, an athlete who lacks such a capability pays the costs in the form of increased time for processing as well as less appropriate responses.

Finally, despite early beliefs that attention cannot be allocated or split among various sources simultaneously, recent research showed that attention could be divided when two targets were discriminated from each other by a given property or when all stimuli were suddenly presented. In open and dynamic environments such as ball games, anticipatory processes simplify the selection process by allocating visual attention to the more expected location and, at the same time, by using a context control strategy, which leaves other attention channels on alert for possible unexpected imperative stimuli to arrive. In such a manner, the skilled player secures a response with higher probability for success and at the same time leaves room for modification and refinement if circumstances require. These mechanisms help to explain how, for example, an elite basketball player can prepare to shoot a layup and, at the last moment, notice a defender’s position, alter his decision, and pass the ball to a teammate.

### **3.2. Priming, Attentional Flexibility, and Dimensions of Attention**

In the typical sport environment, cues are assigned a probability value in terms of response priming. That is, based on the visual information received, one response is given a higher probability of selection than the others. Responses are primed when their associated stimuli are detected. The primed response provides a “benefit” in terms of the reduced time required to process and select it, whereas the “cost” is that it then takes longer to process and prepare an alternative response for unexpected stimuli. The cost–benefit ratio determines the flexibility of the attention process.

Research using the expert–novice paradigm in sport has shown that the higher the skill level of the athlete, the more one is able to optimize this effect (i.e., to

decrease the cost and increase the benefit of the stimuli-associated information processing). Precueing has a positive effect on the RS process because priming increases an athlete’s awareness of upcoming events as well as response readiness. Scientific evidence also suggests that athletes can use automatic and voluntary signal detection modes and can shift between the two when required. With increasing expertise, the athlete acquires both a “smoother” shift and more pronounced flexibility process. In sport, the type and intensity of priming are essential for RS. In competitive situations, athletes choose the priming options that have the highest utility value; that is, they select those that ensure the best chances of triggering the appropriate responses at the proper times.

Because different sports require different attentional processes, no single strategy is ideal for all situations. Some sports (e.g., archery and golf) consist largely of stationary positions, requiring the athlete to focus attention on both internal and external signals. Sports that are more open and dynamic (e.g., basketball and football) require a wider and more flexible attentional style. Moreover, some sports include conditions that require a wide, external attention (e.g., a playmaker in basketball) but others that require a narrow internal attention focus (e.g., shooting a free throw). In these sports, attentional flexibility, which is the ability to alter one’s focus, is a major asset. Neidefer’s model of attentional style accounts for the requirements and shifts of attention needed in various sport contexts. Accordingly, he recognizes two dimensions of attention: (i) width, from wide to narrow, and (ii) direction, from internal to external. Athletes must acquire the skill of adopting the appropriate attentional width and direction that enables them to make effective decisions. Failure to shift attention increases the probability of missing the relevant information needed for optimal DM.

### **3.3. Anticipatory Mechanisms**

The attentional system, which detects the most salient environmental cues, enables the athlete to anticipate events in advance. Repeated responses to game maneuvers enhance the athlete’s perceptual anticipatory capability, which allows for a quick elaboration with the stored program representations in LTM. Fast access to one’s knowledge base becomes possible with only partial information. Because practice in the form of task-specific skill develops ready-access scenarios, these preexisting schemes are available for access even when a limited amount of the information contained in

the game situations “yet to be” is captured by the visual system. This anticipatory capability becomes automated and effortless with the development of expertise. For example, an elite volleyball player may notice the movement of just one of six players on the opposing team. However, since she has a complex set of stored schema developed through experience, she is able to predict or anticipate the movements of the remaining players.

Scientific studies of fast ball games have demonstrated the advantage of expert athletes over their novice counterparts in anticipatory decisions. Advanced anticipatory capability reduced the time required for a response, particularly when temporal conditions were extremely short. Anticipatory capability stems from both knowledge structure and advanced visual attention strategies. During the initial stages of scanning the environment for relevant stimuli, anticipatory decisions alter depending on the events taking place at the time. In expert athletes, visual attention is directed to pattern recognition through enduring eye fixations, and fast eye movements occur only under certain environmental constraints. At this stage, attention is simultaneously diverted to several sources of information, including the opponent’s moves, teammates’ locations, and time left on the play clock. Lack of experience and target types of visual scan result in impaired anticipatory capability. At the final stages of the anticipatory process, the novice athlete is engaged in decisions about what and where moves and actions will be performed, whereas the expert makes decisions about what action to perform in response to the opponent’s moves and/or situational demands. A novice wrestler, for example, decides where his opponent might step, whereas the expert selects the move he will use to score a takedown because he “knows” where the opponent’s leg will be. Finally, RS appropriateness depends not only on the cognitive process but also on the proficiency of the motor system and the confidence one has in his or her ability to carry out the response. In the example cited previously, the wrestler must possess sufficient strength, quickness, and confidence for the decision to be effective.

### **3.4. Memory Representations, Information Delivery, and Knowledge Structure**

When an athlete’s knowledge base is rich, environmental distracters have limited impact on the DM process. A limited knowledge base considers these distracters as

relevant information, resulting in slowed DM via inhibition of the athlete’s information processing. The seminal studies of DeGroot and Chase and Simon on chess players in the 1960s and 1970s, and follow-up studies on a variety of sports since then, demonstrated how information is encoded, processed, and retrieved as a function of skill level and duration of visual exposure. It was assumed in these studies that recall accuracy following a short exposure to structured and unstructured situations would be a reliable indicator of both the visual attention strategy and the knowledge structure of the participants. Differences were consistent in that more skilled athletes could more accurately recall structured/logical/meaningful positions presented in a relatively short time than could unskilled athletes. The “chunking hypothesis” emerged as an explanation for this advantage in memory representation. This hypothesis rests on the premise that experience in the form of knowledge base and structure enables the player both to capture larger chunks of information in the form of logical patterns and to use these patterns as retrieval paths in a later stage. The chunking hypothesis was used to account for many research results, although recent scientific evidence from dancers and musicians has demonstrated significant memory superiority of skilled performers not only in meaningful information structures but also for memorizing random pieces of information. Thus, the chunking hypothesis has been called into question. In 1995, Ericsson and Kintsch proposed the existence of long-term working memory (LTWM), which may account for unusual memory capability in experts that is task-specific and cannot be generalized to other domains.

Short-term memory appears to have limited influence on experts’ memory capabilities, and working memory was considered a constraining factor because only small amounts of information could be temporarily stored while other information was processed. The concept of LTWM overcomes the notion that memory storage capacity is limited and decays immediately after stimulus onset. The constraint of a limited capacity working memory system seems to be irrelevant in expert memory performance, wherein concurrent activity seems to present limited interference. The new LTWM conceptualizes that when one becomes an expert in a given domain, such as sport, he or she can utilize LTM as a means for expanding short-term working memory.

LTWM develops through the creation of a domain-specific retrieval structure that is used to enhance storage and to maintain information in a more accessible

and less interference-prone state. A specific knowledge base is developed through practice, which enables task-specific memory traces to be rapidly and efficiently retrieved. The advantage of LTWM is that it relies on fast encoding and storing processes in a ready retrieved form from LTM. Domain-specific stored information in LTM is used as a retrieval cue that allows one to “unpack” the requested retrieval structure. Although decisions in sport are required, these same memory traces are used to retrieve a response and keep another response alternative on alert for possible and immediate alteration, refinement, or modification. When the stored information is richer and readily accessible, the DM process is smoother and of higher quality.

For a RS process to be efficient, task-specific declarative and procedural knowledge should be structured in a form that allows for easy and fast access. French and McPherson explored the network of conceptual knowledge required to make and execute tactical decisions in the dynamic and changing environments of ball games. The knowledge base consists of memory representations for action profiles, game situation prototypes, competitive scripts, and sport-specific strategies. The knowledge base is structured hierarchically and consists of macro- and microlevel routes. The microlevel is managed by the working memory and enables the athlete to attend to ongoing events in the sport environment. Advanced environmental visual scanning and anticipatory capabilities enhance the flow of information for further processing. The macro knowledge level refers to action plan profiles and specific event profiles, which enable the athlete to control and regulate the sensory systems and efficiently select and execute a response.

### 3.5. Decision Alteration

Several studies incorporating overt (e.g., reaction time, choice reaction time, and accuracy) and covert behaviors (e.g., evoked potentials) simultaneously have indicated that correct and incorrect decisions can be detected in brain activity and described in neural codes. Although not examined in actual competitive situations, these results can be extrapolated into the sport realm. It is suggested that with practice and experience, the athlete incorporates a decision strategy so that he or she can successfully replace one response by another response when the conditions necessitate such an endeavor (i.e., decision alteration). The psychophysiological research clearly indicates that beyond a given neural activation threshold, response alteration is almost impossible. Nevertheless, to reduce the probability of

error, the expert athlete holds alternative decisions “on alert” and activates them when an alternative decision is required. Thus, expert athletes, unlike their nonelite counterparts, activate alternative pathways that allow an “effortless” alteration process, which occurs mainly in complex and time-limited conditions.

An explication of decision alteration among experts was presented in Shallice’s “selection process” theory. According to this explanation, two attentional mechanisms control DM and action: contention scheduling (CS) and supervisory attentional system (SAS). The CS is automatic in nature and selects responses for action. The multiple schematic responses are activated and are in mutually inhibitory competition for selection. The final selection of a response is made when one of the alternative decisions is activated beyond a given threshold. The SAS mechanism can access the neural representations in the form of an intention for execution. When the selected response by the CS is proved to be inadequate during the time sequence, SAS modifies, stops, or, if not too late, alters the RS. Research shows conclusively that despite the advanced anticipatory capabilities of the expert athlete, RS is taking place only in the last stage of the process when uncertainty is minimized so that decision alteration can be made without tremendous processing costs. When an alternative response is ready and required, the alteration takes place.

## 4. DM IN TEAMS

One key advantage that a team has over an individual is a greater supply of human resources available for DM. For example, teams have an advantage over individuals on some measures of information processing (e.g., storage) owing to the extra information processing capabilities provided by multiple team members. Accordingly, in teams, the cognitive labor associated with DM tasks can be distributed over team members. Thus, the team can be considered as a cognitive system in itself. However, the cognitive properties of a team are more than just the sum of the cognitive properties of the team’s constituent members; information processing is affected by the way the team communicates and is coordinated and organized. Consequently, DM efficacy increases to the extent that a team invests effort into establishing strategies for efficient communication, coordination, and organization.

Evidence supports the notion that the more a DM task requires intramember information exchange between team members, the more important are effective

communication strategies for performance. However, studies have revealed a tendency for teams that are under time pressure or that are newly formed to discuss only information shared by all team members and not information unique to individual team members. Teams may be subject to DM biases under these conditions. Similarly, studies on crew communication during ship navigation and subsequent use of computer simulations of communication have shown how confirmation bias, a phenomena known to affect DM in individuals, can be accentuated in teams. In 1991, Hutchins showed that even while holding the cognitive properties of individuals constant, groups may display quite different cognitive properties depending on how communication is organized within the group and over time. In summary, DM benefits from the extra resources a team can provide only to the extent that teams factors such as coordination, organization, and communication are taken into account.

## 5. AFFECTIVE STATES AND DM

DM in practice or competitive situations often occurs in the presence of psychological pressure and under conditions that elicit physiological arousal. Therefore, the cognitive process that leads to RS depends on the (i) perceived stress experienced at the time, (ii) coping strategies and self-regulations used to cope with the emotional and physical stress, and (iii) confidence the athlete has in successfully anticipating and selecting a response.

### 5.1. Perceived Stress and DM

DM is strongly affected by the perceptions of stress (i.e., anxiety) the athlete experiences during competition. Both somatic and cognitive anxieties affect the attention processes, which are vital for the DM to be effective. Under an excessively low-anxiety state, attention is directed internally, which may result in the athlete ignoring or avoiding crucial cues that, under an ideal state, would trigger a response. Under a high-anxiety state, attention narrows to a level at which important environmental cues are missed. The detrimental effects of an elevated anxiety level are most prominent in complex, open, and dynamic situations in which a plethora of environmental stimuli are required for effective DM. In closed and explosive-type sports, such as Olympic weight lifting, high levels

of perceived anxiety are less harmful and, indeed, can even be beneficial.

In line with major theories of cue utilization and attention narrowing under conditions of high perceived anxiety, various studies have shown that DM under low anxiety relies on processing relevant and irrelevant cues alike. Under such conditions, the most important cue is not always primed to trigger a response. Thus, selectivity under such a condition is unfocused and unlimited. Under high-arousal conditions, the opposite is true. That is, an athlete's attention is narrowed, and DM occurs with too few cues. Under moderate levels of anxiety, attention processes are optimal in that only the most relevant cues are primed for processing and DM. Highly anxious athletes more frequently conduct "false alarm" DM errors, (i.e., make unnecessary decisions), whereas athletes with low arousal levels conduct both false alarms and correct RS. Moderate anxiety levels also limit the problems that many athletes have concerning sustained alertness (i.e., vigilance). Optimal anxiety levels are associated with "alertness preservation" (i.e., high concentration) over a given period of time. Thus, when mental alertness is optimal, it enables one to process the environmental stimuli so that no major stimulus is missed.

A concern arises when "choking" under pressure is evident even in the expert athlete. Two theories account for DM and performance decrements under psychological pressure conditions. Distraction theory claims that once under pressure, shifting attention to task-irrelevant cues results in a move from automatic to conscious attention control, resulting in poor DM. Alternatively, self-focus theory or explicit monitoring theory proposes that pressure causes the performer to pay attention to and control the automaticity of the skill. When a skill is learned and is best executed under automatic control, a shift to intentional control negatively affects DM and motor response processing, resulting in choking. In short, the athlete "thinks too much." Increased arousal levels result in a narrowing and self-focused attention that shifts away from the relevant cues and ultimately affects the DM process. Furthermore, choking seems to be most evident in tasks that require procedural knowledge and less so in tasks that primarily consists of an explicit knowledge base.

A recent development with regard to the emotions-performance linkage indicates that RS depends on a larger range of emotions than merely cognitive and somatic anxiety. It is the individual's perceptions of the intensity, pleasantness, and functionality of emotions that determine the specific resources recruited,

which generate energy required for optimal DM. When such an emotional state is not experienced, DM, and subsequently performance, may decline. Thus, each individual has a unique emotional-related performance zone within which the probability of making sound decisions and implementing them is highest. In other words, optimal, moderate, and distracted decisions are more probable under different emotion zones for each athlete. These performance zones may overlap to some degree and may change with experience and skill level.

### **5.2. Coping Strategies and Self-Regulation**

When emotions are elevated, coping strategies are used automatically as part of the process of protecting the self. Thus, different emotions produce different hormonal reactions, which trigger certain responses to cope efficiently with current conditions.

The methods athletes use to regulate themselves through the use, misuse, or avoidance of coping strategies determine how emotions are appraised and, by extension, the effectiveness and efficiency of the DM process. Repeated exposure to similar physically and emotionally stressful situations enables an athlete to better tolerate and monitor processes of attention and DM under pressure. Successful self-regulation can be attained and has been proven through the measurement of electroencephalographs, heart rate/pulse, and other physiological responses of the autonomic nervous system. Sustained alertness has been enhanced by avoidance of sleep deprivation; supplementation with nutritional ingredients; elevated physical conditioning; task-related strategies, such as breathing control, relaxation, imagery, and stress inoculation techniques; and conscious control over one's actions.

Coping has two basic forms: emotion-focused and problem-focused. Both strategies are aimed at changing the individual's appraisal of the current situation. An appraisal change results in a new meaning given to the situation, and attentional processes change as a result. Therefore, strategies such as avoidance, denial, and distancing imply one's intention to disengage from interfering stimuli to promote efficient use of visual attention, information processing, and other components necessary for effective DM. In other cases, when emotions are extremely elevated, such strategies may detract from information processing efficiency. Instead, approach strategies may be more

helpful through the use of mental techniques such as relaxation, imagery visualization, and the setting of process and outcome goals. Research has consistently shown that the approach coping strategies are more helpful in monitoring the appraisal process under high perceived pressure, enabling the cognitive system to make decisions more efficiently. The more complex and dynamic the situation, the more beneficial their effect.

### **5.3. Self-Efficacy and DM**

The social cognitive theory, particularly Bandura's self-efficacy theory, has important implications in sport DM. The beliefs athletes hold in their ability to make the right decision at the right time can be viewed as a coping strategy for handling decisions under competitive conditions. Self-efficacy in the context of sport DM has two dimensions: (i) the confidence one has in successfully anticipating and selecting the right response and (ii) the efficacy one has in executing the selected response. The limited research in this area leaves much to speculation. Current research indicates that self-efficacy in anticipatory decisions differs among novice, intermediate, and expert athletes. In fast ball games, experts display moderate levels of self-efficacy in their anticipatory decisions at early stages of response preparation. After the racquetball contact, anticipatory self-efficacy increases sharply. In novice and intermediate players, anticipatory self-efficacy remains moderate throughout the full duration of the task. These findings appear to support the notion that highly skilled athletes hold several alternative anticipatory decisions on alert under extremely short visual exposure frames, although this alters with the course of time—a pattern not typical of less skilled athletes.

## **6. SUMMARY: A HOLISTIC VIEW OF DM IN SPORT**

DM in sport is a process of information processing in natural environments that impose unique constraints on the decision makers. When temporal conditions allow, the athlete relies on his or her experience and practice-related knowledge base to locate cues in the environment that are used for anticipatory DM. With experience and skill development, athletes shift from serial to parallel processing and from target to context visual strategies. These processes enable an athlete to

identify patterns of essential stimuli, elaborate on them efficiently via stored motor programs, and retrieve a response while keeping other responses on alert for possible decision alteration when needed. This process is efficient when the appropriate attentional style is employed throughout variable environmental constraints. The process of DM is more efficient when the athlete perceives the pressure of the sport environment as facilitative and less efficient when debilitating pressure is perceived. Being in the optimal emotional performance zone and feeling a high sense of efficacy for anticipating decisions and selecting responses increase the probability of making the correct decision. To secure an efficient DM process, the successful athlete employs the appropriate methods and mechanisms that underlie a skillful performance.

### See Also the Following Articles

Coping ■ Decision Making ■ Stress

### Further Reading

- Abernethy, B. (2001). Attention. In R. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (pp. 53–85). New York: Wiley.
- Bloom, B. S. (1985). *Developing talent in young people*. New York: Ballantine.
- Ericsson, K. A. (1996). *The road to excellence: The acquisition of expert performance in arts and sciences, sports and games*. Mahwah, NJ: Erlbaum.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725–747.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305.
- Fisher, A. C. (1984). Sport intelligence. In W. F. Straub, & J. M. Williams (Eds.), *Cognitive sport psychology* (pp. 42–50). Lansing, NY: Sport Science.
- French, K. E., & McPherson, S. L. (1999). Adaptations in response selection processes used during sport competition with increasing age and expertise. *International Journal of Sport Psychology*, *30*, 173–193.
- Moran, A. P. (1996). *The psychology of concentration in sport performers: A cognitive analysis*. East Sussex, UK: Psychology Press.
- Ripoll, H. (Ed.) (1991). Information processing and decision making in sport. *International Journal of Sport Psychology*, *3/4*, 187–406.
- Starkes, J. L. (2000). The road to expertise: Is practice the only determinant? *International Journal of Sport Psychology*, *31*, 431–451.
- Straub, W. F., & Williams, J. M. (Eds.) (1984). *Cognitive sport psychology*. Lansing, NY: Sport Science.
- Tenenbaum, G. (Ed.) (1999). The development of expertise in sport: Nature and nurture [Special issue]. *International Journal of Sport Psychology*, *2*, 113–301.
- Tenenbaum, G. (2003). Expert athletes: An integrated approach to decision-making. In J. Starkes, & A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 192–218). Champaign, IL: Human Kinetics.
- Tenenbaum, G., & Bar-Eli, M. (1993). Decision making in sport: A cognitive perspective. In R. N. Singer, M. Murphey, & L. K. Tennant (Eds.), *Handbook on research in sport psychology* (pp. 171–192). New York: McMillan.
- Tenenbaum, G., & Bar-Eli, M. (1995). Personality and intellectual capabilities in sport psychology. In D. Sakulufske, & M. Zeidner (Eds.), *International handbook on personality and intelligence* (pp. 687–710). New York: Plenum.
- Tenenbaum, G., Levy-Kolker, N., Bar-Eli, M., & Weinberg, R. (1994). Information recall among skilled and novice athletes: The role of display complexity, attentional resources, visual exposure duration, and expertise. *Journal of Sport Psychology*, *12*, 529–534.
- Tenenbaum, G., Levi-Kolker, N., Sade, S. S., Lieberman, D., & Lidor, D. (1996). Anticipation and confidence of decisions related to skilled performance. *International Journal of Sport Psychology*, *27*, 293–307.
- Tenenbaum, G., Sar-El, T., & Bar-Eli, M. (2000). Anticipation of ball locations in low and high skill performers: A developmental perspective. *Psychology of Sport and Exercise*, *1*, 117–128.
- Tenenbaum, G., Tehan, G., Stewart, G., & Christensen, S. (1999). Recalling a floor routine: The effects of skill and age on memory for order. *Applied Cognitive Psychology*, *13*, 101–123.
- Williams, A. M., Davids, K., & Williams, J. G. (1999). *Visual perception and action in sport*. Routledge, London.



# Dementia in Older Adults

Marie E. Cowart

Florida State University, Tallahassee, Florida, USA

1. Senility: What It Is and What It Isn't
  2. Myths and Stereotyping
  3. Extent of the Problem
  4. Preventive Measures
  5. Early and Late Recognition
  6. Misdiagnoses and Comorbidities
  7. Interventions and Caring
  8. Future Prospects
- Further Reading

## GLOSSARY

**cognitive function** Intellectual abilities such as the dimensions of attention, language, memory, problem solving, orientation, visual-spatial abilities, judgment, and psychomotor skills.

**cognitive impairment** A general term referring to alterations in cognitive functioning that occur in some elders and that affect behavior and impair the ability to function in usual activities such as activities of daily living (ADLs); such alterations may or may not be reversible, depending on the causative factors.

**dementia** The permanent and progressive deterioration of intellectual function caused by changes in the brain that results in impaired thinking and altered behavioral impulses.

**memory disorders** A general term referring to the wide variety of structural and functional changes in the aging brain as well as a range of symptoms of memory impairments.

**normal changes in aging cognition** The normative changes in cognitive functioning as assessed from a population-based

research perspective; there is a wide individual variability, with some individuals showing no changes at all.

**Parkinsonism** A chronic progressive disease of the nervous system characterized by a slowly spreading tremor, muscle weakness, and (during the later stages) dementia.

**Pick's disease** A brain disorder involving atrophy of the cerebral cortex and resulting in asthenia (i.e., loss of strength), loss of speech, and progressive dementia.

**senescence** A general term referring to the process of growing old.

**senile dementia (or Alzheimer's disease)** A progressive disease characterized by neuritic plaques and neurofibrillary tangles that leads to cognitive decline; it is most common during older adulthood.

**senile plaques** Extracellular brain tissue changes described by Alzheimer in 1907 and found in the brains of persons with Alzheimer's disease and, to a lesser extent, in normal-aged brains.

**senility** A term referring to a generalized deterioration in mental or cognitive skills, leading to a decline in the ability to perform social or occupational activities.

**vascular dementia** The second most common dementia, caused by an interruption in cerebral blood flow due to stroke.

This article deals with the concept of senility using a broad frame of reference. Cognitive effectiveness and age-related cognitive decline or impairment have been the subject of much research, especially during the past decade or so. At this time, there is not a consensus regarding all ideas about screening for dementia and preventive treatment to delay progression of the

disease. During the coming decade, more research findings will help to answer questions about the most reliable screening methods and the efficacy and ethics of using them. Controlled longitudinal studies will help to determine the value of pharmacological and nonpharmacological interventions. This article explores the meaning of senility, what it is and what it is not, who is at risk, prevention, early and late recognition, comorbidities and misdiagnoses, interventions, and future prospects.

## **1. SENILITY: WHAT IT IS AND WHAT IT ISN'T**

### **1.1. History**

Senile dementia was mentioned in the writings of Hippocrates and Pythagoras during the Greco-Roman era. It was recognized by Bacon in 1290, by dePratis during the 1500s, and by Pinel during the 1800s when he reformed the treatment of the mentally ill. At the turn of the 20th century, Alzheimer defined clinical and physiological brain lesions of senile dementia, now referred to as Alzheimer's disease.

### **1.2. Definition of Normal and Impaired Cognitive Functioning**

It is important to differentiate between normal changes in aging cognition and senility or dementia. In normal aging, cognitive processes are slower and less accurate for more effortful or novel tasks than they are during younger years. Therefore, the older person may take longer to absorb the communication message, retrieve a thought, and respond. Memory is a special problem in aging because the older person takes longer to store information so that it may be recalled. Some have likened the aging brain to an overworked computer. The processor of the older computer is slow and the hard drive is overloaded with stored data, thereby making retrieval a more complex and longer process. Sometimes the files get disorganized, particularly when the computer is tired and overused. The new computer has a fast processor and little data stored in it so that retrieval is rapid. In terms of memory retrieval and learning new information, the older person processes information more slowly than does the younger person. The older person's cognition may be likened to the old computer's processor.

Working memory becomes less efficient in the older person. Working memory may be thought of as the ability to retain or temporarily store some information while at the same time accessing or using other information to perform a task. A complex environment, such as a loud and busy room, may interfere with working memory performance in the elderly. Working memory is needed in problem solving.

In terms of problem solving, the older person has more experiences to draw on, so decision making may be more effective in the older person than in the younger and less experienced person. Some refer to this as judgment or wisdom.

In dementia, there is a breakdown in several dimensions of cognition. Judgment is impaired and memory function is disrupted in terms of storage retrieval and planning for future conversations or decisions. Communication breaks down because the person is unable to problem solve or think ahead about the flow of words to use. Confusion and frustration may result, leading to behavioral and emotional changes.

In 1994, the American Psychological Association defined dementia as having the following deficits: memory impairment, cognitive disturbances (e.g., aphasia, apraxia, agnosia), and disturbances in executive functioning (i.e., goal setting) that result in significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning. According to this widely accepted definition, senility or dementia becomes a disability when social or occupational functioning is impaired.

## **2. MYTHS AND STEREOTYPING**

Unfortunately, "ageism" or labeling sometimes can cause an older person to be called senile or demented when perhaps he or she is only hard of hearing or has received a large volume of information rapidly so that there is not time to encode it for memory retention and recall. Such stereotypes are common and permeate our society.

Consider the example of an 85-year-old and a 35-year-old talking with a group about a party they had attended the previous night. Both met a number of new acquaintances, but each could recall only one or two new names. The older person was quickly labeled senile, whereas the 35-year-old was viewed as just not having paid attention.

### 3. EXTENT OF THE PROBLEM

There is a clear association between cognitive decline and aging. This association is confirmed with cross-national prevalence data. Because of longevity, an increased percentage of the population is seen as having cognitive decline compared with earlier generations when the life expectancy was shorter. In 1996, Sloog and colleagues reported that 1% of 65- to 70-year-olds are diagnosed with moderate to severe memory disorders, whereas that number increases to 21% in 85-year-olds. Danish studies published in 2001 reported that among centenarians, 21% show no dementia symptoms, whereas 40 to 63% have mild to severe dementia.

#### 3.1. Positive Risk Factors

There is much discussion that keeping an active mind deters memory and other cognitive function decline, although these approaches have yet to be tested longitudinally with large samples. Prevalence studies do indicate that a higher educational level, high occupational status and workplace complexity for older ages, an intact marriage to a well-educated spouse, a stimulating environment, and engagement in cultural and educational activities are associated with higher intellectual functioning into old age.

#### 3.2. Negative Risk Factors

A number of disease conditions increase the risk of cognitive decline in individuals. Conditions such as a previous head injury, Parkinsonism, and multiple small brain infarcts or hemorrhages are common conditions that are likely to progress to cognitive impairment.

Major risks for dementia or cognitive decline are chronic diseases. Longer life expectancies are accompanied by increased prevalence of chronic conditions, resulting in more dementia in old age today than was the case during earlier decades when life spans were shorter.

Withdrawal from intellectual engagement, such as retirement from an intellectually challenging and complex job, can be a negative risk factor. These and other associations with cognitive functioning provide positive arguments for continued intellectual activity into old age as a deterrent or delay strategy for dementia.

There is some disagreement in the literature as to whether a busier and more complex life can lead to

fatigue that interferes with cognition or, if kept in balance, can provide intellectual stimulation. A lifestyle that allows for self-pacing provides more time for reflection and provides an opportunity to encode new information and retrieve old information without hurrying.

### 4. PREVENTIVE MEASURES

The research literature is clear in that there are no absolute preventive measures for senility or age-related dementia. If genetic predisposition occurs, preventive measures might only delay symptoms. However, studies have shown repeatedly that prevalence rates of dementia are lower in cases where certain associations have occurred. Some of these are discussed in this section.

#### 4.1. Nutrition and Nutritional Supplements

Various studies have shown associations between the prevalence of Alzheimer's disease and the content of food intake. Vitamins that fall into the antioxidant category are considered to be beneficial in preventing dementia due to their ability to repair or replace molecules that are damaged by free radicals. The National Institute on Aging currently supports several large clinical trials that look at the effects of Vitamins A and C, selenium, betacarotene, folate, and multivitamins on cognitive decline and on the prevalence of Alzheimer's disease. The outcome of these large trials will provide additional empirical basis for evaluating the effect of vitamins and food supplements on aging cognition.

Reduced intake of saturated fat and maintenance of a low cholesterol level are believed to lower the risk of the onset of dementia and Alzheimer's disease. Studies have shown some association of lowered prevalence of cognitive decline among elders who eat a diet that is high in fish and other unsaturated fats. Fish oils taken in large amounts are reported to have a positive effect on the retention of cognitive function. These and other food substances have not been well tested with controlled and longitudinal studies on large samples.

#### 4.2. Pharmacological Effects

Repeated studies have also shown some positive association between cognitive functioning and long-term use of nonsteroidal anti-inflammatory drugs such as

the regime that an individual with arthritis might take. People on prednisone, a steroidal drug, have not shown protection against dementia.

There have been a number of studies on estrogen showing mixed results regarding its effect on protecting cognitive functioning in women. The pattern seems to be that the larger trials, such as the Women's Health Study, show that there is no protective effect of hormone replacement therapy in older women on the prevalence of dementia and Alzheimer's disease.

### **4.3. Physical and Mental Activity**

The concept of neuroplasticity applies to the ability of the adult brain to remodel pathways as occurs in the rehabilitation process in persons with stroke or brain injury. Magnetic resonance imaging has helped researchers to visualize brain mapping and to observe the reconfiguration of active pathways. Although confirming research trials have not been conducted to show the brain's ability to remap cognitive and memory function, the concept of "use it or lose it" has gained strength during recent years. Studies showing associations of cognitive function with higher education levels, as well as greater intellectual and social activity, have provided incentives that encourage people to engage in cognitive activities throughout their lifetimes so as to prevent age-related dementia. Although more longitudinal studies are needed, recommendations to engage in lifelong learning, even into old age, are widespread. However, the lasting effects of such behavior are not known.

## **5. EARLY AND LATE RECOGNITION**

Ethicists have discussed the efficacy of screening for dementia. Who decides to screen? What is done with the information? Who decides to share the results of dementia testing? What are the implications for labeling a person with the Alzheimer's disease diagnosis? Who has the right to refuse screening? Does screening make a person disability eligible, or do the results of screening disqualify a person for the workplace or even for health insurance? Before screening for dementia, clinicians should ask themselves these and other questions and discuss the implications of screening with their colleagues and patients.

### **5.1. Early Recognition and Intervention**

In looking for signs of the onset of dementia, it is important to differentiate between normal changes in memory that occur with aging and changes that may be associated with dementia such as Alzheimer's disease. In normal aging, the older person takes longer to learn new information, and the encoding process requires more time for new material to be retained, than is the case in younger persons. Therefore, an apparent inability to learn or to remember what was just heard is not necessarily a sign of early dementia; rather, it may be a sign of slowed cognition that occurs in all aged persons.

Changes with dementia, on the other hand, emerge with difficulty in learning new information. The person often gives up trying to learn. Difficulty in learning subsequently affects the ability to recall, that is, memory. This applies to verbal information as well as spatial information such as remembering directions.

### **5.2. Clinical Picture**

The person with early age-related dementia may be unable to initiate new ideas into conversation, appears to be repetitive in speech, engages mainly in familiar routines to minimize stress, and shows behavior changes such as impatience, frustration, and withdrawal or depression. The person may encounter difficulty in concentrating as well. Complex skills, such as language planning and problem solving, may be shunned if the person with early impairment is aware of his or her own difficulty in cognitive functioning. Later stages show more memory impairment and language skill deterioration. The advanced stage of dementia will show impairment of all mental functions, including changes in personality and emotions exhibited by changes in behavior.

### **5.3. Screening and Diagnostic Tools**

It is important that an assessment of an individual's cognitive functioning not be based on one instrument or tool. A multifaceted assessment of cognition will require an interdisciplinary team approach because special training is necessary to administer these and other approaches to evaluation. Where teams are absent, the use of a variety of assessment tools is

important. A basic assessment regime might include the following procedures:

- History and physical examination
- Neuroimaging and electroencephalography (EEG)
- Blood and spinal fluid laboratory tests
- Social history from a close family member
- Pharmacological and food supplement history
- Evaluation of activities of daily living (ADLs) and instrumental activities of daily living (IADLs)
- One or more of the following cognition data collection instruments:

1. *Mini-Mental Status Examination.* The Mini-Mental Status Examination (MMSE) is a common screening test that is used because of its reliability and validity in evaluating dementia, although it may show some bias for education and socioeconomic status. The 30-item instrument covers orientation, registration, attention, calculation, short-term recall, language, and construction. The test is not sensitive to early dementia.

2. *Clock Drawing Test.* In the Clock Drawing Test, the patient is provided an examiner-drawn circle and asked to put numbers in it and then to draw the clock hands at a given time (e.g., 11:10, 2:45). The properly administered test requires abstraction. Although it is sensitive, it should not be used alone as the sole screen or measure of cognition.

3. *Depression scales.* Although a number of depression scales are available, two seem to have common applicability for the elderly. The Geriatric Depression Scale (GDS) has several versions (5–15 items per version) and is a self-report of how the individual has felt during the past week. The Beck Depression Inventory (BDI) is a 21-item self-report scale that includes some somatic symptoms that may also be found in other chronic conditions. A depression assessment is important when determining dementia in older persons due to the frequency with which depression occurs in the elderly and concern that signs of depression may mimic or cloud those of dementia.

4. *Activities of daily living.* An important part of the dementia assessment is an evaluation of the individual's ability to carry out routine ADLs such as bathing or showering, dressing, eating, transferring, and toileting. Each item is rated as minimal assistance, moderate assistance, maximal assistance, or total assistance. This information may be gained from direct observation, interview of the caregiver or close relative, or self-report. A measure of ADLs can be used to rate the severity of the psychomotor or behavioral accompaniments of dementia.

5. *Instrumental activities of daily living.* The IADL tool provides higher level functions than does the ADL tool. IADLs include food preparation, home maintenance, laundry, financial management, communication (including use of a telephone), use of transportation, and medication administration. The IADL tool has been criticized for having a gender bias due to the food preparation and household items.

## 6. MISDIAGNOSES AND COMORBIDITIES

Careful screening and diagnosis is important in determining the presence of dementia, particularly during the early stages when signs and symptoms might not be obvious. Vision or hearing problems can interfere with verbal or spatial input and can impair the individual's ability to communicate accurately, thereby obscuring some cognitive functions such as the date and time. Sorting out such conditions is facilitated by multidisciplinary evaluation and repeated assessments over time to ensure an accurate diagnosis.

Frequently with the onset of multiple chronic conditions and diseases, it is difficult to determine whether comorbidities precede or follow the onset of dementia. This is particularly true in the case of senility or dementia. There are some frequently observed conditions that are strongly associated with dementia or are known to predispose to cognitive dysfunction. A number of common ones are mentioned in this section.

### 6.1. Head Injury

Depending on the location and severity of a head injury, or in repeated head injuries such as those that boxers may experience, dementia can occur.

### 6.2. Stroke, Atherosclerosis of Cerebral Arteries, High Cholesterol, and Hypertension

Vascular dementia is an outcome of some strokes or advanced atherosclerosis of cerebral arteries causing cerebral emboli. Persons with advanced diabetes can experience vascular dementia.

### **6.3. Oxygen Depletion of Cardiovascular or Respiratory Origin**

Decreased cerebral oxygenation can lower brain metabolism and, over time, can lead to cognitive impairment.

### **6.4. Parkinsonism**

During the advanced stages of Parkinsonism, this neurodegenerative disease can exhibit cognitive impairment, particularly in working memory.

### **6.5. Chronic Mental Illness**

Some long-term mental health problems may show a slow emergence of dementia that may be due to the mental condition or to the effects of long-term therapy.

### **6.6. Drugs or Other Toxic Substances**

Long-term exposure to certain toxic substances can affect cognition. Short-term drug toxicity can cause the person to exhibit changes in cognitive function, particularly orientation, that can be reversible in many instances.

### **6.7. Other Conditions Such as Dehydration, Fatigue, and Relocation**

A fair amount of research points to the effects on working memory and orientation when the older person experiences unusual physical or mental stressors such as those that may occur with dehydration, extreme fatigue, and sudden relocation of the living environment. Research has demonstrated that older people have times of peak functioning during the day, usually the morning hours. Careful assessment of the entire person is essential before applying the label of senility or dementia.

## **7. INTERVENTIONS AND CARING**

There is a great deal of literature on interventions for individuals with moderate and advanced dementia. This article deals with some early interventions and

concerns for the care of persons with cognitive impairment.

### **7.1. Dietary Supplements and Medications**

Aricept (other trade names include Cognex and Exelon) is the most widely used drug to alleviate the symptoms of Alzheimer's disease. It does this by prolonging the effects of acetylcholine release events, thereby raising neurotransmission and increasing neurotransmission effectiveness. Its use is thought to delay the advancement of symptoms rather than to provide a cure, although side effects may interfere with use of the drug.

### **7.2. Memory Aids and Other Nonpharmacological Treatments**

The use of memory cues, such as written notes, is thought to help prolong independence in people who have progressive age-related dementia. Some literature indicates that people with mild or even moderate dementia retain reading skills longer than they do oral skills. There is also evidence that the use of written language becomes less complex in terms of sentence structure and vocabulary as the dementia progresses.

If familiar routines are retained in persons with mild to moderate dementia, independent living can be extended. Conversely, learning new routines can trigger undesirable or stressful behavior. Community-based group programs are often available through a day care format and use a variety of activity approaches to enable the individual to retain a desirable quality of life and to provide respite for the family caregiver.

### **7.3. Management Concerns**

As the dementia condition progresses from mild to moderate levels of impairment, a number of concerns that need to be addressed may arise. Because dementia emerges in very different ways for each individual, not everyone with moderate or advanced dementia will experience these problems. However, an awareness of possible concerns will enable the practitioner to aid the family caregiver in planning ahead to prevent unsafe or disturbing behaviors.

Driving is an activity that is closely identified with a person's sense of autonomy and independence. Yet as the individual's judgment and problem solving become impaired, it is not unusual for the person with

dementia to become lost or to make wrong decisions that can result in an auto accident. Determining when it is no longer safe for the individual to drive is a problem that family members face and that often requires their intervention.

Wandering is another concern for the ambulatory person with moderate or advanced dementia. Assistive devices, such as radio and identifier tags, can help to quickly locate the individual who has the potential for wandering. Failure to locate the person who has wandered may lead to dehydration, hunger, hypothermia, and even death.

When the person with moderate dementia is unable to carry out two or more ADLs, additional support for the family might need to be considered. Such support may be provided by adding additional assistance in the home. If the behavior is unmanageable, a time may come when the family decides to move the individual to a protective environment such as an assisted living facility for persons with dementia. Measures of family coping may assist in determining the amount of caregiving burden that the caring family members are experiencing and may serve as a useful tool in assisting the family with making decisions about care.

Repetitive forgetfulness can pose problems for caregiving. Forgetfulness can include leaving a stove on, leaving a cigarette lit, and not being able to locate familiar objects such as keys.

These are just a few of the concerns that can arise in providing care for the person with moderate dementia. The caregiver can gain many tips on how to deal with caring for the family member with dementia by joining a support group and sharing concerns with others who face similar experiences.

## 8. FUTURE PROSPECTS

Many call the decade of the 1990s as the “decade of the brain.” Although much knowledge has been gained about brain function, there is much more to learn. Further advancements in maintaining cognitive effectiveness are certain to be made, although a cure for age-related dementia is not anticipated. Whether new approaches, such as brain stimulation from electrical or pharmaceutical products, are developed and are effective in lessening the progression of dementia is another area for empirical study.

Recent brain research, summarized in the September 2003 issue of *Scientific American*, indicates that there is evidence that brain pathways for motor control have

the capacity to change. This phenomenon is referred to as plasticity. Future research will demonstrate whether repatterning can occur in intellectual functioning as it has with physical activity.

Impending demographics in terms of increased life expectancy and population aging means that there will be increasing numbers of elders throughout developed and newly industrialized nations. Greater numbers of people living longer will mean that increases in the prevalence of degenerative diseases, including dementia, will be experienced throughout the world. If preventive approaches and proposed therapies are found to be effective, much will be saved in terms of human and economic resources, and the quality of life of large numbers of elders will benefit.

## See Also the Following Articles

Aging and Competency ■ Aging, Cognition, and Medication Adherence ■ Cognitive and Behavioral Interventions for Persons with Dementia ■ Cognitivism ■ End of Life Issues ■ Psychotherapy in Older Adults

## Further Reading

- Blackman, L., Small, B. J., Wahlin, A., & Larson, M. (1999). Cognitive functioning in very old age. In F. I. M. Craik, & T. M. Salthouse (Eds.), *Handbook of aging and cognition* (Vol. 2, pp. 449–558). Mahwah, NJ: Lawrence Erlbaum.
- Chen, P., Ratcliff, G., Belle, S. H., Cauley, J. A., DeKosky, S. T., & Ganguli, M. (2001). Patterns of cognitive decline in presymptomatic Alzheimer's disease: A prospective community study. *Archives of General Psychiatry*, *58*, 853–858.
- Elias, M. F., Beiser, A., Wolf, P. A., Au, R., White, R. F., & D'Agostino, R. B. (2000). The preclinical phase of Alzheimer's disease: A 22-year prospective study of the Framingham cohort. *Archives of Neurology*, *57*, 808–813.
- Fratiglioni, L., & Rocca, W. (2001). Epidemiology of dementia. In F. Boller, & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 6, pp. 193–215). Amsterdam: Elsevier.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *58*, 2034–2039.
- Pope, S. K., Shue, V. M., & Beck, C. (2003). Will a healthy lifestyle help prevent Alzheimer's disease? *Annual Review of Public Health*, *24*, 111–132.
- Rabbitt, P. (1998). Aging of memory. In *Brocklehurst's textbook of geriatric medicine and gerontology* (5th ed., pp. 123–152). New York: Churchill Livingstone.
- Smith, A. (1996). Memory. In J. E. Birren, & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (4th ed., pp. 236–250). San Diego: Academic Press.



# Depression in Late Life

Barry A. Edelstein, Andrea Shreve-Neiger, and Stephanie A. Scheck

West Virginia University, Morgantown, West Virginia, USA

1. Epidemiology of Late-Life Depression
  2. Conceptual and Diagnostic Issues
  3. Relations among Depression, Physical Diseases, and Medications
  4. Assessment of Depression
  5. Instruments for the Assessment of Depression
  6. Treatment of Late-Life Depression
- Further Reading

## GLOSSARY

**age-related changes** Normal physical and psychological changes from young adulthood through old age.

**comorbidity** The co-occurrence of two disorders or syndromes.

**late-onset depression** Depression with an onset at age 65 or older.

**major depression** Major depressive disorder as defined in the *Diagnostic and Statistical Manual of Mental Disorders*. It is characterized by depressed mood for at least 2 weeks and at least four additional symptoms of depression.

**minor depression** Minor depressive disorder as defined in the *Diagnostic and Statistical Manual of Mental Disorders*. It is characterized by depressed mood for at least 2 weeks but with few symptoms and less impairment than in major depressive disorder.

**reminiscence therapy** Involves recall of events in a person's life and is usually done in groups in which participants are encouraged to talk about past events, sometimes with such aids as photos, music, and personal objects.

**somatic symptoms** Fatigue, changes in appetite, changes in sexual activity, and changes in sleep duration.

**subsyndromal symptoms** Symptoms of depression that are not of sufficient frequency, duration, or intensity to meet the formal criteria for an affective disorder.

The 1999 Surgeon General's Report on Mental Health revealed that almost 20% of older adults experience mental disorders. Depression, although not the most prevalent mental disorder among older adults, is a very significant source of suffering. From a clinical perspective, it is a particularly complex and challenging phenomenon, due in part to age-related differences in the experience and presentation of depressive symptoms, risk factors, and the many potential causes of depressive symptoms (e.g., medications, comorbid diseases, and stroke).

## 1. EPIDEMIOLOGY OF LATE-LIFE DEPRESSION

Depression in older adults is often overestimated and underrecognized, and it is not the most prevalent psychiatric disorder among older adults, as many believe. Indeed, the prevalence of depression is lowest among adults aged 65 or older. Prevalence rates for major depressive disorder among community-dwelling older adults (65+ years) range from 1 to 4%. The North Carolina Epidemiologic Catchment Area Study estimated a prevalence of 0.8% for major depression, 2% for dysthymia, and 4% for minor depression. Incidence

estimates of depression in long-term care facilities have ranged from 12.4 to 14%. If one considers clinically significant depressive symptoms that do not meet diagnostic criteria (subsyndromal symptoms), the rate of depression ranges from 8 to 15% for community-dwelling older adults, 25 to 33% for hospitalized older adults, 10 to 15% for nonpsychiatric outpatients, and 30 to 40% for long-term care residents. The Epidemiological Catchment Area Studies reported bipolar I and bipolar II prevalence rates of approximately 0.1% among older adults.

## 2. CONCEPTUAL AND DIAGNOSTIC ISSUES

There is increasing evidence that depression among older adults differs from that of younger adults along several dimensions (e.g., etiology, presentation, and phenomenology). In 1988, Alexopoulos *et al.* suggested that the etiology of depression among older adults appears to be somewhat related to the age of onset. Gatz and Fiske noted that among depressed older adults, approximately 50% are cases of late-onset depression. The etiology of late-onset depression may be quite different from that of depression occurring in earlier years. The experience and presentation of depression among many older adults appears to be different than that of younger adults. For example, depressed older adults are less likely to report dysphoria than younger adults with the same level of overall depression. Gatz and Hurqitz revealed that older adults are also more likely than younger individuals to endorse test items suggesting a lack of positive feelings, whereas younger adults are more likely to endorse test items suggesting depressed mood. Older adults are also less likely to report ideational symptoms (e.g., guilt) than younger adults and more likely to report somatic symptoms. These somatic symptoms, such as fatigue and changes in appetite and sexual activity, may be related to coexisting medical conditions or normal age-related changes. Reports of lack of interest, sleep disturbance, and lowered energy level, however, may be useful in differentiating depressed and nondepressed older adults.

In addition to the content of age-related symptoms of depression, there is evidence that symptom levels and patterns that would not currently meet *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV), criteria for major disorders are nevertheless

problematic for older adults. Many older individuals present with somatic complaints and experience symptoms of depression that do not meet the criteria for depressive disorders. The effects of these subsyndromal conditions may be just as consequential as those of the disorders that do meet diagnostic criteria.

In summary, depression among older adults appears to be different in a number of ways from that encountered in younger adults. The etiology of depression among older adults appears to vary depending on the age of onset. Moreover, the nature and pattern of depressive symptoms of older adults appear to deviate from those one would expect based on depression criteria found in the *DSM-IV*.

## 3. RELATIONS AMONG DEPRESSION, PHYSICAL DISEASES, AND MEDICATIONS

One of the complicating factors faced by clinicians and depression researchers regarding the nature and etiology of depression among older adults is that older adults have a higher incidence of medical disorders than younger adults. Approximately 80% of older adults suffer from at least one chronic health problem. Consequently, it is relatively common for depressed older adults to have comorbid medical disorders; that is, they exhibit the co-occurrence of two disorders or syndromes.

Although the relations between depression and comorbid disorders can be rather complex, they are generally well established. Depression can precede, follow, or simply co-occur with unknown directional relations. Depression is a risk factor for physical disease (e.g., cardiovascular disease) and diminished cell-mediated immunity. Symptoms of depression can also arise directly from various physical diseases (e.g., hypothyroidism, stroke, Parkinson's disease, and pancreatic cancer) and from the psychological reaction to a disease (e.g., Alzheimer's disease). Depressive symptoms may also follow a disease process (e.g., myocardial infarction). Similarly, depression is very common among cancer patients, with rates varying from 25 to 50%; there is a greater prevalence among individuals experiencing higher levels of pain, physical disability, and illness. When considering the relations between physical disease and depression, it is important to understand that one need not meet criteria for depression for the symptoms of depression to affect physical health.

The complexity of the relations between depression and physical diseases is often amplified by the adverse effects of medications used to treat medical diseases (e.g., parkinsonism, pain, cancer, arthritis, and hypertension) because many of these medications can produce symptoms of depression. Older adults consume more medications than any other age group and are at higher risk of adverse drug reactions than any other age groups because of age-related changes in physiology and increased use of multiple medications.

#### 4. ASSESSMENT OF DEPRESSION

The assessment of depression among older adults can be quite challenging, particularly in light of the issues discussed previously. Age-related differences in the presentation of depression as well as the potential for coexisting physical ailments, medication side effects, and natural age-related changes increase the complexity of diagnosing depression when assessing older adults.

Methods of assessment for depression include self-report, clinical interview, report by others, physical examination, and direct observation. Given the intricacy of identifying depression in older adults, there may be no single best method for assessing depression. Each method has its own relative strengths and weaknesses; however, combining methods during the assessment process can enable a clinician to obtain a more complete understanding of depressive symptomatology in older adults. Here, discussion is limited to the most commonly used self-report and clinician-rated instruments.

### 5. INSTRUMENTS FOR THE ASSESSMENT OF DEPRESSION

The most widely used instruments for assessing depression use the self-report method due to the relative ease of administration and brevity. Clinician rating scales, although often taking slightly longer to administer, offer the advantage of objective observation of the depressive symptoms, which can be particularly useful with cognitively impaired older adults.

#### 5.1. Self-Report Instruments

##### 5.1.1. Beck Depression Inventory

The Beck Depression Inventory (BDI) is a 21-item, multiple-choice inventory. Respondents are asked to

rate each item based on four response choices according to the severity of the symptoms, ranging from the absence of a symptom to an intense level, during the past week. A 13-item version of the BDI is also available. Both the original and short forms have reasonable internal consistency for normal and depressed older adults and adequate test-retest reliability in older adult patient and nonpatient populations.

A second version of the BDI (BDI-II) was published in 1996 so that items would be consistent with the revised criteria for depressive disorders in the *DSM-IV*. The number of items on the BDI-II is the same as that for the original version. Edelman *et al.* suggest that the BDI-II be used with caution with older adults given the lack of empirical support for its use with this population.

##### 5.1.2. Center for Epidemiological Studies-Depression Scale

The Center for Epidemiological Studies-Depression Scale (CES-D) is a 20-item instrument with good psychometric properties. A 10-item version of the CES-D also has adequate psychometric properties and appears to be a good instrument for screening for depression in community-dwelling older adults.

##### 5.1.3. Geriatric Depression Scale

The Geriatric Depression Scale (GDS) is a 30-item scale designed specifically for use with older adults. The scale uses a yes-no format and has well-established psychometric properties with community-dwelling older adults, outpatient and hospitalized older adults, and geriatric stroke patients. A 15-item form of the GDS yields scores that are strongly correlated with those of the long version. Overall, the GDS appears to be a useful instrument for assessing depression in older adults.

##### 5.1.4. Zung Self-Rating Depression Scale

The Zung Self-Rating Depression Scale contains 20 items that are rated on a 4-point scale by the patient according to the amount of time he or she currently experiences the symptom. Psychometric properties with older adults are less than desirable.

## 5.2. Clinician Rating Scales

### 5.2.1. Hamilton Rating Scale for Depression

The Hamilton Rating Scale for Depression (HRSD) is used to quantify the severity of symptoms of depression and is one of the most widely used and accepted instruments for assessing depression. The standard version of the HRSD is designed to be administered by a trained clinician, and it contains 17 items rated on either a 3- or 5-point scale, with the sum of all items making up the total score. The HRSD may be a useful scale for cognitively impaired patients who have difficulty with self-report instruments.

### 5.2.2. Geriatric Depression Rating Scale

The 35-item Geriatric Depression Rating Scale (GDRS) combines the severity rating format of the HRSD with the content of the GDS. The GDRS probably requires less experience and training to administer than the HRSD and has good psychometric properties with hospitalized, outpatient, and community-dwelling older adults.

## 5.3. Instruments for Assessing Depression with Coexisting Dementia

Individuals with dementia pose a particular challenge when assessing for depression due in large part to the overlap of symptoms of depression and dementia as well as the effects of cognitive impairments and memory deficits. Several instruments have been developed to address the complicated nature of assessing depression with coexisting dementia, including the Cornell Scale for Depression in Dementia (CS) and the Dementia Mood Assessment Scale (DMAS).

### 5.3.1. Cornell Scale for Depression in Dementia

The CS contains 19 items that assess for the severity of depression based on a 3-point scale with content similar to the that of the HRSD. Psychometric properties of the CS are adequate. The CS appears to be a useful measure for assessing severity of depressive symptoms in older adults with dementia or cognitive impairments.

### 5.3.2. Dementia Mood Assessment Scale

The 17-item DMAS is a brief measure of mood in patients with dementia that includes a semistructured

interview and direct observation of the individual over time. The psychometric properties of the DMAS are not strong. Its advantage lies in the combination of interview and direct observation.

In summary, there are a wide range of assessment instruments available for the assessment of depression among older adults, with each having strengths and weaknesses. Although many of the instruments were developed for use with younger adults, they have proven reliable and valid with older adults with a wide range of clinical characteristics.

## 6. TREATMENT OF LATE-LIFE DEPRESSION

Despite the availability of treatments for depressed older adults, only a few older adults utilize such treatments. Depressed older adults are most frequently first seen by a primary care physician, usually for somatic symptoms and complaints of low energy. Unfortunately, many older adults and their physicians attribute these symptoms to “normal aging” or a physical ailment and fail to make a mental health referral.

Although presentation and diagnostic considerations for depression appear to differ in older adults when compared to their younger counterparts, Karel and Hinrichsen’s review of the literature suggests that treatment interventions effective for younger adults are also efficacious for older adults. This section focuses on psychosocial interventions, with an emphasis on evidence-based treatments for older adult depression.

### 6.1. Evidence-Based Psychosocial Treatments

A number of psychosocial interventions have proven effective for treatment of older adult depression. These treatments include psychodynamic, cognitive, behavioral, cognitive-behavioral, and interpersonal therapies. Although the rationale for each form of therapy is derived from varying conceptualizations, a number of studies have demonstrated the effectiveness of each of these therapies for treatment of depressed older adults.

#### 6.1.1. Psychodynamic Therapies

Psychodynamic therapies focus on intrapsychic conflicts that may be affecting the individual’s coping and adjustment. Emotional insight, gained through the therapeutic

process, is the goal of treatment for resolution of these conflicts and establishment of more effective coping styles. In addition, this form of therapy focuses on resolution and acceptance of social and physical losses that are characteristic of old age and addresses unresolved issues from various stages of development that may be contributing to the person's distress. Psychodynamic therapies also frequently incorporate a focus on the "self" and maintaining self-esteem through the trials and tribulations of the aging process.

### **6.1.2. Behavior Therapy**

Behavior therapy relies on the premise that overt activity and mood are closely linked. In 1974, Lewinsohn, who offered one of the principal behavioral theories of depression, demonstrated that the lack of sufficient positive reinforcement from one's environment can negatively affect one's mood state. Zeiss and Lewinsohn adapted Lewinsohn's behavioral approach for older adults. This form of therapy typically involves behavioral activation, scheduling of pleasant events into the person's life, decreasing one's participation in aversive activities, and problem solving and social skills training. In addition, "homework" assignments are frequently given to facilitate the achievement of clearly developed goals and skills and activate the person for change.

### **6.1.3. Cognitive Therapy**

Cognitive therapy, as developed by Beck *et al.* in 1979, assumes that maladaptive and irrational cognitions are contributing to, if not causing, the depressed mood. This therapy focuses on identifying distorted or irrational thoughts and beliefs that may be contributing to the mood state. During therapy, the person is encouraged to monitor, challenge, and eventually replace negativistic and irrational thoughts with realistic, more positive ones. Practice through completion of homework assignments is integral to this form of therapy, and clients are encouraged to monitor and challenge irrational or negativistic cognitions in their daily lives once they have practiced in a therapeutic setting.

### **6.1.4. Cognitive–Behavior Therapy**

Gallagher-Thompson and Thompson combined behavioral and cognitive therapies, which they termed cognitive–behavior therapy, for the treatment of depression in older adults. This form of therapy is

based on the rationale that irrational cognitions and/or maladaptive behaviors contribute to and maintain a depressed mood state. It is assumed that negative cognitions or thoughts, especially about one's self, future, and experiences, create a negative lens through which one sees the world. In addition, the depressed older adult is not engaging in enough positive mood-enhancing activities, which tend to exacerbate the negative cognitions. This form of therapy is generally psychoeducational, goal-directed, structured, and brief, designed to give the depressed older adult the cognitive and behavioral skills (e.g., relaxation techniques, assertiveness training, and problem-solving skills) he or she is lacking.

### **6.1.5. Interpersonal Psychotherapy**

Finally, interpersonal psychotherapy (IPT), developed by Klerman *et al.* in 1984, is a time-limited form of therapy that emphasizes interpersonally relevant issues that may have preceded or resulted from depression. Specifically, this therapy stresses the importance of family relations in older adults' lives and, as such, emphasizes the nature of important relationships and focuses on one of four potential interpersonal problems: grief, interpersonal dispute, role transition, and interpersonal deficits. Treatment strategies are then implemented to address each of these problems and reach identified treatment goals. In a review of the use of IPT with older adults, Hinrichsen found several studies that demonstrated the effectiveness of this therapy for older depressed adults.

### **6.1.6. Other Psychosocial Treatments**

Although the previously described therapies have been well established as effective for use with older adults, a number of other therapeutic approaches are used, although little empirical evidence is available for their efficacy in this population. It should be noted, however, that simply because an intervention has not been empirically studied does not indicate that it is not useful or effective in the treatment of older adult depression. Several of these forms of treatment have been examined through preliminary studies that have rendered promising results.

**6.1.6.1. Life Review Therapy** Life review therapy, a form of reminiscence therapy, encourages the older adult to reflect on his or her life and is thought to facilitate his or her progress through Erikson's final developmental task and emerge with a sense of integrity

or despair. Although it is used for the treatment of depression in older adults, some caution its use with this population because it could prove more harmful than helpful, especially for older adults for whom a life review would not be a positive experience. Researchers continue to debate the efficacy of this form of therapy.

**6.1.6.2. Group Therapy** Group therapy, especially psychoeducational group therapy, is another form of treatment for older adult depression. Group therapy captures the interpersonal or social component not seen in other forms of therapy and thus allows for older adults to share with and learn from each other while acquiring behavioral and cognitive coping skills. This form of therapy can be quite effective with depressed older adult family caregivers.

**6.1.6.3. Less Common Forms of Treatment** Less common forms of treatment for older adult depression include bibliotherapy and self-help therapy, both of which emphasize the individual's role in treatment and overcoming depression without a great deal of intervention from a therapist. Typically, in both of these forms of therapy, the older adult is educating himself or herself through readings. In 1989, Scogin *et al.* conducted one of the few studies on the efficacy of bibliotherapy and found promising and generally positive results.

A number of treatments or therapies for older adult depression have been described briefly in this section. Although there are several well-established, effective treatments for depression in this population, there is a need for additional research to empirically examine some of the other less common but potentially advantageous forms of therapy. There is also a need to make treatments more available and acceptable to older adults so that when referred by physicians, depressed older adults will seek treatment.

### See Also the Following Articles

Aging and Competency ■ Elder Caregiving ■ Psychotherapy in Older Adults

### Further Reading

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Blazer, D. (2003). Depression in late life: Review and commentary. *Journal of Gerontology: Medical Sciences*, 58A, 249–265.

Blazer, D. G. (2002). *Depression in late life*. New York: Springer.

Coon, D. W., Rider, K., Gallagher-Thompson, D., & Thompson, L. (1999). Cognitive-behavioral therapy for the treatment of late-life distress. In M. Duffy (Ed.), *Handbook of counseling and psychotherapy with older adults* (pp. 487–510). New York: Basic Books.

Edelstein, B. A., Kalish, K. D., Drozdick, L. W., & McKee, D. R. (1999). Assessment of depression and bereavement in older adults. In P. Lichtenberg (Ed.), *Handbook of assessment in clinical gerontology* (pp. 11–58). New York: Wiley.

Fiske, A., Kasl-Godley, J. E., & Gatz, M. (1998). Mood disorders in late life. In B. Edelstein (Ed.), *Clinical geropsychology* (Vol. 7, pp. 193–229). New York: Elsevier.

Frank, E., Frank, N., Cornes, C., Imber, S. D., Miller, M. D., Morris, S. M., & Reynolds, C. F. (1993). Interpersonal psychotherapy in the treatment of late-life depression. In G. L. Klerman, & M. M. Weissman (Eds.), *New applications of interpersonal psychotherapy* (pp. 167–198). Washington, DC: American Psychiatric Press.

Gallagher-Thompson, D., & Thompson, L. W. (1996). Applying cognitive-behavioral therapy to the psychological problems of later life. In S. H. Zarit, & B. G. Knight (Eds.), *A guide to psychotherapy and aging: Effective clinical interventions in a life-stage context* (pp. 61–82). Washington, DC: American Psychological Association.

Gatz, M., & Fiske, A. (2003). Aging women and depression. *Professional Psychology: Research and Practice*, 34, 3–9.

Gatz, M., Fiske, A., Fox, L., Kaskie, B., Kasl-Godley, J., McCallum, T., & Wetherell, J. (1998). Empirically-validated psychological treatments for older adults. *Journal of Mental Health and Aging*, 41, 9–46.

Gatz, M., & Hurwicz, M. (1990). Are old people more depressed? Cross-sectional data on Center for Epidemiological Studies Depression Scale factors. *Psychology and Aging*, 5, 284–290.

Hinrichsen, G. (1999). Interpersonal psychotherapy for late-life depression. In M. Duffy (Ed.), *Handbook of counseling and psychotherapy with older adults* (pp. 470–486). New York: Wiley.

Jamison, C., & Scogin, F. R. (1992). Development of an interview-based geriatric depression rating scale. *International Journal of Aging and Human Development*, 35, 193–204.

Karel, M., & Hinrichsen, G. (2000). Treatment of depression in late life: Psychotherapeutic interventions. *Clinical Psychology Review*, 20, 707–729.

Karel, M. J., Ogland-Hand, S., & Gatz, M. (2002). *Assessing and treating late-life depression: A casebook and resource guide*. New York: Basic Books.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy of depression*. New York: Basic Books.

Knight, B. G., Santos, J., Teri, L., & Lawton, M. P. (1995). The development of training in clinical geropsychology. In B. G. Knight, L. Teri, P. Wohlford, & J. Santos (Eds.),

- Mental health services for older adults: Implications for training and practice in geropsychology.* Washington, DC: American Psychological Association.
- Mulsant, B. H., & Ganguli, M. (1999). Epidemiology and diagnosis of depression in late life. *Journal of Clinical Psychiatry, 60*(Suppl. 20), 9–15.
- Parmalee, P. A., Katz, I. R., & Lawton, M. Pl. (1993). Anxiety and its association with depression among institutionalized elderly. *American Journal of Geriatric Psychiatry, 1*, 46–58.
- Storandt, M., & VandenBos, G. R. (Eds.). (1994). *Neuropsychological assessment of dementia and depression in older adults: A clinician's guide.* Washington, DC: American Psychological Association.
- Weissman, M. M., Bruce, M. L., Leaf, P. L., Florio, L. P., & Holzer, C., III. (1991). Affective disorders. In L. N. Robins, & D. A. Regier (Eds.), *Psychiatric disorders in America: The epidemiologic Catchment Area Study* (pp. 53–80). New York: Free Press.



# Developmental Counseling

David A. Jepsen

University of Iowa, Iowa City, Iowa, USA

1. Introduction
  2. The Developmental Viewpoint
  3. Collaborative Relationships
  4. Guiding Principles for Developmental Counseling Processes
  5. Appraisal in Developmental Counseling
  6. Developmental Counseling Interventions for Resolving Current Developmental Problems
  7. Developmental Counseling Interventions for Preventing or Reducing Developmental Problems
- Further Reading

*interventions for prevention* Interventions designed to prevent problems from either becoming more serious or expanding to other areas of functioning.

*life cycle* Differences in a person's activities across time.

*life structure* Differences in a person's activity pattern or configuration at a particular point in time.

Developmental counseling is a collaborative and catalytic relationship formed between a professional counselor, who is trained in both counseling and human development, and a client, who is currently experiencing or is expected to experience problems that impede progress toward mastering personally and/or culturally valued functions for the purpose of co-constructing new meanings of current actions and/or new possibilities for future actions.

## GLOSSARY

*appraisal* Collecting and synthesizing data to construct a description of a client's life history and current problems.

*collaborative relationship* A counselor and client working together jointly toward agreed-on goals.

*developmental change* Relatively enduring irreversible differences over time in a person's activity pattern judged as movement toward mastery of culturally valued functions and social competencies.

*developmental problem* A "gap" between present experiences and experiences desired by the developing person or by people who have a stake in the person's development.

*interventions* Counselor-initiated acts that literally "go between" events sequenced along a client's life span, between the client's problem-saturated activity and future growth/movement toward resolving problems and addressing new challenges.

## 1. INTRODUCTION

Most people experience problems while growing up, that is, problems that impede their progress toward mastering personally and/or culturally valued functions. Most contemporary cultures value successful functioning as a friend, a learner, a producer/worker, a spouse or partner, and a citizen. Furthermore, the specific cultural group with which a person identifies—not necessarily the dominant culture—sets a "timetable" for growing up that specifies the period in life during which the person's attentions to these functions should begin, peak,

and diminish. Problems arise when people perceive a “gap” between their current experiences and the experiences desired by themselves, by important others in their lives, and/or by cultural standards. People may experience developmental problems such as making and keeping friends, finding and sustaining meaningful work, achieving intimacy, and defining and implementing a sense of unique personal identity.

Described in another way, a person’s development is problematic when his or her activity is not organized effectively to perform major life functions that the person’s culture considers to be fundamental to the human condition. Although most cultures seem to agree on some essential human functions, they differ on others. For example, achieving a sense of personal identity is more prominent in cultures that emphasize individualism and autonomy than in those that emphasize community and belonging. Cultures also differ in their prescribed cultural pathways to achieving competence. For example, some cultures promote pathways of personal independence, whereas others promote pathways of social interdependence.

Over the past 50 years or so, several writers have advocated an approach to counseling that addresses developmental problems. The aggregate scholarship on developmental counseling remains somewhat informal and eclectic. Perhaps the most systematic treatment has been developmental counseling and therapy (DCT) advanced by Ivey and colleagues. Given such wide-ranging ideas and practices, this article represents a composite definition amalgamated from many sources.

Developmental counseling represents a broad counseling approach designed to help a wide range of people who are experiencing problems of growing up across the life span. Because few counselors can master all of the skills and knowledge necessary to serve all people with developmental problems, developmental counselors usually specialize in working with clients from a particular age group (e.g., counselors for children), address a particular cultural function (e.g., career counselors), or identify with a particular subculture (e.g., counselors for women).

The delivery of developmental counseling can be distinguished from other counseling approaches. Most notably, developmental counselors work with people to activate their movement toward developmental change rather than to remediate pathology. Developmental counselor–client contacts usually are for a limited number of meetings and may be intermittent over longer periods of time, similar to the relationships that many adults maintain with their dentists or accountants.

Developmental counselors serve as catalysts, rather than as the sole providers or vehicles, for clients to change and mobilize every helping resource in their environment—sometimes called activating a network of resources—so as to stimulate and support developmental changes. Client growth is expected to occur both outside and within the counseling relationship.

Developmental maturation or growing up is given meaning both from the perspective of the person who is experiencing it and from the perspective of external observers. If the person who is experiencing developmental problems seeks the help of an expert “external observer” called a professional counselor, that collaboration is called developmental counseling. Sometimes, developmental counselors will work with groups of people who are about to experience common developmental problems in an effort to prevent more serious problems. Professional counselors provide, among other initiatives, a fresh perspective on developmental problems experienced by clients, a perspective that is informed by knowledge of human development and cultural standards. Although developmental counselors employ many of the same ideas and techniques as do counselors with other theoretical preferences, the former are most clearly set apart by subscribing to a developmental viewpoint.

## **2. THE DEVELOPMENTAL VIEWPOINT**

Development generally refers to changes in patterns of overt and covert activity as people grow up or mature in their capacity to perform several culturally valued life functions. Individual development involves multiple changes as the person strives to master many life functions simultaneously. When such changes are judged as movement toward cultural ideals of maturity or healthiness, there is general agreement that the person made developmental changes. Developmental change is progressive and incremental, usually involving movement away from simple, dependent, self-centered activity patterns that are typical of younger members of a society and movement toward complex, independent, society-centered activity patterns that are typical of mature members of a society. The metaphor for developmental change is usually a spiral rather than an inclined plane, thereby representing back-and-forth movement that is predominantly forward. By contrast, development is not construed as change in traits (e.g., becoming less

introverted or more intelligent), as change in states (e.g., becoming less anxious or more expressive), or as change in specific isolated behaviors (e.g., eliminating socially disruptive behaviors, increasing frequency of approaching peers).

Developmental counselors conceptualize differences in individual clients' activity along two dimensions: the time across the life cycle and the activities within a contemporary life structure. The life cycle traces differences in a person's activities across time, and the life structure specifies differences in a person's activity pattern or configuration at a particular point in time. Development usually connotes quantitative changes (e.g., frequency, intensity, duration) in age-graded normative activity that is typical, rather than unusual, for the age group. For example, the duration of friendships tends to lengthen during late childhood and adolescence. Development also refers to qualitative changes in action patterns (e.g., complexity, coherence) that serve the same broad, culturally valued life functions. For example, patterns of values expressed as reasons for moral or career choices are generally more complex and coherent during adolescence than during childhood. Developmental change is also reflected in the transformation of the organization of activity—the life structure—as well as in the activity content.

Life span simply denotes the sequence of activities across the whole of life from birth to death. In contrast, the developmental life cycle denotes changes that meet three conditions:

- A sequence of changes over the life span that are undergone as people master progressively more complex life functions
- Changes over time through segments of the life span, called life stages, that are characterized by distinctly different activities and expectations
- Stages of change that replicate the progressive mastery experienced by preceding generations

An essential idea in development is the life stage, an age-graded segment of the life cycle that is demarcated by changes in the physical body, in social expectations, and (consequently) in the psychological activity responding to these changes. Each life stage includes a set of developmental tasks that must be mastered, or normative problems that must be solved, within an age range as expected by the culture and, therefore, must follow the culturally imposed timetable. These tasks can be carried out well or poorly. Mastery of tasks within an early stage is essential to mastering other tasks within that stage and as a precursor to mastering tasks within later stages.

Failure to resolve tasks from an earlier stage often is experienced as unfinished business and is likely to produce recurrent problems later in life, sometimes called delays in development. Developmental counselors often encounter young adult clients who are experiencing developmental problems because they did not master the tasks during childhood and adolescence.

In contrast to developmental tasks are developmental marker events that occur at various ages/stages and that require the person to make adaptations in activities and roles. Such events often result from circumstances beyond the person's control (e.g., illness, death of a loved one, accident, job loss, military conscription, natural disaster). Other marker events are likely to be anticipated (e.g., marriage, childbirth, job changes, retirement).

Developmental life structure denotes the contemporary pattern or configuration of activities organized to fulfill culturally valued functions and to achieve age-appropriate developmental tasks. Activity patterns are often distinguished as social roles (e.g., friend, student, jobholder, citizen), each of which represents aggregate responses to cultural expectations and developmental tasks. The multiple social roles that one person plays often introduce problems such as conflicting demands from two or more roles. For example, a worker may be expected to meet production goals that require time and energy usually devoted to family. As people age and developmental tasks change, the activity that is considered appropriate for the age and role also changes; therefore, life structures are reorganized into new activity patterns. For example, when high school or college graduates move to different communities, start new jobs, and start families, their activity patterns become reorganized. The transition from one life structure to another presents a person with an opportunity to address new developmental tasks and problems.

### 3. COLLABORATIVE RELATIONSHIPS

The developmental counselor and client develop a working relationship, sometimes called a working alliance, that represents a collaboration between two perspectives on the client's experiences, the client's internal subjective perspective, and the counselor's external "objective" perspective. The collaboration recognizes that the counselor is influenced by the client and also exercises influence on the client; the client tends to talk and act in the manner in which the counselor

listens and responds. Their shared goal is to accelerate movement toward achieving mastery of developmental changes (e.g., expanding action possibilities within the current structures, reorganizing structures). Co-constructing a developmental goal is often a first step in the developmental counseling process. Developmental counseling goals emphasize constructing and not repairing, working toward new client successes in mastering life functions and not recovering the functions—or, in short, promoting healthy development rather than treating psychopathology.

#### **4. GUIDING PRINCIPLES FOR DEVELOPMENTAL COUNSELING PROCESSES**

Three principles guide the counseling processes selected by developmental counselors. A paramount principle is that counselors adapt their processes to fit clients' current developmental stages and structures. For example, developmental counselors who work with children will adapt their manner of communication as well as the content of their language to fit the children's cognitive developmental levels. Because people under extreme stress often regress to forms of communication appropriate for earlier stages (e.g., they cry or scream like children), counselors must be careful to initiate responses that are appropriate matches regardless of clients' chronological ages or current social standing.

Such counselor adaptation is sometimes called the "problem of the match" in developmental counseling or instruction. The principle was probably derived from education (rather than from psychology or medicine), where the phrase "developmentally appropriate practice" is frequently invoked to describe instructor selection of learning climates and pedagogical methods that match the students' cognitive developmental level. Some counselors are very specific in matching cognitive developmental structures (e.g., assessing for client Piagetian stages), whereas others match on the broad dimension of conceptual complexity. Developmental counselors believe that their effectiveness is linked to their ability to adapt counseling processes to match their clients' developmental levels.

A second principle guiding developmental counseling processes is that counselors' stances toward their clients cycle between support and challenge. Counselors begin by supporting their clients' full expression of problem-filled stories. Then, counselors

challenge—without confronting—their clients to move toward actions that promote developmental growth. Later, counselors support their clients' efforts to think or act in new ways that will contribute to growth. Some developmental counselors attempt to challenge their clients at "one-half step" beyond the clients' current positions on a theoretical developmental continuum. The rationale is, in part, to present the next steps in solving the developmental problem as a way in which to encourage client changes.

Third, most developmental counselors adhere to a principle captured by the following phrase: "the symbiotic nature of appraisal and intervention." The principle holds that both processes are ongoing within the counseling process and that each contributes to and profits from the other. For example, appraisal data are the material for counselor or client interpretive interventions, and client responses to interpretations become data integrated into client portraits constructed from appraisal data.

#### **5. APPRAISAL IN DEVELOPMENTAL COUNSELING**

Appraisal in developmental counseling is the conjoint collecting and synthesizing of data about the client's life structure and life cycle. The data are used to construct a descriptive portrait of the client's development and current problems. The assumption is that current developmental problems are best understood when the person's contemporary activity patterns and the sequence of past events are fully examined. Appraisal of a client's life structure involves identifying patterns or configurations of contemporary roles, relationships, and/or activities. For example, adult clients are usually asked to describe the work, family, and community roles that they play, whereas young clients are often asked to describe their family roles and family compositions, sometimes in the form of a family tree. Specifically, the client and counselor review the life structure for incidents of both the support and stressors.

Life cycle appraisal involves describing the sequence of events that the client deems to be critical in his or her development. In addition, the developmental counselor inquires about how the client has experienced developmental tasks within past and current life stages. Life cycle appraisal processes may involve, for example, a life history questionnaire or a life review interview. The client's history is assumed to affect current experiencing.

Therefore, developmental counseling appraisal techniques emphasize the person's assets and coping strategies as well as his or her limitations and barriers.

When making an appraisal, the developmental counselor gives careful attention to the context or surroundings of the client's activity. This attention often yields important observations about person-environment interactions that are critical to understanding the client's development. Questions about social and physical surroundings are frequently included in early discussions with the client. Parents, teachers, and/or friends may also be consulted for independent observations of the client's surroundings. Attention is paid to the hierarchical organization of social systems, especially the differential exercise of authority and power, and to the horizontal structure for opportunities for peer relationships. Specifically, the contextual barriers to action and opportunities for actions are identified.

## 6. DEVELOPMENTAL COUNSELING INTERVENTIONS FOR RESOLVING CURRENT DEVELOPMENTAL PROBLEMS

Interventions are counselor-initiated events that literally "go between" events sequenced along the life span, that is, between the client's current problem-saturated activity and his or her future growth or movement toward resolving problems and addressing new developmental tasks. Developmental interventions usually focus on the client's cognitions or behaviors and use techniques similar to other counseling approaches such as interpretations, teaching, and role-playing.

The developmental counselor frequently constructs interpretations and invites the client to construct interpretations that stimulate new client insights. These interpretations emphasize explanations derived inductively from descriptions rather than derived deductively from classification systems. The developmental counselor may also teach the client a new action sequence that is derived from the counselor's knowledge about the action sequences, called developmental pathways, which are common to people who have mastered a developmental task. Role-playing is another example of how the counselor teaches the client a new pattern of activity. The counselor demonstrates the activity, and the client imitates it, practices it, rehearses it, and finally implements it under real-life conditions. The developmental counselor also encourages the client to engage in

self-directed activities (e.g., self-help books, computer-assisted programs) or to take advantage of learning opportunities outside the counseling sessions.

## 7. DEVELOPMENTAL COUNSELING INTERVENTIONS FOR PREVENTING OR REDUCING DEVELOPMENTAL PROBLEMS

The developmental counselor often engages in interventions to prevent people's problems either from becoming acute or from expanding to interfere with more life functions. Primary prevention occurs when the counselor engages in educating groups of people who are addressing a common developmental task. For example, the counselor may instruct groups of people who are preparing for retirement or, in developmental terms, are preparing to cope with the task of assigning meaning to their lifetimes of productive activity. The instructional goal is to reduce acute problems associated with the loss of work. By learning coping strategies and constructing action plans, these people may avoid persistent regrets or sadness about retiring.

Secondary prevention occurs when the developmental counselor works with a small group of people selected because the members have a common developmental problem. The goal is to prevent the spread or further complication of problematic actions or situations. For example, the developmental counselor may work with young adults who have few friends and are considered to be "shy" in an effort to enrich their current friendships and possibly increase the number of friends or acquaintances, thereby forestalling more serious relationship problems.

### See Also the Following Articles

Lifespan Development and Culture

### Further Reading

- Arbona, C. (2000). The development of academic achievement in school aged children. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed.). New York: John Wiley.
- Blustein, D. L., Juntunen, C. L., & Worthington, R. L. (2000). The school-to-work transition: Adjustment challenges of the forgotten half. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed.). New York: John Wiley.

- Crespi, T. D., & Generali, M. M. (1995). Constructivist developmental theory and therapy: Implications for counseling students. *Adolescence, 30*, 735-744.
- Greenfield, P. M., Keller, H., Fuligni, A., & Maynard, A. (2003). Cultural pathways through universal development. *Annual Review of Psychology, 54*, 461-490.
- Hesketh, B. (2000). Prevention and development in the workplace. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed.). New York: John Wiley.
- Hill, R. D., Thorn, B. L., & Packard, T. (2000). Counseling older adults: Theoretical and empirical issues in prevention and intervention. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed.). New York: John Wiley.
- Howard, G., Nance, D., & Myers, P. (1986). Adaptive counseling and psychotherapy: An integrative eclectic model. *The Counseling Psychologist, 14*, 363-442.
- Ivey, A. E. (1986). *Developmental therapy: Theory into practice*. San Francisco: Jossey-Bass.
- Ivey, A. E., & Goncalves, O. F. (1988). Developmental therapy: Integrating developmental processes into the clinical practice. *Journal of Counseling and Development, 66*, 406-413.
- Juntunen, C. L. (2002). Development, developmental concerns, and counseling. In C. L. Juntunen, & D. R. Atkinson (Eds.). (2002). *Counseling across the lifespan: Prevention and treatment*. Thousand Oaks, CA: Sage.
- Juntunen, C. L., & Atkinson, D. R. (Eds.). (2002). *Counseling across the lifespan: Prevention and treatment*. Thousand Oaks, CA: Sage.
- Vera, E. M., & Reese, L. E. (2000). Preventive interventions with school-age youth. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed.). New York: John Wiley.



# Diagnostic and Statistical Manual of Mental Disorders

Victoria del Barrio

*Universidad Nacional de Educación a Distancia, Madrid, Spain*

1. Antecedents
2. DSM-I (1952)
3. DSM-II (1968)
4. DSM-III (1980)
5. DSM-III-R (1987)
6. DSM-IV (1994)
7. Categorical Development of the DSM  
Further Reading

## GLOSSARY

- axis** A reference line in a coordinate system.
- category** Each division in a system of classification.
- classification** Grouping of elements according to classes.
- code** Number for identification of disorders.
- mental disorder** A clinically behavioral or psychological syndrome associated with stress or disability or with increased risk of suffering death, pain, or an important loss of freedom.
- syndrome** A group of associated disease symptoms.

The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) classification system for mental disorders, developed by the American Psychiatric Association and now employed worldwide, has its historical roots in previous systems dating back over several centuries, from the Greek Hippocrates in the fourth-century BC to the nineteenth-century German psychiatrist Emil

Kraepelin. This article traces the development of the DSM, describing successive changes and improvements over the course of its six editions, as well as its relationship with the other main system, the European-based *International Classification of Diseases*.

## 1. ANTECEDENTS

Classification is both the process and the result of arranging individuals into groups formed on the basis of common characteristics. It is important in any science, but it is usually a difficult task that raises a variety of problems for experts to deal with, including mixture (heterogeneity or homogeneity of a group), discrimination (between-group differences), and identification (assignment of an individual to a group).

Historically, the first attempt to put the variety of psychological disorders into some kind of order was made by the Greek physician Hippocrates (4th century BC); he coined the terms *phrenesis*, *mania*, and *melancholia*, which were maintained by Galen in Rome (1st century AD). Much later, during the Renaissance, Barrough (1583) divided mental disorders into mania, melancholia, and dementia. The work was continued in Europe by the German philosopher Immanuel Kant in his *Anthropologie* and by the Frenchman F. Pinel in his *Nosologie Philosophique* (1789), which divided mental

disorders into mania with and without delirium, melancholia, dementia, and idiocy. The Swede, K. Linnaeus, better known for his classifications of plants and animals, also extended his work to the field of the human mind, employing such categories as *Ideales* (delirium, amentia, mania, melancholia, and vesania), *Imaginnarii* (hypochondria, phobia, somnambulism, and vertigo), and *Patheci* (bulimia polydipsia, satyriasis, and erotomania).

At the end of the 19th century, the German, Emil Kraepelin—often considered the founder of modern scientific psychiatry—in his *Lehrbuch der Psychiatrie* [*Handbook of Psychiatry*] (1899) included a highly influential classification. He aimed to identify groups of patients through symptom clusters that would define different syndromes, and he attributed disorders to four organic roots: heredity, metabolic, endocrine, and brain disease. His categorization laid the bases for systems such as the first official American Psychiatric Association (APA) classification, which included mental disorders by brain traumatism, mental disorders by encephalopathy, mental disorders by drugs, mental disorders by infectious agents, syphilis, arteriosclerosis, epilepsy, schizophrenia, manic-depressive psychosis, psychopathy, psychic reaction, paranoia, oligophrenia, and other cases.

In the 1960s, neo-Kraepelinian psychiatrists called for improvements to the old classification, and men such as Spitzer, Endicott, and Feighner began to prepare the construction of a new American psychiatric classification.

Although the roots of modern American classification are in Europe, the first official classification was a census of mental disorders drawn up in the United States in 1840, with just two categories: idiocy and madness. Another important antecedent of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) was the work of the Committee on Statistics of the American Psychiatric Association (1917), led by Dr. T. Salmon, which recorded mental disease statistics from the psychiatric services of hospitals. Its result was the *Statistical Manual for the Use of Hospitals for Mental Disease*. A few years later, the New York Academy of Medicine organized the National Conference on Disease Nomenclature (1928) to unify the terminology and nosology used by practitioners; its *Standard Classified Nomenclature of Diseases* (1932; rev. ed., 1934) originally referred only to physical diseases, but a section on Diseases of the Psychobiological Unit was included later, and the 1951 edition was published for use in mental hospitals. All of these classifications owed a debt to Kraepelin's system.

World War II brought so many diagnostic problems to military psychiatrists that they urged the construction of a *Classification of Mental Problems*. The psychological nomenclature was introduced in 1944 by the Navy and a year later by the Army. However, the existing *Standard Nomenclature* (1932) proved inadequate for many situations, and the Armed Forces asked for a revision of the three systems in use. The result was the creation of the DSMs.

In Europe, a parallel movement, that of mental hygiene, held its first meeting in Paris in 1932. A European classification would unify psychiatric nomenclature, with the following categories: congenital mental disorders, endogenous psychosis, reaction psychosis, symptomatic psychosis, encephalopathies, epilepsy, toxicomania, and non-alienated mental disorders. However, this proposal was not well accepted by the mental health community.

Today, the two most well-known categorical classifications of mental disorders are the *International Classification of Diseases* (ICD) and the DSM. The ICD, created in Europe, has been promoted by the World Health Organization (WHO) since 1900, but its original version included only physical diseases; the DSM, published by the APA since 1952, has always focused on mental disorders. The two classifications attempt to deal with the same behaviors and coincide to a large extent, but there are some important differences.

The first ICD versions included only physical disorders (Table 1), but the ICD-6 (1948) saw the incorporation of mental diseases, bringing it closer to the DSM-I. The ICD-8 (1967) and DSM-II (1968) also coincided in many aspects, but the APA's decision to develop the DSM-III generated new tensions. Much of this divergence was resolved through better communication, and the DSM-IV and ICD-10 reestablished the close relationship between the WHO and the APA.

## 2. DSM-I (1952)

The DSM was first published in 1952 (DSM-I). Influenced by its European antecedents, and by the views of A. Meyer (a biologically oriented psychiatrist) and K. Menninger (who was psychoanalytically oriented), it took into account not only biological but also social and psychological elements, and it offered a multidimensional consideration of disorders. The process of developing the manual involved the collaboration of the National Institute of Mental Health and different psychiatrists and military professionals.

TABLE I  
Relationship over Time between the DSM and the ICD

Year	DSM (APA), United States	ICD (WHO), Europe
1900		ICD-1 <sup>a</sup>
1910		ICD-2 <sup>a</sup>
1920		ICD-3 <sup>a</sup>
1929		ICD-4 <sup>a</sup>
1938		ICD-5 <sup>a</sup>
1948		ICD-6
1952	DSM-I	
1955		ICD-7
1966		
1967		ICD-8
1968	DSM-II	
1975		ICD-9
1978		ICD-9-CM
1980	DSM-III	
1987	DSM-III-R	
1992		ICD-10
1994	DSM-IV	
1996		ICD-10 Children
1998	DSM-IV-R	
2000	DSM-IV-TR	

<sup>a</sup>ICD with only physical diseases.

Its impact was quite limited. Smith and Fonda, in a study using the *DSM-I* criteria, found interrater agreement to be high for organic psychosis, melancholia, and schizophrenia but quite low in many other categories.

Other studies were also critical. For example, the omission of criteria was criticized by H. J. Eysenck and R. B. Cattell; in their view, common nomenclature in itself would not improve clinical diagnosis, as clinicians would understand different things from the same labels. Furthermore, P. E. Meehl questioned the reliability of clinical judgment, whereas L. Cronbach and Meehl drew attention to the urgent need for accuracy in diagnostic identification.

With the publication of the *ICD-7*, some of the differences between the WHO and APA classifications were highlighted. The *DSM-I* categories were also adopted for a computer program designed for assessment tasks (DIAGNO), but this presented some problems of application. The need to improve the system was evident, and it was Endicott, Guze, Klein, Robins, and Winokur who assumed the task of reviewing and improving the manual.

### 3. DSM-II (1968)

This was published as the *American National Glossary* to the *ICD-8*. An APA committee, with the help of some experts, carried out an in-depth revision of its first version. Meyer's previous influence was drastically reduced, whereas more room was given to psychoanalytical and Kraepelinian ideas in order to improve its acceptance by clinicians. It was based on the medical illness model of different syndromes made up of clinical symptom clusters.

Many unsuccessful attempts were made to coordinate the *DSM* and *ICD* systems, but the differences eventually increased, both in diagnoses and in terminology. For instance, the *DSM-II* had 39 more diagnoses than the *ICD-8 10/3*, and regarding terminology, terms associated with certain theoretical frameworks (such as "schizophrenic reaction" for schizophrenia) were avoided but others, such as "neuroses" or "psychophysiological disorders," remained.

The *DSM-II* has frequently been referred to as "old wine in a new bottle" since the influences of Meyer and Kraepelin were still present but many of the new research perspectives were absent. One year after its publication, Jackson described it as "a hotchpotch of different bases of classification." However, studies on schizophrenia diagnosis in various countries greatly stimulated the task of revision, and efforts were redoubled to clarify basic diagnostic criteria, paving the way for the construction of the *DSM-III*.

### 4. DSM-III (1980)

Criticism of the previous editions had noted the lack of an organizing criterion and the overlapping of categories. Efforts to improve the instrument culminated in the publication of the *DSM-III*.

A task force chaired by R. Spitzer, with international cooperation and the participation of N. Andreasen, J. Endicott, D. F. Klein, M. Kramer, Th. Millon, H. Pinsker, G. Saslow, and R. Woodruff, prepared the work. New members were added subsequently, and a draft was prepared at the important St. Louis conference in 1976. Collaborators were recruited from many institutions, including the Academy of Psychiatry and the Law, the American Academy of Child Psychiatry, the American Academy of Psychoanalysis, the American Association of Chairmen of Departments of Psychiatry, the American College Health Association, the American

Orthopsychiatry Association, and the American Psychological Association. Misunderstandings arose between the American Psychological Association and the American Psychoanalytical Association, but all were eventually solved.

The goals of these teams were (i) to expand the classification, maximizing its utility for the outpatient population; (ii) to differentiate levels of severity and cause within syndromes; (iii) to maintain compatibility with the *ICD-9*; (iv) to establish diagnostic criteria on empirical bases; and (v) to evaluate concerns and criticisms submitted by professional and patient representatives. Subsequently (1977–1979), drafts were sent to experts and practitioners for study and suggestions before the definitive modifications were made. The major characteristics of the *DSM-III* are (i) operational or explicit criteria for assigning patients to diagnostic categories and (ii) the implementation of a multiaxial framework.

#### 4.1. Explicit Diagnostic Operational Criteria

Diagnostic criteria make explicit the signs and symptoms required for a diagnosis on the basis of empirically validated rules, and they are not only descriptive but also phenomenological and behavioral. They would permit greater effectiveness and reliability in diagnoses. Also included for each disorder was its description, the usual age at which it begins, mean duration, prognosis, rates by sex, and risk factors. Criteria for Axes I and II are basically the same as the Feighner Criteria, based only on disorder definitions, as in the *DSM-I* and *DSM-II*.

Such criteria were used when creating the *Research Diagnostic Criteria*, which included 23 disorders (schizophrenia, schizoaffective disorder–manic, schizoaffective disorder–depressed, depressive syndrome superimposed on residual schizophrenia, manic disorder, hypomanic disorder, bipolar with mania, bipolar with hypomania, major depressive disorder, minor depressive disorder, intermittent depressive disorder, panic disorder, generalized anxiety disorder, cyclothymic personality, labile personality, Briquet's disorder, antisocial personality, alcoholism, drug-use disorder, obsessive–compulsive disorder, phobic disorder, unspecified functional psychosis, and other psychiatric disorders).

An operative description of clinical disorders was included, employed in Axes I and II of the *DSM-III* to differentiate between syndromes. Unlike those of the

biological classifications, *DSM-III* criteria are polythetic (patients do not need to have all the characteristics in order to be included in a category) and intentional (patient's characteristics are listed).

#### 4.2. Inclusion of Five Diagnostic Axes

A new feature in the *DSM-III* was the inclusion of five diagnostic axes in order to obtain greater sensitivity and accuracy in diagnoses and treatments and permit the consideration of psychosocial aspects. The idea of such axes was introduced by Essen, Möller, and Wohlfahrt (in Sweden in 1947). They proposed separating the syndromes from all etiological conceptions, creating two principal axes ("phenomenological" and "etiological") plus three others ("temporality," "social functioning," and "others"). In 1969, Rutter *et al.* in the United Kingdom constructed a multiaxial classification system for children (with axes for clinical syndromes, delay of development, mental retardation, medical condition, and psychosocial situation), and other multiaxial systems were proposed in the United Kingdom by Wing in 1970, in Germany by Helmchen in 1975, and in the United States by Strauss. Apart from the *DSM-III*, the multiaxial formula was applied to the *ICD* in 1992.

Axes were assigned as follows: Axis I to clinical syndromes; Axis II (with two sections) to child maturational problems and adult personality disorders; Axis III to physically rooted problems; Axis IV to the intensity and severity of psychological stressors; and Axis V to the patient's level of adaptive functioning in the past year, objectively evaluated. Axes II, IV, and V link the syndromes with environmental determinants and reflect a shift toward more psychological conceptions.

*DSM-III* constitutes an authentic work of description on psychopathology (including differential diagnosis, etiology, treatment, prognosis, and management). The *DSM-III*, although atheoretical in principle, maintained a biological interpretation of the field. Although praised for its achievements, it was also criticized for including inessential characteristics of disorders and for employing too many diagnostic categories (265); psychoanalysts also criticized the lack of an axis that took into account defense mechanisms and ego functions, and they complained that the replacement of the term neuroses by anxiety disorders situated them at a greater distance.

In sum, the *DSM-III* represents a key stage in the evolution of the *DSM* classification. This change was both quantitative (from 104 syndromes to 395, and

from 132 pages to 500) and qualitative. Since 1980, with the adoption of the axes system, the corpus of categories has remained practically constant. Moreover, *DSM-III* made important progress in relation to aspects such as reliability, intracategory coherence and differentiation, validity, and behavioral data. It also favored the construction of new assessment instruments based on its criteria (self-rating questionnaires, semistructured interviews, and so on). Despite all these innovations and improvements, further revision was necessary, and this arrived in the form of the *DSM-III-R*.

### 5. DSM-III-R (1987)

This appeared in 1987 as the final result of revisions carried out by a work group headed by Spitzer. Innovations included the reorganization of categories; some improvements in Axes IV (on psychosocial stress) and V [inclusion of the Global Assessment of Relational Functioning (GARF) scale]; and the incorporation of aspects related to drug abuse, homosexuality (now included in unspecified sexual disorders), and hyperactivity. Classification of affective disorders was reorganized. This version had a dramatic impact worldwide and became more commonly used than the *ICD*, even in Europe.

### 6. DSM-IV (1994)

A task force led by Allen Frances working with experts and scientific groups from all over the world prepared this version. The goal was to improve its cultural sensitivity and to improve compatibility with the *ICD*.

Features of the new instrument included (i) brevity of criteria set, (ii) clarity of language, (iii) explicit statements of its constructs, (iv) it was based on up-to-date empirical data, and (v) better coordination with the *ICD-10*. The *DSM-IV* also presented a series of changes with respect to its predecessor. Categories such as organic mental disorders disappeared, whereas others such as eating disorders, delirium, dementia, and amnesic and other cognitive disorders, and some severe developmental disorders (Rett's syndrome and Asperger's syndrome), were incorporated; the categories of child and sexual disorders were restructured.

In the multiaxial system, the changes occurred in Axis IV, which includes more stress-generating events, and Axis V, in which other scales were added for determining level of adaptation—the Social and

Occupational Functioning Assessment Scale and the GARF scale.

Criticism of the *DSM-IV* has not been lacking, especially with regard to its focus on consensus more than on data. However (and despite claims that there would never be a *DSM-V*), remodeling of the *DSM-IV* is currently in progress.

## 7. CATEGORICAL DEVELOPMENT OF THE DSM

### 7.1. Important Changes

Over time, the *DSM* categories have undergone significant changes. Whereas in 1952 there were two main categories, in 2000 there were 17. This increasing complexity is even more evident in the case of subcategories (Table II).

Each diagnostic category is identified by a numerical code. The current *DSM* codes the different disorders with its own code and also with that of the *ICD*, making it possible to compare and contrast diagnoses made with these two different categorical systems, thus facilitating communication between them.

### 7.2. Critical Positions

The *DSM* classification system has attracted criticism from various theoretical points of view. Those from the antipsychiatry current have always been against the use of classifications in psychiatry since they consider labeling a dangerous procedure. Another notable critic was Eysenck, who spoke of the fundamental weakness of any scheme “based on democratic voting procedures rather than on scientific evidence.”

Further criticisms of the *DSM* refer to (i) the cultural biases of all classification systems; (ii) its extreme individualism—only individual diagnosis is taken into account; (iii) the influence on it of old-fashioned medical classifications, despite modern developments in psychology and psychiatry; and (iv) the “softness” of the categories, which are based more on description than explanation, despite the crucial importance of the latter.

However, despite such criticisms, categorical classifications have made possible comparisons and inferences that have helped to advance clinical knowledge. They may have their imperfections, but they have made an enormous contribution to diagnostic reliability and to understanding among mental health professionals. They have also helped to promote the creation of more

**TABLE II**  
**Evolution of DSM Categories**

<i>DSM-I (1952)</i>	<i>DSM-II (1968)</i>	<i>DSM-III (1980)</i>	<i>DSM-III-R (1987)</i>	<i>DSM-IV (1994)</i>	<i>DSM-IV-TR (2000)</i>
B-Mental deficiency (6 diagnoses)	I-Mental retardation (6 diagnoses)				
Transient and situational personality disorders (8 diagnoses)	IX-Behavioral disorders of childhood and adolescence (7 diagnoses)	Disorders usually first evident in infancy and adolescence (45 diagnoses)	Disorders usually first diagnosed in infancy and adolescence (46 diagnoses)	Disorders usually first diagnosed in infancy and adolescence (46 diagnoses)	Disorders usually first diagnosed in infancy and adolescence (48 diagnoses)
			Delirium, dementia (1)	Delirium, dementia, and amnesic and other cognitive disorders (33)	Delirium, dementia, and amnesic and other cognitive disorders (29)
A-Organogenic disorders 1-Acute brain disorders (13 diagnoses) 2-Chronic brain disorders (25)	II A-Psychoses associated with organic brain syndrome (40 diagnoses) II B-Non psychotic organic brain syndrome	Organic mental disorders (58)	Mental disorders due to a general medical condition not elsewhere classified (61)	Mental disorders due to a general medical condition not elsewhere classified (3)	Mental disorders due to a general medical condition not elsewhere classified (3)
		Substance use disorders (119)	Substance-related disorders (22)	Substance-related disorders (133)	Substance-related disorders (140)
C 1-Psychotic disorders (17)	III-Psychoses not attributed to physical conditions listed previously (24)	Schizophrenic disorders (5)	Schizophrenia disorders (5)	Schizophrenia and other psychotic disorders (16)	Schizophrenia and other psychotic disorders (20)
		Paranoid disorders and other psychotic disorders (8)	Other psychotic disorders (5)		
		Affective disorders (9)	Mood disorders (9)	Mood disorders (16)	Mood disorders (16)
C3-Psychoneurotic disorders	IV-Neuroses (11 diagnoses)	Anxiety disorders (10)	Anxiety disorders (9)	Anxiety disorders (12)	Anxiety disorders (12)

C 2-Psychophysiological autonomic and visceral disorders (10)	VI-Psychophysiological disorders (10)	Somatoform disorders (5)	Somatoform disorders (7)	Somatoform disorders (10)	Somatoform disorders (10)	
		Factitious disorders (3)	Factitious disorders (3)	Factitious disorders (2)	Factitious disorders (2)	
		Dissociative disorders (5)	Dissociative disorders (5)	Dissociative disorders (5)	Dissociative disorders (5)	
		Psychosexual disorders (22)	Sexual disorders (20)	Sexual and gender identity disorders (31)	Sexual and gender identity disorders (31)	
	VII-Special symptoms (10)			Eating disorders (3)	Eating disorders (3)	
		Disorders of impulse control not elsewhere classified (6)	Sleep disorders (12)	Sleep disorders (17)	Sleep disorders (17)	
			Impulse-control disorders (6)	Impulse-control disorders not elsewhere classified (6)	Impulse-control disorders (6)	
		Psychological factors affecting physical condition (1)				
	C 4-Personality disorders (18)	VIII-Transient situational disturbances (5)	Adjustment disorders (8)	Adjustment disorders (9)	Adjustment disorders (6)	Adjustment disorders (6)
		V-Personality disorders (33)	Personality disorders (12)	Personality disorders (12)	Personality disorders (11)	Personality disorders (11)
X-Condition without manifest psychiatric disorders and nonspecific conditions		V-Codes for conditions not attributable to mental disorders that are a focus of attention or treatment	Other conditions that may be focus in clinical attention (9)	Other conditions that may be focus in clinical attention (32)	Other conditions that may be focus in clinical attention (36)	
XI-Non diagnostic term for administrative use			Other psychological factors affecting physical condition (1)			
		Additional codes	Additional codes	Additional codes	Additional codes	
Total no. of syndromes 97	146	316	242	382	390	

homogeneous and accurate assessment instruments that will constitute the source of future progress in the field.

### See Also the Following Articles

Depression in Late Life ■ Forensic Mental Health Assessment  
 ■ Mental Measurement and Culture

### Further Reading

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 529, 281–203.
- Eron, L. D. (1966). *Classification of behavior disorders*. Chicago: Aldine–Atherton.
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Muñoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26, 57–63.
- Haynes, S. N., & O'Brien, W. H. (1988). The Gordian knot of *DSM-III* use: Interacting principles of behavioral classification and complex causal models. *Behavioral Assessment*, 10, 95–105.
- Millon, Th. (1996). The *DSM-III*: Some historical and substantive reflections. In Th. Millon (Ed.), *Personality and psychopathology*. New York: Wiley.
- Schmith, H. O., & Fonda, C. (1952). The reliability of psychiatry diagnosis. *Journal of Abnormal and Social Psychology*, 52, 262–267.
- Spitzer, R. L., Endicott, J., & Robins, E. (1975). *Research diagnostic criteria (RDC)*. New York: New York State Psychiatric Institute.
- Spitzer, R. L., Gibbson, M., Skoldol, A. E., Williams, J., & First, M. B. (1994). *DSM-IV Casebook: A learning comparison to the Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Press.
- Spitzer, R. L., & Willson, P. T. (1975). Nosology and the official psychiatry nomenclature. In A. M. Freeman, H. I. Kaplan, & B. J. Sadock (Eds.), *Compressive textbook of psychiatry* (Vol. 2). Baltimore: Williams & Wilkins.
- Szasz, T. (1966). The psychiatry classification of behavior: A strategy of personal constraint. In L. D. Eron (Ed.), *The classification of behavior disorders* (pp. 123–170). Chicago: Aldine.
- Widiger, T. A., Frances, A. J., Pincus, H. A., First, M. B., Ross, R., & Davis, W. (1994). *DSM-IV sourcebook*. Washington, DC: American Psychiatric Association.



# Diverse Cultures, Dealing with Children and Families from

Antoinette Halsell Miranda

Ohio State University, Columbus, Ohio, USA

1. Introduction
  2. Racial and Ethnic Identity
  3. Understanding Culture as a Continuum
  4. Cross-Cultural Competence
  5. Cross-Cultural Communication
  6. Diversity in Education
  7. Implications for Practice
- Further Reading

*race* A social construct that attempts to divide human groups based on physical characteristics.

*racism* A belief that people can be validly grouped on the basis of biological traits and that some groups have superior traits, whereas others have inferior ones.

*structural assimilation* A process that occurs when one group can access and experience full participation within majority group organizations such as country clubs and private clubs.

## GLOSSARY

*cultural assimilation* A process that occurs when one group assumes the values, behaviors, and beliefs of another group, often the majority culture.

*culture* The beliefs, values, traditions, behavior patterns, and social and political relationships that are shared by a group of people who are bound together by things such as common history, language, social class, and religion.

*discrimination* Often based on prejudice, a process consisting of differential treatment directed toward certain groups.

*ethnicity* An individual's sense of identification that provides him or her with a sense of belonging to a group.

*power* An important concept when entering into a discussion about diversity given that the struggle for power, and the dominant group having power, has played an integral part in the continued oppression of marginalized groups.

*prejudice* Attitudes and beliefs toward a particular group that are often negative and formed in disregard of the facts.

Working with culturally different children and families requires psychologists to have an understanding of cultural differences and, in the best of circumstances, to become cross-culturally competent.

## 1. INTRODUCTION

It is fairly well known that there has been a dramatic demographic shift in the United States population during the past decade or so. Based on the 2000 census, people of color make up approximately 29% of the U.S. population. By the year 2050, they are expected to account for 47% of the U.S. population. The demographic changes have major implications for many institutions, including schools and the workforce. The change in the ethnic composition of the population is viewed as both an opportunity and a challenge. There is the opportunity

to enrich society with multiple and diverse cultural elements. The challenge is to find effective ways in which to incorporate and respect the multiple diversities that exist while at the same time recognizing that there is an American culture. Historically, White ethnic groups have been “forced” to abandon their cultures and blend into the American culture. Ethnic minorities, on the other hand, have not been so successful. The reason why has not been for lack of trying; rather, they have not fared as well due to the racism and prejudice that exists in the United States. Ethnic minorities are often received and treated as marginalized groups. As marginalized groups, they have not always had full access to the “American dream.” Although the civil rights movement and subsequent civil rights laws in the United States improved opportunities in education and employment for many ethnic minorities and women, people continue to be marginalized and discriminated against.

While many ethnic minorities have been successful in culturally assimilating, they still struggle to experience full structural assimilation in the United States. There has been a growing acceptance and recognition that, as a society, there has not been an appreciation of the rich diversity that exists; instead, there has been an attitude that one must fit in or assimilate to be successful. Exploring how variables such as class, race, ethnicity, and gender interact and intersect to influence behavior has been critical to researchers understanding life in America for marginalized groups. The United States is a pluralistic nation made up of many different ethnic groups whose members desire to retain their cultural identities. Understanding these different cultural identities is beneficial in working effectively with diverse populations. Identity development research has contributed to this knowledge base.

## **2. RACIAL AND ETHNIC IDENTITY**

Identity is complex and influenced by a number of factors. Erickson was instrumental in helping people to understand that individual identity is embedded in social, cultural, and historical contexts. Race and ethnic identity adds to this complexity. The development of racial and ethnic identity for people of color often begins to occur during late adolescence. It is generally this period of time when adolescents start to explore the questions of “Who am I?” and “Who will I become?” Searching for racial and ethnic identity can be a lifelong journey. A focus on certain aspects of people’s identities may vary at different times in their lives, depending on

the circumstances. For example, high school is a time when most adolescents begin to start to develop their identities separate from their families as they self-reflect on who they are. Individuals are faced with attempting to integrate these different components of identity to form who they are. Many aspects of society influence how people view themselves and, consequently, how they develop their identities.

Cross and Helms have researched Black identity and White identity, respectively. They have demonstrated that racial and ethnic identity develops in stages. Most important, these stages are fluid, with people moving back and forth between stages. It should also be noted that not all people move through these stages. Black adolescents and other adolescents of color tend to begin to develop racial identities earlier than do White adolescents because their self-perceptions can be influenced by how society views them. The aspect of people’s identity that others notice and make them most aware of is one that can potentially loom large in their lives. Thus, adolescents of color are subjected to questions and stereotypes about their race and/or ethnicity more often than are White adolescents. Even though some adolescents begin to explore their racial and ethnic identities, research has shown that the movement through the stages most often occurs during late adolescence and early adulthood. For Whites, identity development involves the abandonment of entitlement or White privilege. The key challenge for people of color is the development of a positive identity. Research continues on the development of racial and ethnic identity.

## **3. UNDERSTANDING CULTURE AS A CONTINUUM**

When working with culturally diverse children and families, it is important to understand that culture occurs on a continuum and is influenced by the process of dual socialization, which is the process by which individuals learn how to function in two distinct socio-cultural environments. Most ethnic minorities in the United States are able to function in their own cultures but also learn how to navigate the waters of the majority culture. Individuals who navigate these two different cultures successfully are said to be bicultural. This concept is important and is related to assimilation. In the United States, the dominant group controls most of the social, economic, and political institutions. Thus, the dominant group determines these institutions’

norms, values, and beliefs, which are often referred to as culture capital. Having knowledge of culture capital allows easier movement through mainstream culture. Members of ethnic minority groups often must acquire the necessary cultural traits of the majority group to move up the social and economic ladders. It is often necessary for working-class people and people of color to learn the cultural norms, or to be told explicitly the rules of the culture of the dominant group, for them to acquire power.

However, when working with culturally different people, it is important to remember that individuals will vary along the cultural continuum. Individuals may choose to remain isolated from the mainstream, on one end of the continuum, or choose to relate to both cultures, never quite integrating them into one. Many ethnic minorities find themselves straddling two cultures. The challenge for these individuals is to develop strong self-concepts and positive self-esteem without compromising who they are. Understanding issues of identity and cultural continuum provides a better understanding of culturally different people. Thus, with the acknowledgment that diversity exists, a body of knowledge that encourages people to learn to respect, value, and understand diverse groups has evolved. This has occurred primarily through educational channels such as ensuring that disciplines require a diversity course for graduation and the development of multicultural competencies.

#### **4. CROSS-CULTURAL COMPETENCE**

As with so many disciplines, the demographics of education and psychology do not match those of the U.S. population. In psychology and education literature, there is increasing emphasis placed on training individuals to become cross-culturally competent. Cross-cultural competence means that individuals possess attitudes, abilities, and understandings that allow them to function not only in American culture but also within and across various ethnic, cultural, and linguistic groups. Three common elements of cross-cultural competence are personal awareness, knowledge, and skills. Although different disciplines may advocate diverse ways of attaining cross-cultural competence, these three elements appear to be universal aspects of that process. In education and psychology, having cross-culturally competent individuals is seen as critical to being able to deliver services that are sensitive and effective in meeting the needs of culturally different individuals. Developing cross-cultural competence is a

process. It is generally recognized that becoming cross-culturally competent will not occur by simply studying groups of people. It is necessary to have personal experiences with culturally different groups as well.

Self-awareness consists of individuals engaging in self-exploration that requires them to examine their own culture and socialization process that influences how they view and interact with majority and minority people. Understanding their own culture and its influence on their socialization is the first step in awareness. To understand another culture, individuals must first explore their own culture and its impact on their development. This often requires them to become aware of their biases and prejudices. Knowledge is the second component of cross-cultural competence. This is usually accomplished by studying the concept of diversity as well as racial and ethnic groups. This can be accomplished in a number of different ways, including reading books, talking with individuals from the cultural groups, and interacting with culturally different people. Skills are the final component. Skills involve the ability to work effectively with culturally diverse populations, understanding the influence that issues of diversity may have on people as well as one's ability to interact effectively with them.

Cross-cultural competence also means that individuals are aware of their own cultural limitations. Learning about other cultures requires them to take risks and have a willingness to be open, appreciative, and respectful of cultural differences. Intercultural interactions should be viewed as learning opportunities, recognizing that there is integrity and value in every culture one encounters.

#### **5. CROSS-CULTURAL COMMUNICATION**

One of the most difficult tasks that people face is communicating across differences. There is always the fear that what people intend to say will not be interpreted the way in which they meant it to be interpreted. It is generally recognized that there are high-context cultures and low-context cultures, and there are differences in how individuals from these cultures engage in communication. High-context cultures tend to communicate with fewer words and less emphasis placed on verbal interactions. As a result, individuals from these cultures are better at reading nonverbal cues. Situational and nonverbal cues convey primary meaning. For example, it is not unusual to hear an African American say, "All

my mother needed to give us was that “look.” She didn’t have to say anything. We knew we were in trouble.” Individuals from high-context cultures also place a great deal of emphasis on establishing trust first. Thus, before revealing personal information, individuals from these cultures will want to first establish a relationship. Examples of high-context cultures include Asians, Native Americans, Arabs, Latinos, and African Americans. Low-context cultures, in contrast, tend to exhibit precise and direct verbal communication. They often do not notice, or are not attuned to, nonverbal communication. Meaning in low-context cultures is conveyed primarily through written and spoken words. Communication is often linear. Individuals from these cultures become particularly frustrated with high-context cultures that communicate in a circular fashion. Content of communication in low-context cultures can be best described as “cutting to the chase.” Examples of low-context cultures include European Americans, Swiss, and Germans.

Cultures also differ in other aspects of communication, including proximity and touching. Groups differ in the amount of social distance they tolerate in social situations. For example, European Americans tend to keep a distance of 3 feet when talking to people with whom they are unfamiliar. Latinos and African Americans tend to feel comfortable with less social distance between themselves and others. Touching also varies greatly between groups. For example, handshaking takes on different meanings across cultures. Americans use handshaking as an introduction, whereas other cultures may be reluctant to engage in handshaking.

It is important to recognize that communication can be culturally loaded. Understanding communication differences across cultures and learning to adjust one’s communication style to minimize misinterpretation is the key. Effective cross-cultural communication consists of individuals being willing to learn and understand cultural differences.

## **6. DIVERSITY IN EDUCATION**

Schools are reflective of the larger society in its diversity. The changing face of society and schools was predicted as early as the 1980s. Much of the past 20 years has focused on how to best serve an increasingly diverse population in the schools. An increasing achievement gap between White students and students of color continues to be problematic. Overrepresentation of certain ethnic minority groups in special education categories

continues despite efforts to remedy the situation with a requirement for nondiscriminatory assessment as part of a multifaceted evaluation for special education placement. Overrepresentation of African American males for suspensions and expulsions has been well documented. Students of color are overwhelmingly tracked in lower level classes, are underrepresented in gifted and talented classes, have lower expectations placed on them by educational personnel, and have access to fewer resources than do their more affluent peers. Finally, the limited English-proficient (LEP) population is increasing. All of these issues, as well as others, make it imperative that clinicians have a solid understanding of issues of diversity and their influence and impact in educational settings.

Having cross-culturally competent personnel in education is a must. Understanding the lives of children and their families enables the provision of services that will make positive changes in their academic and personal lives. It is well known that involved parents are important to children’s education. Unfortunately, parents from lower socioeconomic status and minorities have a lower level of contact with the educational systems. The challenge for professionals is to find opportunities to bridge the cultural gap between themselves and parents. Lack of contact does not equate with lack of interest. Epstein has engaged in considerable research that has been informative regarding best practice in how to work with parents. Like other areas of diversity, issues around the distribution of power, resources, and knowledge may prohibit meaningful engagement of parents in the schools. Understanding the influence of the parents’ experiences with schooling may also help to explain their reluctance to become involved. This is especially important with immigrant children whose parents’ ideas of involvement may mirror what they experienced in their native lands and may be inconsistent with beliefs in the United States. In addition, a large percentage of parents in the inner city are young parents who never finished school, or were unsuccessful in school, and who view schools as negative places. These parents often did not have good role models for participation and may be struggling to understand the need for parent involvement in their children’s schooling.

Finding effective ways in which to connect with parents and help them to become full partners in their children’s education and mental health development can be far-reaching. To do that, psychologists need to understand the social and cultural differences that may exist between the home and the school. This understanding alone could prevent miscommunication

between parents and the school. One goal should be to establish two-way communication between the school and families. It could be quite enlightening to know parents' perceptions of the school and how they believe they are (or are not) welcomed. Clinicians are often trained to "cut to the chase" and attack the "problem" when, in fact, what may be most important is to build rapport and the relationship with parents. Instead of telling parents what they should be doing, clinicians could instead ask them, "How can we assist you?" Parents who feel truly invited into the school, have a sense of self-efficacy, and perceive that they play a critical role in their children's educations are more likely to be involved in their children's schooling. It is important to acknowledge parents as important contributors to the lives of their children.

As clinicians seek to close the achievement gap between Whites and ethnic minorities and to improve the overall educational experience, they must also understand and acknowledge conditions that may be barriers to learning. Poverty is one such variable that has far-reaching implications. Children in poverty (of which ethnic minorities are disproportionately represented) experience more problems with physical health, lower cognitive ability, poor school achievement, and the like. Despite the odds against these children being successful, there is evidence that many of these children actually do succeed. Research on resiliency identifies many factors, which are termed "protective," that enable children to beat the odds. It is important for clinicians to be familiar with this literature so that they are skilled at identifying and promoting those protective factors that are present in children's lives. Research has identified the following as protective factors: social skills (e.g., empathy, communication), problem-solving skills, a sense of control, a sense of purpose, caring relationships with adults, high expectations for doing well, and opportunities for meaningful participation in clubs, teams, and/or organizations. For many culturally diverse children and families who are also in poverty, having available resources that include human resources can make an incredible difference in their lives.

## 7. IMPLICATIONS FOR PRACTICE

Most disciplines acknowledge the importance of having their personnel become cross-culturally competent. Although nearly everyone in the United States shares the American culture at some level, there are many who

also have subcultures that are very much a part of their lives. Recognizing the similarities as well as the differences helps to bring clinicians one step closer to providing effective services to a culturally different population. Although it may seem like a daunting task, especially if clinicians are in an area with many different cultural groups, there are some ways in which clinicians can increase their skills in being effective psychologists with culturally different groups. First, they should be open to new ideas and experiences. Too often, clinicians look at the world through their own lenses. They should take time to learn about other groups and attempt to understand the world from their perspective. This does not mean that clinicians need to agree with them so long as they at least understand these other groups' perspectives. Second, clinicians should find a cultural mediator who can guide their understanding about the beliefs and practices of a particular culture with which they may be unfamiliar. This is often done through open discussions and interpersonal sharing. Clinicians should always remember that even though they do not represent everyone from that culture, members of that culture can provide them with some insight. Third, clinicians should see diversity as something to be celebrated rather than as something that is a deficit. Fourth, clinicians should view their learning experience as a journey. There is so much to know. They will not "get it" in one course. Instead, they will continually explore, learn, and explore some more. Fifth, clinicians should explore their own cultural or ethnic identities and examine their socialization with respect to majority and minority people. Until they can understand the impact of their culture and socialization on their attitudes, beliefs, and behaviors, clinicians cannot fully appreciate the cultures of others. Finally, clinicians should always remember to strive to engage in culturally responsive practices that will ultimately benefit clients in a positive way.

The United States is a pluralistic country that is rich in diversity. This is not a phenomenon but rather a reality. Understanding the cultural differences in this country, and how they may influence behavior and aspects of society, is critical. Clinicians should embrace diversity and view it as an opportunity rather than as a challenge.

### See Also the Following Articles

Acculturation ■ Cross-Cultural Psychology, Overview ■ Health Psychology, Cross-Cultural ■ Prejudice and Discrimination ■ Racial and Ethnic Minorities, Counseling of ■ Training, Cross-Cultural ■ Translation

**Further Reading**

- Banks, J. A. (2002). *Teaching strategies for ethnic studies* (7th ed.). Boston: Allyn & Bacon.
- Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York: New Press.
- Friere, P. (1970). *Pedagogy of the oppressed*. New York: Seabury Press.
- Hooks, B. (1994). *Teaching to transgress: Education as the practice of freedom*. New York: Routledge.
- Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York: HarperCollins.
- Ladson-Billings, G. (2001). *Crossing over to Canaan*. San Francisco: Jossey-Bass.
- McIntosh, P. (1989). *White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies*. Working Paper 189, Center for Research on Women, Wellesley College.
- Nieto, S. (1999). *Affirming diversity: The sociopolitical context of multicultural education* (3rd ed.). New York: Longman.
- Tatum, B. D. (2003). *Why are all the Black kids sitting together in the cafeteria?* New York: Basic Books.



# Downsizing and Outplacement

Wayne F. Cascio

University of Colorado, Denver, Colorado, USA

1. Downsizing Defined
2. Mergers, Acquisitions, and the Demise of Loyalty
3. Corporate Financial Performance before and after Downsizing
4. Effects of Downsizing on Knowledge-Based Organizations
5. The Psychological Impact of Downsizing on Victims and Survivors
6. Outplacement: Easing the Career Transition Process  
Further Reading

## GLOSSARY

**burnout** A gradual process of loss during which the mismatch between the needs of the person and the demands of the job grows ever greater.

**distributive justice** The fairness of the outcomes of decisions, for example, in the allocation of bonuses or merit pay or in making decisions about who goes and who stays in a layoff situation.

**downsizing** Planned changes in a firm's organizational structure that affect its use of people; in a broader sense, downsizing refers to reductions in a firm's use of assets (e.g., financial, physical, human, information), so when referring to reductions in the numbers of people in an organization, it is appropriate to use the term "employment downsizing."

**knowledge-based organizations** Collections of networks in which interrelationships among individuals (i.e., social networks) generate learning and knowledge.

**organizational citizenship behaviors** Discretionary behaviors performed outside of one's formal role that help other employees to perform their jobs or that show support for and conscientiousness toward the organization.

**organizational decline** An environmental or organizational phenomenon that occurs involuntarily and results in erosion of an organization's resource base.

**organizational support** The extent to which an organization values employees' general contributions and cares for their well-being.

**outplacement** As an extension of the termination process that typically includes two elements: (a) counseling employees who have lost their jobs for emotional stress resulting from the trauma of termination, and (b) assistance with job search.

**procedural justice** The fairness of the procedures used to make decisions; procedures are fair to the extent that they are consistent across persons and over time, free from bias, based on accurate information, correctable, and based on prevailing moral and ethical standards.

Employment downsizing is synonymous with layoffs, but it differs from organizational decline. Its extent and scope are global, and it is an ongoing feature of many organizations. Downsizing has led to a redefinition of the concept of loyalty and has led organizations to think in terms of 3- to 5-year relationships with employees. Evidence indicates that downsizing does not lead to better long-term performance of organizations, but that it may have negative effects on victims and survivors, particularly if they perceive procedural injustice in the process. Outplacement, which provides counseling for those who have lost jobs as well as assistance in job search, can help to speed the reemployment and readjustment process.

## **1. DOWNSIZING DEFINED**

In everyday conversation, the term “downsizing” is often used as a synonym for “layoffs” of employees from their jobs. Downsizing is commonly the result of a broader process of organizational restructuring that refers to planned changes in a firm’s organizational structure that affect its use of people. Such restructuring often results in workforce reductions that may be accomplished through mechanisms such as attrition, early retirements, voluntary severance agreements, and layoffs. Layoffs are a form of downsizing, but it is important to note that downsizing is a broad term that may include any number of combinations of reductions in a firm’s use of assets—financial, physical, human, and information assets. Therefore, when referring to reductions in the numbers of people in an organization, it is appropriate to use the term “employment downsizing.”

### **1.1. Downsizing and Organizational Decline**

Employment downsizing is not the same thing as organizational decline. Downsizing is an intentional, proactive management strategy, whereas decline is an environmental or organizational phenomenon that occurs involuntarily and results in erosion of an organization’s resource base. For example, the advent of digital photography, disposable cameras, and other imaging products signaled a steep decline in the demand for the kind of instant photographic cameras and films that Polaroid had pioneered during the 1940s. On October 12, 2001, Polaroid was forced to declare bankruptcy.

### **1.2. Extent and Scope of Employment Downsizing**

According to the U.S. Department of Labor, approximately 2 million people have been affected by employment downsizing each year since 1996. In 2001, for example, companies announced nearly 1 million job cuts during the 3 months after the terrorist attacks of September 11. Many firms conduct multiple rounds of employment downsizing during the same year as ongoing staff reductions become etched into the corporate culture. On average, two-thirds of firms that lay off employees during a given year do so again the following year.

The phenomenon of layoffs is not limited to the United States. They occur often in Asia, Australia, and Europe as well. Japan’s chip and electronics

conglomerates shed tens of thousands of jobs during the first few years of the 21st century as the worldwide information technology slump and fierce competition from foreign rivals battered their bottom lines. In Mainland China, more than 25.5 million people were laid off from state-owned firms between 1998 and 2001. Another 20 million were expected to be laid off from traditional state-owned firms by 2006.

The incidence of layoffs varies among countries in Western Europe. Labor laws in countries such as Italy, France, Germany, and Spain make it difficult and expensive to dismiss workers. For example, in Germany, all “redundancies” must by law be negotiated in detail by a workers’ council, which is a compulsory part of any big German company and often has a say in which workers can be fired. Moreover, setting the terms of severance is tricky because the law is vague and German courts often award compensation if workers claim that they received inadequate settlements. In France, layoffs are rare. Even if companies offer generous severance settlements to French workers, as both Michelin and Marks & Spencer did, the very announcement of layoffs triggers a political firestorm.

## **2. MERGERS, ACQUISITIONS, AND THE DEMISE OF LOYALTY**

Worldwide, tens of thousands of mergers and acquisitions take place each year among both large and small companies. In general, after a buyout, the merged company eliminates staff duplications and unprofitable divisions. Employment downsizing is part of that restructuring process. Such restructuring often leads to similar effects—diminished loyalty from employees. Following takeovers, mergers, acquisitions, and downsizings, thousands of workers have discovered that years of service mean little to struggling managers or a new corporate parent. This leads to a rise in stress and a decrease in satisfaction, commitment, intention to stay, and perceptions of an organization’s trustworthiness, honesty, and caring about its employees.

Companies counter that today’s competitive business environment makes it difficult to protect workers. Understandably, organizations are streamlining to become more competitive by cutting labor costs and to become more flexible in their response to the demands of the marketplace. But the rising disaffection of workers at all levels has profound implications for employers. Perhaps one of the most fundamental issues to consider

is the new meaning of loyalty. Today, workers and managers are less loyal to their organizations than ever before. As Reichheld has shown, U.S. companies on average now lose half of their employees in 4 years, half of their customers in 5 years, and half of their investors in fewer than 12 months.

So, what is the new meaning of loyalty? Employees at all levels are still loyal, but instead of being blindly loyal to their organizations, they are loyal to a vision or to a mission. They are loyal to mentors, protégés, or team members. As a result, more and more organizations are now thinking in terms of 3- to 5-year employment relationships. Some are even adopting a strategy long used in the sports world, that is, fixed contracts with options for renegotiation and extension.

### **3. CORPORATE FINANCIAL PERFORMANCE BEFORE AND AFTER DOWNSIZING**

In a series of studies that included data from 1982–1994, 1995–2000, and 1982–2000, Cascio and colleagues examined financial and employment data from companies in the Standard & Poor's (S&P) 500. The S&P 500 is one of the most widely used benchmarks of the performance of U.S. equities. It represents leading companies in leading industries and consists of 500 stocks chosen for their market size, liquidity, and industry-group representation. Cascio and colleagues' purpose was to examine the relationships between changes in employment and financial performance. Each year, they assigned companies to one of seven mutually exclusive categories based on each company's level of change in employment and its level of change in plant and equipment (assets). Then, they observed the companies' financial performance (profitability and total return on common stock) from 1 year before to 2 years after the employment change events. They examined results for firms in each category on an independent basis as well as on an industry-adjusted basis.

In Cascio and colleagues' most recent study, they observed a total of 6418 occurrences of changes in employment for S&P 500 companies over the 18-year period from 1982 to 2000. As in their earlier studies, they found no significant, consistent evidence that employment downsizing led to improved financial performance, as measured by return on assets or industry-adjusted return on assets. Downsizing strategies—either employment downsizing or asset downsizing—did not yield long-term payoffs that were significantly

larger than those generated by “stable employers,” that is, those companies in which the complement of employees did not fluctuate by more than  $\pm 5\%$ .

Thus, it was not possible for firms to “save” or “shrink” their way to prosperity. Rather, it was only by growing their businesses that firms outperformed stable employers as well as their own industries in terms of profitability and total returns on common stock.

#### **3.1. Is Downsizing Always Wrong?**

Sometimes, downsizing is necessary. In fact, many firms have downsized and restructured successfully to improve their productivity. They have done so by using layoffs as part of a broader business plan. In the aggregate, the productivity and competitiveness of many firms have increased during recent years. However, the lesson from the research described previously is that firms cannot simply assume that layoffs are a “quick fix” that will necessarily lead to productivity improvements and increased financial performance. The fact is that layoffs alone will not fix a business strategy that is fundamentally flawed.

Employment downsizing might not necessarily generate the benefits sought by management. Managers must be very cautious in implementing a strategy that can impose such traumatic costs on employees—on those who leave as well as on those who stay. Management needs to be sure about the sources of future savings and carefully weigh those against all of the costs, including the increased costs associated with subsequent employment expansions when economic conditions improve. [Table I](#) shows direct and indirect costs of downsizing.

### **4. EFFECTS OF DOWNSIZING ON KNOWLEDGE-BASED ORGANIZATIONS**

Knowledge-based organizations, from high-technology firms to the financial services industry, depend heavily on their employees—their stock of human capital—to innovate and grow. Knowledge-based organizations are collections of networks in which interrelationships among individuals (i.e., social networks) generate learning and knowledge. This knowledge base constitutes a firm's “memory.” Because a single individual has multiple relationships in such an organization, indiscriminate, nonselective downsizing has the potential to inflict considerable damage on the learning and

**TABLE I**  
**Direct and Indirect Costs of Employment Downsizing**

<i>Direct costs</i>	<i>Indirect costs</i>
Severance pay in lieu of notice	Recruiting and employment costs of new hires
Accrued vacation and sick pay	Low morale; risk-averse survivors
Supplemental unemployment benefits	Increase in unemployment tax rate
Outplacement	Lack of staff when economy rebounds; training and retraining
Pension and benefit payouts	Potential lawsuits from aggrieved employees
Administrative processing costs	Heightened insecurity; reduced productivity
Costs of rehiring former employees	Loss of institutional memory and trust in management

memory capacity of the organization. Empirical evidence indicates that the damage is far greater than might be implied by a simple tally of individuals.

When one considers the multiple relationships generated by one individual, it is clear that downsizing that involves significant reductions in employees creates the loss of significant “chunks” of organizational memory. Such a loss damages ongoing processes and operations, forfeits current contacts, and may lead to forgone business opportunities. Organizations that are at greatest risk include those that operate in rapidly evolving industries, such as biotechnology, pharmaceuticals, and software, where survival depends on a firm’s ability to innovate constantly.

## 5. THE PSYCHOLOGICAL IMPACT OF DOWNSIZING ON VICTIMS AND SURVIVORS

Downsizing exacts a devastating toll on workers and communities. Lives are shattered, people become bitter and angry, and the added emotional and financial pressure can create family problems. “Survivors,” or workers who remain on the job, can be left without loyalty or motivation. A common experience is that employee morale is the first casualty in downsizing. Workplaces tend to be more stressful, political, and cutthroat than before the downsizing. Local economies

and services (e.g., human services agencies, charitable organizations) become strained under the impact on the community. Survivors often experience “burnout,” that is, a gradual process of loss during which the mismatch between the needs of the employees and the demands of the job grows ever greater. Research indicates that each person expresses burnout in a unique way but that the basic themes are the same:

- *An erosion of engagement with the job.* What started out as important, meaningful, and fascinating work becomes unpleasant, unfulfilling, and meaningless.
- *An erosion of emotions.* The positive feelings of enthusiasm, dedication, security, and enjoyment fade away and are replaced by anger, anxiety, and depression.
- *A problem of fit between the person and the job.* Individuals see this imbalance as a personal crisis, but it is really the workplace that is in trouble.

Violence and sabotage are possible among victims of the downsizing. In fact, the most common precipitator of workplace violence is a layoff or firing. Sabotage, especially cyber-sabotage, has emerged as a new threat among disgruntled ex-employees. Recently terminated workers have posted a company’s payroll on its intranet, planted data-destroying bugs, and handed over valuable intellectual property to competitors. Although exact numbers are hard to ascertain, computer security experts say that this is fast becoming the top technical concern at many companies.

### 5.1. Roles of Procedural and Distributive Justice in Employment Downsizing

Procedural justice focuses on the fairness of the procedures used to make decisions. Procedures are fair to the extent that they are consistent across persons and over time, free from bias, based on accurate information, correctable, and based on prevailing moral and ethical standards.

Distributive justice focuses on the fairness of the outcomes of decisions, for example, in allocating bonuses or merit pay or in making decisions about who goes and who stays in a layoff situation. In simple terms, it is the belief that all employees should “get what they deserve.”

In the wake of decisions that affect employees, such as those involving pay, promotions, and layoffs, employees often ask, “Was that fair?” Judgments about the fairness or equity of procedures used to make decisions (i.e.,

procedural justice) are rooted in the perceptions of employees. Strong research evidence indicates that such perceptions lead to important consequences such as employee behavior and attitudes. As noted previously, when employees believe that they have not been treated fairly, they may retaliate in the form of theft, sabotage, or even violence. In short, the judgments of employees about procedural justice matter.

Procedurally fair treatment has been demonstrated to result in reduced stress and increased performance, job satisfaction, commitment to an organization, and trust. It also encourages organizational citizenship behaviors, that is, discretionary behaviors performed outside of one's formal role that help other employees to perform their jobs or that show support for and conscientiousness toward the organization. These include behaviors such as the following:

- Volunteering to carry out activities that are not formally a part of one's job
- Persisting with extra enthusiasm or effort when necessary to complete one's own tasks successfully
- Helping and cooperating with others
- Following organizational rules and procedures, even when they are personally inconvenient
- Endorsing, supporting, and defending organizational objectives

Procedural justice affects citizenship behaviors by influencing employees' perceptions of organizational support, that is, the extent to which the organization values employees' general contributions and cares for their well-being. In turn, this prompts employees to reciprocate with organizational citizenship behaviors. These effects have been demonstrated to occur at the level of the work group as well as at the level of the individual. In general, perceptions of procedural justice are most relevant and important to employees during times of significant organizational change. When employees experience change, their perceptions of fairness become especially potent factors that determine their attitudes and behaviors. Employment downsizing is a major change for people, whether victims or survivors. Thus, considerations of procedural justice will always be relevant.

## 6. OUTPLACEMENT: EASING THE CAREER TRANSITION PROCESS

Outplacement is an extension of the termination process. It typically includes two elements: (a) counseling

the employee who has lost his or her job for emotional stress resulting from the trauma of termination, and (b) assisting with job search. Its premise is that the sooner the terminated employee is reemployed, the less time he or she has to become disgruntled, to file a lawsuit, and/or to cause problems for those who remain, particularly through violence or sabotage.

Employment downsizing often creates anxiety or resentment among remaining employees. Some workers might feel that the departing employee received unjust treatment. Others may doubt their own job security and resent the upheaval and additional workload caused by a coworker's departure.

Outplacement can help to address these issues. It provides firing managers with training and procedures to minimize trauma to the workers affected and disruption within the affected department. It helps survivors to see their organization as a fair and considerate employer. Among those downsized, outplacement can mitigate the damaging effects of unemployment on family life by including the terminated employee's spouse in counseling sessions. Subsequent career assessment and job-search assistance may include activities such as an interest inventory as well as "how to" advice on building a résumé and doing well in a job interview. These activities focus on helping the terminated employee to identify his or her strongest skills and career preferences, to negotiate job offers, and to select the best placement.

### See Also the Following Articles

Career Counseling ■ Organizational Justice

### Further Reading

- Cascio, W. F. (1993). Downsizing: What do we know? What have we learned? *Academy of Management Executive*, 7(1), 95–104.
- Cascio, W. F. (2002a). *Responsible restructuring: Creative and profitable alternatives to layoffs*. San Francisco: Berrett-Kohler.
- Cascio, W. F. (2002b). Strategies for responsible restructuring. *Academy of Management Executive*, 16(3), 80–91.
- Cascio, W. F., & Young, C. E. (2003). Financial consequences of employment-change decisions in major U.S. corporations: 1982–2000. In K. P. De Meuse, & M. L. Marks (Eds.), *Resizing the organization: Managing layoffs, divestitures, and closings—Maximizing gain while minimizing pain* (pp. 131–156). San Francisco: Jossey-Bass.

- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology, 86*, 425–445.
- De Meuse, K. P., & Marks, M. L. (Eds.). (2003). *Resizing the organization: Managing layoffs, divestitures, and closings—Maximizing gain while minimizing pain*. San Francisco: Jossey-Bass.
- Greenberg, J. (1997). *The quest for justice on the job*. Thousand Oaks, CA: Sage.
- Gutknecht, J. E., & Keys, J. B. (1993). Mergers, acquisitions, and takeovers: Maintaining morale of survivors and protecting employees. *Academy of Management Executive, 7*(3), 26–36.
- Kanovsky, M. (2000). Understanding procedural justice and its impact on business organizations. *Journal of Management, 26*, 489–511.
- Kleinfeld, N. R. (1996, March 4). The company as family no more. *The New York Times*, pp. A1, A8–A11.
- Maslach, C., & Leiter, M. P. (1997). *The truth about burnout: How organizations cause personal stress and what to do about it*. San Francisco: Jossey-Bass.
- Reichheld, F. F. (1996). *The loyalty effect*. Boston: Harvard Business School Press.
- Schweiger, D. M., & Denisi, A. S. (1991). Communication with employees following a merger: A longitudinal field experiment. *Academy of Management Journal, 34*, 110–135.



# Driving, Risky

David W. Eby

University of Michigan, Ann Arbor, Michigan, USA

1. Factors Related to Risky Driving Behavior
  2. Theories of Risky Driving Behavior
  3. Countermeasures for Reducing Risky Driving Behaviors
  4. Conclusion
- Further Reading

increase his or her chances of being in a crash or becoming severely injured in a crash.

**sensation seeking** The behavior associated with a psychological need to seek out and engage in activities for the purpose of stimulation.

## GLOSSARY

**countermeasures** Programs or other activities designed to solve or alleviate a problem.

**graduated licensing** A legislative countermeasure designed to reduce risky driving and crashes among young, beginning drivers.

**mobile phones** Wireless communication devices that are also known as cellular telephones or cell phones. A mobile phone is either hand-held, requiring the driver to handle the phone while engaging in a conversation, or hands-free, allowing the driver to have a conversation without physically holding the phone.

**occupant crash protection devices** Devices that are designed to reduce the severity of injury incurred by a motor vehicle occupant during a crash. The two most common devices are safety belts and airbags.

**perceived risk** A driver's perception of the probability of a negative outcome for a certain driving behavior. The two most important negative outcomes for traffic safety are crashes and citations.

**risky driving behaviors** Actions that increase, above a particular threshold, the objective likelihood of a crash or the severity of injury should a crash occur. A driver may not consider his or her actions to be risky even though they

The automobile provides unprecedented personal mobility. The costs of this convenience, however, are traffic crashes and the deaths and injuries caused by these crashes. In the United States, for example, there were 6,394,000 police-reported traffic crashes, resulting in non-fatal injury to 3,189,000 people and fatal injury to 41,821 people, in the year 2000. In fact, motor vehicle crashes are the leading cause of death for people aged 4 to 33 years.

For decades, traffic safety professionals have worked to reduce the frequency of motor vehicle crashes and the severity of resulting injury without compromising personal mobility. While this problem has been approached from several perspectives, such as designing safer roads and developing occupant crash protection devices for cars, perhaps the most important viewpoint comes from behavioral scientists. Regardless of the safety features present in cars or on the roads on which they travel, cars are operated by people whose behaviors are influenced by a multitude of psychological factors. Therefore, an important component of the traffic safety problem is understanding, predicting, and modifying the behaviors of drivers.

Because of its clear influence on crashes, much research has focused on risky driving behavior. Estimates suggest that risky driving behavior causes or contributes to at least 40% of crashes. Risky driving behaviors are those actions that increase, above some threshold, the objective likelihood of a crash or the severity of injury should a crash occur. As such, a driver may not consider his or her action to be a risky one even though it increases his or her chances of being in a crash or becoming severely injured in a crash. This definition of risky driving behavior assumes a threshold from which to assess the increase in risk or crash severity. This objective threshold is set by societal standards. In the case of speeding, for example, the threshold may be the speed limit, the speed of traffic flow, or the speed that is safe for the current road or weather conditions. Thus, drunk driving, speeding, running red lights, talking on a mobile phone, and lack of safety belt use are all examples of risky driving behaviors.

## **1. FACTORS RELATED TO RISKY DRIVING BEHAVIOR**

Certain groups of people seem to engage in risky driving more often than others. Reviewed here are four of the most common demographic and personality factors associated with risky driving.

### **1.1. Age**

Observational and self-reported studies of driver behavior show that young people, more frequently than others, speed, travel with shorter headways, run traffic lights, use hand-held mobile phones, and fail to use safety belts. Drinking and driving is also common among young drivers, peaking between 21 and 35 years of age. These empirical studies mirror U.S. crash statistics that show that the number of motor vehicle crashes per licensed driver is highest for young drivers and declines with age, suggesting that the frequency of risky driving decreases with age.

### **1.2. Gender**

It is well established that men, more often than women, drive after consuming alcohol, tailgate, speed, and fail to use safety belts. Interestingly, however, use of hand-held mobile phones while driving is more common among women than men, although data on the contribution of

mobile phone use to crash involvement have not yet been routinely or accurately collected. Crashes involving alcohol, speeding, and lack of safety belt use are more common for men than women. Collectively, these data show that risky driving behavior, in general, is more common among men than women.

### **1.3. Sensation Seeking**

A large body of work has documented the fact that certain people, referred to as sensation seekers, have a psychological need for a higher level of arousal than others, leading to negative traffic safety consequences. Sensation seekers engage in new, complicated, or emotionally intense activities, such as taking risks, because of the resulting increased arousal. Thus, a sensation seeker might engage in risky driving behavior simply to experience a situation in which physiological arousal will be elevated. Sensation seeking can be measured with a test, the Sensation Seeking Scale (SSS), in which behaviors related to sensation seeking are self reported. Several studies have shown that men score higher on the total SSS than women. Studies have also documented that SSS scores tend to increase with age up to about 19 years and then decline gradually through the life span. SSS scores also correlate significantly with frequency of motor vehicle crashes, traffic citations, and many observed and self-reported risky driving behaviors such as drinking and driving, speeding, and lack of safety belt use. Thus, there appears to be a link between risky driving behaviors and the biological need to seek arousal.

### **1.4. Attention Deficit/Hyperactivity Disorder (ADHD)**

There is some evidence that people diagnosed with ADHD may engage in risky driving behavior more often than others. Although few studies have investigated the role of ADHD in risky driving behavior, those that have reported that people with ADHD were more likely than control subjects to speed, disregard traffic control devices, drive without a license, have multiple crashes, and have multiple moving and nonmoving violations on their record. They were also more likely to have records of failure to appear in court and failure to pay fines, suggesting that people with ADHD tend to not follow through on the consequence of their violations. Because of methodological weaknesses of these studies, these results should be considered as preliminary. In

addition, the relationship between risky driving behavior and successfully treated ADHD is unknown.

## 2. THEORIES OF RISKY DRIVING BEHAVIOR

The last two decades have seen a profusion of theories on risky driving behavior, with no single one emerging as clearly superior. Collectively, however, these theories show various perspectives on how risk is conceptualized and its influence on driving. In addition, there are many commonalities among the theories that can help practitioners develop countermeasures for risky driving behaviors. The following sections briefly describe a sample of theories chosen to illustrate the breadth of thinking about risky driving behavior.

### 2.1. Zero Risk

The theory of zero risk assumes that driving behaviors result from motives, a primary one being the need to be mobile. As people drive, they experience subjectively risky situations, which over time they learn to avoid. In addition, as drivers gain experience, previously threatening situations that cannot be avoided begin to be perceived as less risky the more they encounter these situations without adverse consequences. The theory posits that people both learn to avoid risk and adapt to risky situations they encounter by a reduction in perceived risk, to the point of experiencing zero risk while driving. Risky driving behaviors, therefore, occur either because of inexperience or because of inappropriate amount of perceived risk for certain behaviors.

### 2.2. Risk Homeostasis

The theory of risk homeostasis maintains that people in a population have a preferred level of risk for a given activity. Rather than attempting to minimize risk while driving, people change their driving to reach the preferred level of risk. If the perceived risk of crash or citation is reduced, through roadway improvements or reductions in police presence, for example, drivers will simply increase their risky driving behaviors to compensate.

### 2.3. Utility Maximization

Theories of utility maximization posit that people choose driving behaviors based upon the outcome of

a cost–benefit analysis, in which the cost is the perceived risk associated with certain driving behaviors. People engage in the behavior that maximizes their benefit, or utility, at the present time. Therefore, risky driving behaviors could arise from (1) assigning a low cost to the risky outcome, such as a crash, (2) perceiving a low risk for the behavior, or (3) believing that the benefits of the risky behaviors, such as saving time, outweigh the costs of the behavior.

### 2.4. Problem Behavior

The theory of problem behavior posits that all risky behaviors of adolescents and young adults, not just driving, are interrelated. The theory suggests that risky behaviors result from the combination of influences from three general components of an individual's life: the psychological, social, and behavioral characteristics of the individual; the social environment; and the specific situation for the risky behavior. Each of these components has influences that make risky behavior more likely and less likely. These components interact and result in a tendency to engage in or not engage in risky behaviors. An important aspect of this theory is that behavior is motivational in nature; individuals may wish to engage in risky behaviors because they fulfill certain developmental needs, such as the need for peer approval. According to the theory, risky driving behavior results from both a motivation to engage in the behavior and a lack of inhibiting influences in the personal, social, or situational contexts to prevent the behavior.

### 2.5. Decision Making

Decision making theories conceptualize risky driving behaviors as resulting from a decision-making process in which risk is just one of several factors that may be considered. Therefore, the cognitive abilities of people, such as the speed at which they can process information, their memory capacity, and their general knowledge of the world, influence the selection of a given behavior. Various components of the decision, including the perceived risk, are combined to determine a subjective worth for the behavior among the set of behaviors available for a given situation. In these theories, risky driving behaviors can occur for the following reasons: (1) the risky driving behavior is selected over a less-risky behavior because it affords the person greater perceived benefit, or (2) the person, relative to others, is lacking cognitive skills or abilities for good decision making.

### **3. COUNTERMEASURES FOR REDUCING RISKY DRIVING BEHAVIORS**

In general, traffic safety countermeasures are not based on theory. However, designing psychology-based countermeasures that are effective in reducing the incidence of risky driving behavior is of great interest to traffic safety professionals. These countermeasures fall into three general classes: legislation/enforcement, licensing, and educational programs.

#### **3.1. Legislation/Enforcement**

Most traffic safety legislation is designed to prevent risky driving behavior by making it illegal to engage in these behaviors. We are all familiar with the fact that it is illegal to drive over the speed limit, fail to stop at a stop sign, or not use a safety belt. This legislation, however, cannot be effective without enforcement. Legislation and enforcement countermeasures work on the principle of using punishment, or the threat of punishment, to reduce the likelihood of risky driving behaviors. The effectiveness of such programs has been shown in studies of drinking and driving, safety belt use, bicycle helmet use, and motorcycle helmet use. Although not commonly used, positive reinforcement has also been shown to be effective in reducing risky driving behavior. At least one study has shown the effectiveness of a variable reinforcement schedule in increasing safety belt use. In this study, police randomly pulled over drivers and rewarded them with prizes for wearing safety belts. Legislation coupled with enforcement is the most effective way to combat risky driving behavior.

#### **3.2. Licensing**

Countermeasures based on driver licensing are designed to legally restrict driving privileges until the driver meets certain criteria. Because of the high crash rate and high incidence of risky driving among young people, for example, many states have implemented a graduated licensing system. With graduated licensing, young drivers gradually gain skills, experience, and knowledge over a period of time under controlled, less risky conditions. Generally, there are three phases to graduated licensing. In the first phase, the young driver can operate a vehicle only under adult supervision. In the second phase, the young person can drive without an adult present, but his or her unsupervised driving is subject

to restrictions, such as daytime driving only. In the third phase, the driver is given full driving privileges. Usually, these programs incorporate motivation for safe driving behavior by requiring the driver to be citation- and/or crash-free for a period of time before graduating to the next phase. These programs also occasionally include educational components in which the driver attends and passes a class and/or demonstrates driving proficiency via a driving test. In summary, graduated licensing is designed to reduce risky driving among young people by rewarding safe driving behaviors and giving young drivers the opportunity to gain experience, acquire knowledge, and mature under conditions in which risky driving behavior is less likely. Evidence suggests that graduated licensing does reduce young driver crashes by as much as 25%, probably by reducing the incidence of risky driving among young people.

Suspended licensing is also used as a sanction following a conviction for certain driving violations, usually drunk driving. The removal of a person's driving privileges is intended to both separate the person from his or her vehicle so that the undesirable driving behaviors cannot occur for some period of time, and to punish the person by taking away a privilege. This sanction has been found to be modestly effective for reducing the incidence of drunk driving, both during the sanction period and after.

#### **3.3. Educational Programs**

Educational countermeasures attempt to reduce the incidence of risky driving behaviors and crashes by presenting drivers with information relevant to traffic safety. The topics of these programs vary widely. Probably the most familiar program is the driver education course that is required in most jurisdictions for beginning drivers. The effectiveness of these programs has been extensively researched and the unfortunate conclusion is that these programs do not reduce crashes or risky driving; they do, however, teach people the rules of the road and how to operate a car. Similar findings are reported for motorcycle training programs. Other programs have attempted to teach young drivers advanced skills, such as controlling a vehicle during a skid. It has been found that when compared to a control group, young drivers completing the course actually had more crashes, although the difference was not statistically meaningful. Educational programs have attempted to teach people about the negative consequences of risky driving. The most common of these programs are victim impact panels, in which a convicted drunk driver is required to listen to

statements from the family of the victim impacted by that driver's behavior. Again, evaluations of these programs have shown that they do not reduce drunk driving behavior or crashes. As a final example, another common countermeasure involves public information and education about the presence and enforcement of a particular traffic safety law. These programs are designed to inform or remind the public about specific traffic safety legislation, such as a mandatory safety belt law, and tell the public that police are making a special effort to enforce the law, that is, to create an increase in the perceived threat of enforcement. These programs have been found to be effective, when coupled with actual enforcement, in increasing safety belt use, decreasing drunk driving, and decreasing aggressive driving.

The disappointing conclusion from these examples is that, to date, few educational programs positively impact risky driving behaviors, except those coupled with legislation and enforcement. One should not, however, draw the conclusion that educational programs cannot be effective, but rather that more research is needed in order to develop effective programs. As discussed earlier, there is no unifying or even leading theory of risky driving behavior. Lacking such theory, program development is forced to proceed in an *ad hoc* fashion. Another challenge is developing programs that use cognitively appropriate language for the target group. As recently recognized by the National Highway Traffic Safety Administration (NHTSA), traffic safety programs targeted toward young people may not accommodate their cognitive abilities. Consequently, NHTSA has begun a program to apply the principles of cognitive development in the construction of young driver programs. Finally, it should be noted that measurement of the effectiveness of an educational program has several inherent difficulties. The typical outcome measures are either infrequent (crashes), requiring several years of follow-up data, or difficult to measure on the road (risky driving behaviors). Furthermore, many programs may not reach the proper audience or may suffer from a self-selection bias, compromising the validity of evaluations.

#### 4. CONCLUSION

Advances in our understanding and prevention of risky driving behaviors lie ahead. Many of these advances will undoubtedly result from the application of psychological theory and principles. Further research is needed to develop and test theories of risky driving behaviors and continued efforts are needed to integrate these theories

into effective countermeasures. Success in these areas will greatly enhance the prevention of the death and injury toll caused by motor vehicle crashes.

#### See Also the Following Articles

Accidents in Transportation ■ Attention Deficit/Hyperactivity Disorders (ADHD) ■ Decision Making ■ Driving Safety ■ Transportation Systems, Overview ■ Travel Behavior and the Environment

#### Further Reading

- Bloomquist, G. (1986). A utility maximization model of driver traffic safety behavior. *Accident Analysis & Prevention*, 18, 371–375.
- COMSIS Corporation & the John Hopkins University. (1995). *Understanding youthful risk taking and driving*. Report No. DOT-HS-808–318. Washington, DC: U.S. Department of Transportation.
- Eby, D. W., & Molnar, L. J. (1999). *Matching safety strategies to youth characteristics: A literature review of cognitive development*. Report No. DOT-HS-808–927. Washington, DC: U.S. Department of Transportation.
- Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. New York, NY: Academic Press.
- Jonah, B. A. (1986). Accident risk and risk-taking behavior among young drivers. *Accident Analysis & Prevention*, 18, 255–271.
- Jonah, B. A. (1997). Sensation seeking and risky driving: A review and synthesis of the literature. *Accident Analysis & Prevention*, 29, 651–665.
- Konecni, V. J., Ebbesen, E. B., & Konecni, D. K. (1976). Decision processes and risk taking in traffic: Driver response to the onset of yellow light. *Journal of Applied Psychology*, 61, 359–367.
- Mayhew, D. R., Simpson, H. M., Williams, A. F., & Ferguson, S. A. (1998). Effectiveness and role of driver education and training in a graduated licensing system. *Journal of Public Health Policy*, 19, 51–67.
- Näätänen, R., & Summala, H. (1976). *Road-user behavior and traffic accidents*. New York, NY: Elsevier.
- National Highway Traffic Safety Administration. (1995). *Understanding youthful risk taking and driving*. Report No. DOT-HS-808–318. Washington, DC: U.S. Department of Transportation.
- Shope, J. T., Molnar, L. F., Elliott, M. R., & Waller, P. F. (2001). Graduated driver licensing in Michigan: Early impact on motor vehicle crashes among 16-year-old drivers. *Journal of the American Medical Association*, 286, 1593–1598.
- Simpson, H. M. (1996). Summary of key findings: Research and information needs, program and policy priorities. In H. Simpson (Ed.), *New to the road: Reducing the risk for*

- young motorists* (pp. 1–17). Berkeley, CA: The Regents of the University of California.
- Wasielewski, P. (1984). Speed as a measure of driver risk: Observed speeds versus driver and vehicle characteristics. *Accident Analysis & Prevention*, 16, 89–104.
- Wilde, G. J. S. (1982). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis*, 2, 209–225.
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. New York, NY: Cambridge University Press.
- Zuckerman, M., & Neeb, M. (1980). Demographic influences in sensation seeking and expressions of sensation seeking in religion, smoking, and driving habits. *Personality and Individual Differences*, 1, 197–206.



# Driving Safety

Fredrick M. Streff

University of Michigan, Ann Arbor, Michigan, USA

1. Driving Safety as a Behavioral Target
2. Speeding
3. Alcohol-Impaired Driving
4. Safety Belt Nonuse
5. Driving Safety among Older Drivers
6. Driving Safety among Younger Drivers
7. Conclusion

Further Reading

This article describes behavioral targets and strategies to improve driving safety. Speeding, alcohol-impaired driving, and safety belt use behaviors are discussed, as are unsafe driving behaviors among young (age 14–18 years) and older drivers (age 70+ years). For each behavior or set of behaviors, those factors that support safe driving or reduce poor driving and those that undermine safe behaviors or support unsafe driving behaviors are discussed.

## GLOSSARY

**alcohol-impaired driving** Instances in which at least one driver involved in a crash was impaired by alcohol.

**alcohol-involved crash** A crash in which at least one involved driver had been drinking prior to the crash, but the level of impairment is unknown.

**crash severity** Measured in terms of the extent of the damage sustained by the vehicle involved in a crash and the severity of the injuries sustained by the people involved in the crash.

**deterrence** Describes avoidance behaviors of drivers that are formed by perceptions of enforcement levels as well as punishment likelihood and severity.

**graduated licensing** A licensing procedure by which young drivers are permitted driving privileges incrementally based on the successful completion of each licensing phase, allowing young drivers to experience more complex driving situations gradually.

**speed-related crashes** Crashes in which a driver was traveling in excess of the posted speed limit or the speed was higher than the safe driving speed for the conditions.

## 1. DRIVING SAFETY AS A BEHAVIORAL TARGET

Driving safety is measured objectively by the presence or absence of motor vehicle crashes and/or the severity of motor vehicle crash outcomes. Crash severity is typically measured in terms of the damage sustained by the vehicle involved in the crash and the severity of the injuries sustained by the people involved in the crash. Driving safety is also assessed by the presence or absence of precursor behaviors demonstrated to be related to crash involvement and/or severity.

According to the National Highway Traffic Safety Administration (NHTSA), in 2001 there were 6,323,000 traffic crashes in the United States resulting in nearly 40,000 deaths and an additional 2,003,000 injuries. Annually, there are approximately 20 persons killed and an additional 1585 persons injured for every 100,000 licensed drivers. The NHTSA estimates that

the total economic cost of traffic crashes in 2000 was \$230.6 billion.

Although improvements in vehicle and road design have contributed much to improving driving safety, perhaps the most significant advances in driving safety in the past 15 years are behavior change programs designed to reduce the likelihood of drivers engaging in unsafe behaviors and programs designed to increase the likelihood of safe driving behaviors. Specifically, most efforts in behavior change to improve driving safety have focused on five target areas: speeding, alcohol-impaired driving, safety belt use, and driving behaviors of older drivers (age 70+ years) and young drivers (age 14–18 years).

## 2. SPEEDING

The NHTSA estimates that speeding (traveling in excess of the posted limit or traveling too fast for conditions) is a contributing factor in 30% of fatal crashes, resulting in approximately 12,000 lives lost in speeding-related crashes per year. Speeding reduces a driver's ability to steer around curves or objects in the roadway, extends the distance that is needed to stop a vehicle, and increases the distance a vehicle travels as a driver reacts to a situation. Moreover, the dynamic forces generated in a crash increase exponentially with the speed of the vehicle.

Speeding behaviors are motivated and reinforced by a wide array of conditions. Speeding behaviors can result from consciously accepting risks inherent in the behavior or unconsciously by misinterpreting speed cues and subsequent risks. Drivers often speed to achieve their goal of reaching a given location at a given time. Some speed because they are not attending to available speed cues. A speedometer serves as an objective and reliable cue of vehicle speed. However, subjective cues in the driving environment (e.g., the rate at which trees on the roadside pass by as one is driving) may or may not be perceived by a driver. Even if perceived, a driver may not understand the meaning of these objective and subjective speed cues, the risks associated with the cues, and what the safe driving speed may be for a given situation. Variations in risk perceptions and acceptance among drivers compound these relationships.

Expectancies play a significant role in a person's perception and acceptance of speed. For example, drivers turning into a traffic stream on a road with a 30 mph limit would not generally expect to have a vehicle

approaching at 100 mph. A misperception of closing speed in this case could easily cause a traffic crash when the turning vehicle crosses into the path of the oncoming car that should have been farther away when the turn began. Another expectancy that many drivers have is that any given driver will get to his or her destination much sooner by passing traffic traveling slowly in a cue, like one would find when two lanes of a highway are merged into one because of road construction. Although there can be a measurable time savings from speeding, the gains are minimal. A vehicle traveling in an open lane at 50 mph passing a 3-mile-long line of traffic traveling at 10 mph will gain less than 2.5 min (and no doubt anger a number of people). In fact, the faster car has to pass approximately 700 vehicles to save the 2.5 min. Passing slower vehicles may create a perception of getting to the destination faster than is warranted by objective information.

The primary means applied to limit speeding behavior has been traffic law enforcement. The theoretical roots for this approach are found in learning theory. Enforcement strategies to control speeding behaviors target change in a driver's perception of the likelihood of arrest, prosecution, conviction, and the severity of the subsequent sanction. That is, the programs change the perceived reinforcing and punitive qualities of the behavioral consequences. These perceptions serve as antecedent conditions signaling a heightened probability of negative consequences resulting from speeding. That is, speed can be regulated by decreasing the perceived utility/reinforcing properties of speeding. For example, perhaps drivers would perceive fewer benefits associated with driving on an adjoining lane of slowly moving traffic if they were made aware of the small gain made. On the other hand, if drivers perceive the environment as being too dangerous to speed, they will try to avoid the negative consequences by traveling at safe, legal speeds.

When the level of enforcement activity and its concomitant publicity decrease, changes in drivers' perceptions of the risks and rewards of speeding will result in a return to the speed behaviors that were reinforced (or at least not punished) prior to the special enforcement program. These changes may occur quickly, as is the case among drivers using two-way radios to alert other drivers to the threat of enforcement. These informed drivers will slow while in the enforcement zone and resume speed once the threat has passed. These changes may also occur slowly, as is the case when enforcement patrols target an area using an intermittent enforcement schedule.

### 3. ALCOHOL-IMPAIRED DRIVING

According to the NHTSA, 41% of persons killed in traffic crashes in the United States in 2001 were killed in crashes involving alcohol. That is, 17,448 persons were killed in alcohol-involved crashes in 2001. This seemingly dangerous activity is motivated by the desire to travel from point A to point B and is in part controlled by the extent to which drivers perceive risk in the behavior or reward in alternate behaviors. Behavior change programs to reduce the likelihood of alcohol-impaired driving have taken the forms of reducing the response cost for alternate behaviors, programs to reduce the social acceptability of drinking and driving, interventions to reduce alcohol dependency, and enforcement campaigns to increase the perceived threat of drinking and driving behavior.

Some of the most innovative and successful programs to reduce alcohol-impaired driving have been those that reduce the response cost and even reward alternate behaviors incompatible with alcohol-impaired driving. Incentives in the form of free food and nonalcoholic beverages for persons who serve as the designated driver for a group have proven to be successful. An interesting program that prevents a person from driving while impaired is the "tipsy taxi" program. In this program, a patron who has had too much to drink is given a ride home by a volunteer. The patron or a friend then picks up the car from where it was parked at a later time.

The primary challenge to these programs is that the response cost for the desired behaviors can be perceived to be quite high. A designated driver may resent not being able to drink or become angry at boisterous drinking friends. The designated driver may feel uncomfortable to the extent that he or she can be identified and treated poorly by others in an establishment because he or she is the designated driver. In addition to the embarrassment of having to have someone else drive one home, the tipsy taxi has a significant response cost obstacle to overcome: People do not like to leave their cars on the street or in a parking lot near a bar.

Perhaps the most common programs to reduce impaired driving are those intended to persuade people that alcohol-impaired driving is an unacceptable behavior. The foci of these persuasion programs are attitudes, perceptions, and social norms associated with drinking and driving. These programs use mass media distribution of materials to educate and persuade readers to adopt the desired set of values and to engage in the desired behaviors. Although these types of

information and persuasion strategies are an effective, almost necessary, adjunct to other types of programs, persuasion programs are not in themselves an effective means to change driving behavior.

As is the case for speeding behaviors, policies, laws, and their enforcement are significant components of alcohol-impaired driving reduction programs. Strengthening laws and increasing sanctions may increase their deterrent value in two ways. First, a law may serve to discourage a group of persons from engaging in the illegal act (general deterrence). Second, a law may serve to affect the behavior specifically of persons who have had a previous experience with violating the law (specific deterrence). For example, a new law that affects only repeat offenders may have a general effect (a reduction in drinking and driving among persons without a conviction) in addition to its desired effect on the target group of repeat offenders (specific deterrence).

Behavior change programs targeting a reduction in alcohol-impaired driving also include interventions designed to reduce alcohol dependency. The most significant problem faced by treatment-based programs to reduce alcohol-impaired driving is the fact that often participants are ordered by the court to attend the treatment sessions that the court prescribes. The result is a client who has little motivation to participate actively in the change program imposed by the court and is being asked to make a major life change using a strategy that may well be inappropriate for that client.

### 4. SAFETY BELT NONUSE

When used, lap/shoulder belts reduce the chance of death or serious injury from a traffic crash by 45–60%. The NHTSA estimates that in 2001, use of safety belts prevented more than 12,000 fatalities that would have otherwise been caused by traffic crashes. Despite the demonstrated effectiveness of safety belts for preventing death and reducing injury, a substantial proportion of people do not use safety belts. Although NHTSA observation studies have shown that safety belt use has increased considerably from its 1994 level of 58% to its 2002 level of 75%, one out of four vehicle passengers still travel without this proven life-saving device.

Programs to promote safety belt use have focused primarily on managing short-term consequences of safety belt use. The most significant consequence, loss of life, is an extremely rare event when considered on a

trip-by-trip basis. Because of its infrequency, this consequence does little to promote belt use. In fact, prior to the implementation of laws mandating safety belt use, the national rate of belt use was approximately 14%. This rate shows that other, short-term consequences are necessary to increase belt use. Perhaps the most underused programs for promoting safety belt use are incentive programs in which a person or group of persons are rewarded for belt use.

Incentive programs have proven to be effective in achieving belt use in the 90–100% range at industrial sites and within communities. At the heart of the programs is the delivery of rewards for belt use rather than punishments for belt nonuse. Some jurisdictions have used police patrols to pull vehicles out of the stream of traffic to reward drivers, similar to the manner in which drivers are given tickets. These programs often create resentment toward police because the circumstances surrounding the delivery of the reward are perceived to be more punitive than the value of the reward. Strategies in which the delivery of rewards is concurrent with another activity in which belt use can be assessed easily by observation (e.g., fast-food drive-throughs) are highly successful.

The regularity and generalization of belt use can be accomplished with contingency management. Delaying the reward for belt use until multiple examples of the behavior are observed or using an intermittent reinforcement schedule can serve to extend the effects of the program past the time the program has ended. An example of delayed reward is a program in which persons who are observed wearing their belts receive a single bingo letter each time they order from the drive-through. The completed card may be redeemed for a product from the participating store. An intermittent schedule is achieved when belt users select a card randomly from a set of cards in which there are winning and losing cards.

Using a mixed cognitive and behavioral approach provides another strategy for promoting belt use. In this type of program, persons visiting participating stores may be given an opportunity to complete a form that pledges them to use their safety belt on every trip they take. The form is divided to create a raffle ticket that is entered in a weekly drawing for nominal prizes and a rear-mirror tag displaying their commitment to belt use. The mirror tag serves two purposes. First, it serves as a cue to the driver to put on the safety belt. Second, volunteers observe belt use on streets in the area and record the plate number of every person with a mirror tag who is using a belt. This is done throughout the program period, after which a

small number of desirable prizes are awarded to persons with the highest number of observed belt use. This mixed strategy is most effective at sites where the number of observations that can be made of all participants is approximately the same (e.g., parking lots).

The effectiveness of incentive programs is often improved by using group contingencies. When group contingencies are applied, the reward is earned only when each member of the group achieves the target level of the desired behavior. In this type of program, motivation is not derived from desire to obtain the reward alone. There are additional motivations resulting from the group asserting its influence on a person who may be preventing the group from receiving its reward.

There are also a variety of belt promotion strategies that employ punishment or the threat thereof to achieve program goals. Changes in enforcement policy mandated by laws requiring belt use have resulted in the most dramatic increases in belt use among the general population. Many, if not most, who begin to use their belt after this type of law change do so because they obey laws generally. Others respond to the fact that being pulled over is unpleasant and buckle up to avoid the stress of getting a ticket. The group of nonusers consists of persons who believe it is a personal decision; those who derive satisfaction merely from “tweaking the system”; and those who, for personal reason (e.g., fear of entrapment and discomfort), choose not to use belts.

Efforts to increase belt use through the exclusive use of mass media strategies have proven to be universally ineffective. These programs are based on increasing knowledge and shaping attitudes conducive to belt use. Although ineffective when used in isolation, media strategies can be highly effective and essential components to promoting an incentive- or disincentive-based safety belt promotion program.

## 5. DRIVING SAFETY AMONG OLDER DRIVERS

Achieving and maintaining safe driving behaviors among older drivers is going to be a growing challenge in the future. The NHTSA reports that in the United States there were 18.9 million licensed drivers age 70 or older in 2000, a 36% percent increase since 1990. During the same time period, the general population of older persons increased 6%, whereas the number of older drivers involved in fatal crashes increased at three times this rate (18%).

The basic motivations for driving at age 70 do not differ from those of younger drivers. One drives to get from one point to another. However, driving allows older persons the mobility required to maintain their physical and mental health. Without independent transportation, older persons may feel isolated and depressed because they cannot get together with family or friends without help; their access to medical treatment and fresh foods is reduced, further increasing their risk of an early death; and often they lose feelings of maturity and value as their driving abilities restrict their driving. The privilege to drive is often an important rite of passage from youth to adulthood in the United States. In order to avoid these negative consequences, many older persons drive in situations they have difficulty managing and may not freely admit to family members or others that they have difficulty driving in these situations.

Changes in vision, reduced range of motion, and slowed reflex and reaction times are among the effects age has on the body. Cognitively, there is a reduction in attention resources required for effective perception and decision making. Moreover, as people age, the ability to divide attention between several tasks and to identify and ignore irrelevant information decreases. As drivers age, they tend to perceive a given situation as being more threatening and their ability to respond lower than is actually the case. Older persons often take a variety of medications that may affect their ability to perceive or act in a manner consistent with safe driving. These changes affect the number, cause, and severity of crashes experienced by older drivers.

Compared to younger drivers, older drivers have a lower rate of alcohol involvement in fatal crashes but are increasingly under the influence of multiple prescription drugs that may affect driving in ways that have not yet been determined. Older drivers have a crash profile similar to that of younger drivers, but the causes of these profiles are different. As drivers age, their perceptions and assessment of risk are greater than they are in reality, and their perceived skill at meeting these challenges is lower than it is objectively. Comparing crash configurations for older and younger drivers, the NHTSA has shown that vehicles with drivers age 70 or older are three times more likely to be the vehicles struck in fatal collisions and six times more likely to crash while turning left than is the case for younger drivers.

Worsening vision and flexibility caused by aging may make assessments of an oncoming vehicle's approach speed more difficult, thus requiring more time to scan the environment. This reduces the time available for decision making. Also, age-related deficits

in decision-making processes may result in drivers making decisions too slowly to maneuver to avoid a collision. Finally, once the decision is made cognitively, the ability of older persons to act on that decision behaviorally is also impaired by age.

The gradual degradation of abilities necessary for safe driving creates a significant challenge. Some programs have attempted to address problems associated with age by using training sessions to increase skeletal flexibility and mental dexterity. Although these programs have had limited success, it may be useful to consider a licensing system that allows drivers to exit the driving population gracefully and in a manner consistent with their condition. However, one must keep in mind that alternatives to driving oneself from point A to point B may be extremely limited. This is especially true outside large metropolitan areas. Because there are few opportunities for traveling locally for day-to-day necessities, states are reluctant to modify licensing laws that may further restrict an elderly person's ability to be independently mobile and drive.

## 6. DRIVING SAFETY AMONG YOUNGER DRIVERS

According to the NHTSA, motor vehicle crashes are the leading cause of death for persons age 15–20 in the United States. In 2001, 3608 drivers in this age group were killed in fatal crashes and traffic crashes injured an additional 337,000. More disturbing is the fact that although drivers age 15–20 constitute only 7% of the driving population, they are involved in 14% of all fatal crashes and approximately 20% of all reported crashes. Research has shown that teens are three or four times more likely to be involved in a crash during their first 2 or 3 years of driving compared to the general driving population.

For drivers age 15–20 involved in traffic crashes, police recorded the cause of the crash to be driver error or other factors related to the driver's behavior in 75% of cases. Compared to drivers older than age 20, young drivers are more likely to exceed posted speed limits, drive too fast for conditions, follow too closely, cross traffic lanes, pass other vehicles, and approach signals at higher speeds. Young drivers are also less likely to use their safety belts and are affected by lower levels of alcohol than are older drivers. One of the central causes of these problems is the fact that younger drivers perceive a given situation to be less

risky and their ability to handle a given situation to be greater than is actually the case.

Until recently, driver education classes were the primary means by which young drivers were trained to drive. These programs are excellent at teaching basic skills, such as accelerating, stopping, and steering, but do little to improve the necessary perceptual and decision-making skills required for safe driving. Research has shown that driver education may at best result in a small and short-lived improvement in driving safety but appears to have no effect on crash involvement in the long term. Recently, states have begun to implement a driver licensing protocol called graduated licensing.

Young drivers are involved in a higher rate of crashes because of their lack of experience driving and their heightened levels of risk taking. Graduated licensing is a licensing procedure by which young drivers are given driving privileges incrementally based on the successful completion of each licensing phase. This allows young drivers to experience more complex driving situations gradually, reducing the chance that drivers will encounter situations beyond their skill levels. Because graduated licensing practices restrict young people from driving at hazardous times, there is an additional safety benefit associated with reduced exposure.

Differences in laws from state to state make it difficult to specify each stage of graduated licensing, but the process is generally the same for all states. The process generally starts with a highly restricted learner permit that permits driving only in daylight with an adult in the car. In the second stage, drivers earn an intermediate license that limits driving without a guardian to daylight hours and still requires supervision at night. The final stage is full licensure, which occurs when a person is age 17 or older, has completed at least 6 months of the intermediate license condition, and has remained violation and crash free for the prior 3 months. Research has shown that the implementation of a graduated licensing system results in approximately 25% fewer crashes involving young drivers compared to the number of crashes before the institution of a graduated license system.

## 7. CONCLUSION

Safe driving requires that appropriate behaviors take place in the correct manner at the correct time. Risky driving behaviors, such as speeding, alcohol-impaired driving, and safety belt nonuse, have been proven to be best controlled by the application of learning theory models. The most successful intervention models have

been those that managed relationships between the behavioral antecedents, the desired/undesired behavior, and the consequences associated with the behavior. By restricting driving to relatively safe conditions and extending driving experience over a long period of time prior to full licensure, the crash risk of young, novice drivers can be reduced significantly.

## See Also the Following Articles

Accidents in Transportation ■ Traffic Safety Assessment  
■ Transportation Systems, Overview

## Further Reading

- Cooper, P. (1997). The relationship between speeding behavior and crash involvement. *Journal of Safety Research*, 28(2), 83–95.
- Eby, D., & Molnar, L. (1998). *Matching traffic safety strategies to youth characteristics*, NHTSA Report 808927. Washington, DC: U.S. Department of Transportation.
- Evans, L. (1994). Safety belt effectiveness: The influence of crash severity and selective recruitment. In *Proceedings of the 38th annual conference of the Association for the Advancement of Automotive Medicine* (pp. 25–42). Des Plaines, IL: Association for the Advancement of Automotive Medicine.
- Federal Highway Administration. (1998). *Older driver highway design handbook: Recommendations and guidelines*. Washington, DC: U.S. Department of Transportation.
- Koppa, R., Holmgren, J., & Salzmann, A. (2001). *Adaptive equipment to enhance older driver performance: A guidebook*. College Station: Texas Transportation Institute.
- National Highway Traffic Safety Administration. (2001). *Achieving a high seat belt use rate: A guide for selective traffic enforcement programs*. Washington, DC: U.S. Department of Transportation.
- National Highway Traffic Safety Administration. (2002). *Traffic safety facts 2001—Overview*, NHTSA Report 809476. Washington, DC: U.S. Department of Transportation.
- Sadler, D., Perrine, M., & Peck, R. (1991). The long-term safety impact of a pilot alcohol abuse treatment as an alternative to license suspensions. *Accident Analysis and Prevention*, 23(4), 203–224.
- Shope, J., Molnar, L., Elliott, M., & Waller, P. (2001). Graduated licensing in Michigan. Early impact on motor vehicle crashes among 16-year-old drivers. *Journal of the American Medical Association*, 286(13), 1593–1598.
- Streff, F., & Eby, D. (1994). *An evaluation of the impact and effectiveness of Michigan's drunk and impaired driving laws*. Ann Arbor: University of Michigan.
- Yu, J. (1994). Punishment celerity and severity: Testing a specific deterrence model on drunk driving recidivism. *Journal of Criminal Justice*, 22(4), 355–366.



# Drug Abuse

Charles E. Dodgen

Caldwell, New Jersey, USA

1. Introduction
  2. Risk Factors and Causes of Drug Abuse
  3. Course of Drug Abuse
  4. Assessment of Drug Abuse
  5. Treatment of Drug Abuse
  6. Relapse Prevention
- Further Reading

## GLOSSARY

**pharmacology** The study of the effects of chemical substances on living systems.

**psychoactive** The ability of a drug to induce subjective effects on mood and feeling.

**psychosocial** Reflecting the combined influences of psychological and social factors.

**relapse** Resumption of drug use following a period of abstinence.

**withdrawal** A maladaptive behavior change that occurs when blood or tissue concentrations of a drug decline in an individual.

Drug abuse, also referred to as substance or chemical abuse, is the recurrent use of a drug despite the experience of problems caused by the drug use. Difficulties arising in certain areas of a user's life are of more importance to researchers and treatment professionals than other areas for identification of a drug abuse problem. The following are types of problems that signify drug abuse: impairment meeting

major responsibilities in life, such as those regarding school, work, or home; difficulties with the law and social behavior; and aggravation of physical/medical conditions due to drug use. Drug abuse is to be contrasted with drug (chemical/substance) dependence. With drug dependence, use is considered compulsive and beyond the willful control of the user. That is, someone who is drug dependent is addicted; this is thought to be a more severe condition than drug abuse. Treatment of drug abuse is accomplished primarily using a variety of counseling and psychotherapeutic techniques employed to assist the abuser to stop using the drug, to develop new behavioral and mental coping skills, and to rehabilitate his or her life from the damage caused by the substance abuse.

## 1. INTRODUCTION

In the discussion of drug abuse, it would be easy but inaccurate to label any regular use of a substance as abusive. Drug use in the United States is commonplace. Many people are capable of consuming drugs without developing problems. Drugs such as caffeine and alcohol, as well as prescription pharmaceutical products such as pain killing agents or antianxiety medication, are routinely and openly consumed every day in the United States (and in other countries as well). The various drugs affect the body differently and are used for specific purposes. For example, caffeine is used to remain alert and to enhance concentration, and

tranquilizers are used to quell anxiety and for relaxation. However, drugs of abuse all have in common the property that they are psychoactive. For the sake of discussion, drugs may be classified with respect to different properties; one commonly employed system is in terms of the effect of the drug on the central nervous system (CNS). The following is one such classification system, with examples of drugs in each category:

1. CNS stimulants: Cocaine, amphetamine, and caffeine
2. CNS depressants: Alcohol, barbiturates, benzodiazepines, and solvent inhalants
3. Psychotomimetics (also known as psychedelics or hallucinogens): Marijuana, LSD, and mescaline
4. Narcotics/Opioids: Opium, heroin, codeine, morphine, and methadone

Substance use typically begins in adolescence. Adolescent substance use does not appear to be random; that is, it follows a fairly predictable pattern. Adolescents tend to start using substances that are legal and widely available to adults: alcohol and tobacco. Due to the fact that these drugs are the starting point for substance use, they are referred to as “gateway drugs.” In 1975, Kandel developed a stage model of progression of drug use that has since been revised:

1. Beer or wine use
2. Hard liquor or cigarette use
3. Marijuana experimentation
4. Alcohol abuse
5. Prescription drug use
6. Opiates and other illegal drugs

The vast majority of adolescents experiment with the gateway drugs at least one time. However, although most individuals try alcohol and tobacco, only for a minority of adolescents does use advance to abusive levels. As the stages advance, progressively fewer adolescents are found in each category. For example, alcohol will be tried by approximately 9 out of 10 students by their senior year in high school and cigarettes by approximately 6 out of 10 students by senior year. Opiates, at the last stage of the model, will be tried by only 1 out of 100 students by senior year.

Due to the high prevalence of substance use in the United States, it should be no surprise that substance-related problems are often encountered by mental health clinicians. The relatively high frequency with which substance-related problems are encountered by mental health professionals reflects the influence of the following factors: (i) Drug abuse has the potential to

create or worsen all psychological symptoms, such as anxiety, depression, impulsive behavior, and antisocial behavior; and (ii) people seeking mental health services also tend to be at elevated risk for substance abuse problems. In other words, drug abuse harms people and contributes to psychiatric symptoms, and people experiencing psychological problems are apt to use drugs abusively.

## 2. RISK FACTORS AND CAUSES OF DRUG ABUSE

With any medical or mental health condition, it is desirable to determine the cause or causes of the affliction. Identifying the cause(s) helps to develop prevention strategies to limit or eliminate future cases and treatment strategies for those already affected by the condition. For example, after the discovery that an absence of insulin was responsible for type 1 diabetes, effective treatment of diabetes with externally supplied insulin became possible. In addition, research is under way to develop early identification tests for intervention strategies to prevent later development of diabetes. This research has led to the isolation of faulty antibodies believed to attack the insulin-producing cells of the pancreas. The antibodies can be detected before the person is symptomatic for diabetes; experimental treatments are being used in an attempt to prevent the development of diabetes in these high-risk individuals.

Human behavior is complex and defies easy explanation. Unlike certain physical characteristics (e.g., eye color) or physical disorders that can be traced to single genes, a disorder such as drug abuse likely represents the interaction of multiple genetic and environmental influences. Complicating things further, ethics prevents us from conducting experimental studies (involving environmental or genetic manipulation) that might help us to tease apart various possible influences. One way to attempt to identify possible causes of substance abuse is to study risk factors. Risk factors are those variables associated with increased likelihood of developing a substance use disorder. Classes of risk factors are listed here with examples in each class:

1. Peer: Peer substance use, strong attachment to peers, and positive peer attitudes about substance use
2. Parent/family: Parent substance use, positive attitude about substance use, parent tolerance of adolescent substance use, and family disruption (e.g., divorce)

3. Personal: Early (childhood) behavior problems, poor academic performance, anxiety/depression, and low self-esteem

4. Biological: Genetic predisposition to substance use (e.g., a parent is a substance abuser)

5. Community/social: Low socioeconomic status, high availability of substances, and deviant norms that encourage use of substances

Risk factors help us to understand influences to use substances, but we know many more people use them than become abusers. Therefore, the question as to who will progress beyond experimentation and casual use to the level of abuse is not answered by risk factors alone. It appears that use of substances is more a function of external risk factors, such as peer, social, and family factors; abuse of substances appears to be more a function of personal factors, such as psychiatric, behavioral, and emotional problems.

The biopsychosocial disease model is the most widely accepted model of substance abuse and addiction. It should be clear after reviewing the list of risk factors that biological, psychological, and social factors contribute to substance abuse. The biopsychosocial model is sufficiently comprehensive to include all known contributors to substance abuse.

### 3. COURSE OF DRUG ABUSE

Disease conditions are defined by several common factors, such as having identifiable causes, characteristic symptoms, and established treatments. In addition, diseases have an observable course. It is important to describe the course of an illness in part so that the condition can be identified (i.e., for diagnostic purposes). Also, if the untreated progression of an illness was not known, there would be no way to judge the effectiveness of treatment. Treatment interventions endeavor, essentially, to change the course of a disease. Initial attempts to describe and classify the course of alcohol abuse depicted an ever-worsening condition that eventuated in death, unless the drinking was stopped altogether. As it turns out, the long-term outcome of regular alcohol use is not certain death. Some people who use alcohol never develop problems, some who develop problems (alcohol abusers) never become addicted, and a minority of alcohol abusers (approximately one-third) exhibit the progressive deteriorative pattern of drinking. The same overall trends may be expected with other substances of abuse as with

alcohol. In 1995, Shaffer and Robbins developed a general model to describe the typical course of an addiction, consisting of the following stages:

1. Initiation: Experimentation with a drug is begun.

2. Positive consequences: At this point in the use process, only the pleasurable pharmacological and social effects of the substance are experienced.

3. Negative consequences: For those individuals who continue to regularly use the substance, eventually negative consequences are experienced in terms of health, relationships, work, school, finances, or the law.

4. Turning point: For abusers who continue despite negative consequences, there is some recognition of the damage the substance is causing in their lives and ambivalence ensues.

5. Active quitting: For some abusers, ambivalence is resolved in the direction of stopping use.

6. Relapse prevention: For those who have quit, behavior changes are maintained over time to prevent resumption of drug use.

### 4. ASSESSMENT OF DRUG ABUSE

In order to treat a condition, it must first be determined that a given individual has the condition; in other words, the diagnosis of drug abuse must be made. In medicine, objective tests via technologically advanced equipment (e.g., x-ray and magnetic resonance imagery) are often used to assist the doctor in the diagnostic process. In the evaluation of drug abuse, modern technology is hardly relevant. Biological testing, in the forms of urinalysis and evaluation of saliva and blood samples, may be used but are not the mainstay of assessment. Biological testing can determine if a specific drug or drug metabolite is present in a sample but cannot indicate anything about patterns of use, withdrawal symptoms, compulsive behavior, or consequences of use, all of which are important aspects to assess. Therefore, biological testing is confined to the role of confirming recent abstinence; this information is especially important in certain settings (e.g., criminal justice system and workplace) but of limited use in a drug abuse assessment. Since we are more interested in determining whether a pattern of abusive drug use is present or not, relevant information needs to be gathered. Therefore, the interview is the primary method by which information is acquired to make the diagnosis of drug abuse. Typically, the diagnostic interview is conducted with the person in question as well as with

others in a position to observe relevant behaviors (most often family members and/or close friends). In addition to the interview, information is sometimes acquired via self-report, paper-and-pencil tests. The following information is typically obtained during a drug abuse assessment:

1. List all substances ever used
2. Age of first use of all substances
3. How used each substance (e.g., smoke, drink, snort, etc.)
4. Age of peak use, and amount used, for each substance
5. Number of days use substance per week, for each substance
6. Amount of substance used on a typical day of use
7. Date of last use of each substance
8. List all negative consequences resulting from use of substances

Diagnosing a drug abuse disorder is only one element of the assessment process. It is also necessary to determine as part of the evaluation the most appropriate setting in which treatment should take place (e.g., outpatient, halfway house, or inpatient); the proper intensity of treatment (e.g., daily treatment or monthly treatment); whether other treatment needs exist (e.g., medical and/or psychological disorders); and specific, individual treatment goals for a given person.

## 5. TREATMENT OF DRUG ABUSE

There is no one treatment for drug abuse. This fact is a reflection of the complexity of the condition and its diverse manifestations, and it highlights the importance of the assessment process, which is critical in helping determine the best treatment for a given individual. The treatment of drug abuse may occur in different settings, with varying degrees of professional assistance (e.g., self-help/12-step and professional help) and different modalities of professional services (e.g., individual therapy, group therapy, family therapy, and pharmacological treatment). Drug abuse treatment may be characterized as specialized treatment with one main goal: to stop the use of the substance. Treatment is primarily talking therapy—counseling and psychotherapy; in addition, medications may be employed to manage detoxification from some drugs and/or to treat coexisting psychological or medical conditions. However, regardless of the setting of treatment, the intensity of the contact schedule, or who

renders the treatment, it is ultimately talking therapy that takes place. Especially early in treatment, the focus of discussion is on behavior directly related to drug use and stopping the use of the drug. Most programs and professionals recommend complete abstinence from drugs; some have the goal of harm reduction (allowing use to continue while attempting to reduce drug use to less harmful levels), but they are in the minority. As treatment progresses, and abstinence is achieved and maintained, the emphasis usually broadens to other areas of the person's life that may need repair, such as their decision-making skills, coping skills, emotional state, and relationships. In other words, the individual suffers psychological and social damage from drug abuse and may even have had significant deficits in these areas prior to his or her drug abuse; treatment is designed to improve the psychosocial functioning of the individual once he or she is drug-free.

## 6. RELAPSE PREVENTION

Drug abuse has been described as a chronic, relapsing disorder. Like all chronic conditions, long-term effort must be applied for the individual to maintain abstinence from drug use. Nobody would expect the blood sugar levels of someone with diabetes to be in a healthy range if the person only complied with the prescribed care regimen for 1 month after a visit to the physician. Likewise, if a drug abuser only applies the principles of treatment for a limited period of time, resumption of abusive habits would be expected. One way to attempt to guard against a backslide into prior behavior is to extend treatment as long as possible. In addition, teaching relapse prevention skills that an abuser may use going forward in time is an integral part of drug abuse treatment. Some common elements of relapse prevention programs include identification of high-risk situations that are likely to lead to relapse, development and practice of skills to effectively cope with risky situations, enhancement of self-confidence to be able to apply coping skills when needed, learning to limit a slip to an isolated incident rather than allow it to be the beginning of a process of abuse, drug/alcohol monitoring for abstinence verification, and developing positive behaviors (e.g., working and physical exercise).

### See Also the Following Articles

Alcohol Dependence ■ Diagnostic and Statistical Manual of Mental Disorders ■ Drug Dependence ■ Nicotine Addiction

### Further Reading

- Bukstein, O. (1995). *Adolescent substance abuse: Assessment, prevention and treatment*. New York: Wiley.
- Dodgen, C. E., & Shea, W. M. (2000). *Substance use disorders: Assessment and treatment*. San Diego: Academic Press.
- Gold, M. S. (1991). *The good news about drugs and alcohol: Curing, treating and preventing substance abuse in the new age of biopsychiatry*. New York: Villard.
- Kandel, D. (1975). Stages in adolescent involvement in drug use. *Science*, 190, 912–914.
- Schuckit, M. A. (1995). *Educating yourself about alcohol and drugs: A people's primer*. New York: Plenum.
- Shaffer, H. J., & Robbins, M. (1995). Psychotherapy for addictive behavior: A stage-change approach to meaning making. In A. M. Washton (Ed.), *Psychotherapy and substance abuse: A practitioner's handbook* (pp. 103–123). New York: Guilford.



# Drug Dependence

Douglas H. Ruben

Best Impressions International, Incorporated, Okemos, Michigan, USA

1. Concept of Drug Dependence
  2. Contingencies of Dependence
  3. Family Reasons for Recidivism
  4. Conclusion
- Further Reading

## GLOSSARY

**drug tolerance** The physiological buildup of a drug in the body's metabolic system generating a constant urgency for more of the drug to avoid withdrawal symptoms. Drug abusers "tolerate" more of an ingested chemical as the body habituates to its effects.

**drug-dependent behavioral patterns** The functional and topographical sequence of responses conditioned as a by-product of repeated episodes of drug consumption. The sequence of "behavioral patterns" consists of orderly connected responses overtly and covertly occurring from the onset of cravings to the offset of pleasure gratification. Intravenous heroin users, for example, may (i) search for a syringe (or other paraphernalia), (ii) secure an isolated place for drug preparation, and (iii) go through ritualistic steps to penetrate the vein (tighten a large rubber band around the vein site; hit the vein with two fingers to draw blood flow; insert the needle inside the vein; and compress the plunger, forcing the liquid through the syringe).

**drug dependence-induced disorders** Medical and psychological complications result from chronic drug dependence producing a cluster of symptoms otherwise not present in the addict's life. Common disorders include mood, dementia, Korsakoff's, anxiety, depression, psychosis, and sexual.

**polysubstance addiction** Progression of one drug dependence to another drug dependence so that both drugs produce similar effects, opposing effects, or complementary effects. Cross-addicted users who drink alcohol (depressant) may also enjoy sedative effects of cannabis or contrast alcohol with cocaine (stimulant-euphoric); they may also find that alcohol's calming effects take the edge off of painful crack-cocaine withdrawal.

**recidivism** The gradual relapse of drug addiction after a substantive period of abstinence. Recurrence of drug-dependent behavioral patterns starts the spiral of habits, followed by small amounts of the drug being ingested. As drug-intake levels increase to pretreatment amounts, self-control erodes into compulsive fulfillment of cravings.

**remission** Period of drug treatment when addictive habits and behavioral patterns disappear but are not extinct. Medical or social stressors may force rapid detoxification and drug-free behaviors atypical in the addict's normal life. Incarceration for drug possession and intent to deliver (drug sales), for example, may last several months, dramatically eliminating drug use. Abstinence occurs, but only artificially and temporarily since the addict restores drug dependence once released from jail.

**routes of administration** Entry routes of drug use include major sensory receptors such as inhaling or "snorting" through the nose (smell, olfactory), swallowing through the mouth (taste, gustatory), rubbing on skin (topical, tactile), injecting needles through skin (intravenous, intramuscular, or subdermal), or forceful anal insertion (suppository).

**setting events** Environmental, physiological, and psychological variables mediating the relationship between drug-dependent behaviors and their consequences. Location of drug use (e.g., home and bars), people around whom use occurs (e.g., friends and relatives), time of day, light or

darkness in room (media of contact), nervousness or indigestion (e.g., irritable bowel), migraine headaches, chronic back pain, spousal arguments, and persistent sexual deprivation are all examples of the endless continuum of drug-causing antecedents.

**substitutive effects** Spontaneous regeneration of drug cravings or even pretreatment use levels may result from addicts either consuming what they think are benign compounds or engaging in benign activities, all of which restore the body's climax of pleasure previously evoked by the drug. Recovering cocaine addicts who increase daily numbers of cigarettes smoked or drink more cups of caffeine-loaded coffee may unknowingly reactivate the body's tolerance for stimulation similar to using cocaine.

**titration** Prescribed or self-induced weaning off of one drug while simultaneously ingesting a new drug. The replacement drug may generate milder or controlled effects, such as when highly regulated methadone doses replace opioids (e.g., heroin). Titration from medication dependence may involve a gradual reduction schedule of thinning the dose or altering the frequency of intake.

**withdrawal syndrome** Deliberate or accidental cessation of continuous drug use causing the body to crave replenishment of the drug. Substances taken in greater amounts over a long period of time and then sharply reduced or stopped continue in the bloodstream until slowly diffused; this is also called detoxification.

Drug dependence is a pathological condition marked by biological tolerance and withdrawal symptoms resulting from consuming large amounts of a particular drug through various routes of administration (e.g., oral and intramuscular). Essential features of biological tolerance include intense and compulsive cravings for substances to reach a desired peak effect. This is due to two reasons. First, attainment of effect level, after repeated use, requires higher amounts of the same drug, whereas less amounts produced pleasure during earlier stages of use. Second, pleasurable effects diminish after repeated use of a drug, fooling the user into ingesting higher doses to restore a prior peak effect. Chronic, uncontrolled use may be triggered by life stressors that impede social, financial, occupational, and interpersonal functioning, typically impacting nuclear family members. Life stressors are cues or "setting events" evoking unwanted emotional, behavioral, or cognitive states eliminated through drug use. The second defining feature of drug dependence, withdrawal syndrome, involves physiological reactions upon immediate cessation of the drug. General symptoms observed include restlessness, anxiety, insomnia, and nausea, although severity of body complications depends on the type of

drug discontinued. For example, acute withdrawal symptoms of cocaine, such as fatigue, dysphoric mood, increased appetite, and psychomotor retardation, markedly differ from cannabis withdrawal symptoms, of which there are few except for excessive sedation or appetite stimulation. Abatement of withdrawal symptoms occurs when either abusers reuse the drug in small or large doses or titrate to another drug expected to produce commensurate effects. Users under court order to "drop urine samples" (urinalysis), for example, may cease cocaine use but increase their intake of alcohol for its stimulant properties. In this way, body cravings for peak effects remain the same when satisfied by an equivalent-acting drug.

## 1. CONCEPT OF DRUG DEPENDENCE

Drug dependence is a clinical term denoting significant distress or impairment in cognitive, behavioral, and physiological symptoms of users consuming large amounts of legal or illegal compounds. "Impairment," conceptually, considers the abuser to persist habitual drug ingestion despite realization of the untoward physical or psychological effects produced. Adverse effects are manifold but functionally are uniform. First, abusive use increases in quantity and frequency to match the gratifying or intoxicating effects desired. Mathematically, multiple daily routines build rapid metabolic adjustment or "tolerance" to the substance. The demand for more substance is higher to compensate for the body's adjustment (tolerance) and ensure euphoric or intoxicating sensations. Tolerance develops into dependence in two ways: physically and psychologically.

A physical dependence defines when substance intake levels greatly increase to maintain a desirable, sensory peak effect and prevent withdrawal symptoms. A psychological dependence is not biochemical in origin but contains a learning cycle of achieving optimal levels of functioning while intoxicated. Stimulant properties of high alcohol consumption, for example, may enable socially heroic, gregarious, or flirtatious behaviors in people who are introverts when sober. Withdrawal effects occur due to abrupt cessation of either physical or psychological dependence. Discontinuation causes adverse cognitive, behavioral, and physiological changes due to the decline of blood and tissue concentrations and musculoskeletal constrictions and dilations.

Withdrawal symptoms for alcohol, in particular, are classified as delirium tremens (DTs) because characteristic signs of delirium (e.g., hallucinations, delusions, and autonomic hyperactivity) are present.

### 1.1. Dependence versus Abuse

Drug dependence differs from drug abuse in three ways. First, drug dependence consists of clearly measurable periods of tolerance and withdrawal, whereas drug abuse sets the occasion for tolerance and withdrawal. Abuse means there exist early warning signs of life impairment predictive of physical and psychological dependence. Adults arrested for driving under the influence of alcohol, for instance, often face fears that their uncontrolled drinking may result in fatal vehicular accidents. Second, drug-dependent users show marked cognitive, behavioral, and psychomotor retardation in their daily activities. Drug abusers, however, may or may not suffer apparent cognitive, behavioral, or psychomotor malfunctioning; they may mask intoxication symptoms by performing their jobs and vehicular operations adequately and interacting normally with their families. Third, the phenomenon of withdrawal syndrome applies only to drug dependence. Drug abusers are episodic and not routine users, and their consumptive rates are variable; intermittent and lesser quantity users build resistance to adverse cognitive and visceral repercussions and only suffer mild or no side effects after abrupt cessation of desired substances.

### 1.2. Primary Types of Drug Dependence

The *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition, text revision) nosology of drug dependence cites 11 subclassifications of substances: cannabis, opioids, nicotine, phencyclidine, sedative-hypnotics, alcohol, amphetamines, caffeine, cocaine, hallucinogens, and inhalants. Ingested routes of administration vary for each substance type, as do the dependent and withdrawal symptoms. For cannabis, oral or inhalant routes of intake of the psychoactive component tetrahydrocannabinol (THC) produce sedation, mild euphoria, and altered perception of time and sensations. Effects arise as the drug attacks the nerve cells in the brain and alters their function. Physical withdrawal effects are significant, despite the popular opinion that there are none, and include irritability, anger, depressed mood, headaches, and

restlessness. Similarly, psychological effects include anxiety, paranoia, hypersensitivity, and overeating.

Opioids, whether synthetics (e.g., Vicadon), semisynthetics (e.g., heroin), or synthetics with morphine-like action (e.g., codeine), are orally, inhalantly, and intramuscularly infused. Effects include analgesic, anesthetic, and suppressive changes, whereas withdrawal signs include psychomotor agitation, impaired judgment, drowsiness, and, in severe cases, respiratory depression. Nicotine is absorbed orally (smoking or “chewing”) due to the rapidity of intake and high nicotine content of the substance. Stimulant and nausea-inhibitive properties of dependence counter unpleasant withdrawal effects of insomnia, irritability, anxiety, concentration erosion, agitation, and appetite suppression. Phencyclidine (PCP or angel dust) grew from medical anesthetics in the 1950s to one of today’s street drugs, producing sedative and disinhibitive effects when taken orally or intravenously. Withdrawal effects are uncertain, but stages of dependence result in rage, violence, and bizarre personality defects. Sedative-hypnotics similarly alter mood, judgment, and sexual functioning. Short-acting sedative-hypnotics (e.g., diazepam) may incur weak withdrawal effects, whereas longer acting agents cause autonomic hyperactivity, hand tremors, insomnia, anxiety, and often nausea.

Alcohol dependence depresses the central nervous system (CNS). However, to many, mild sedation under low doses is perceived as excitatory, not inhibitory. At advanced or chronic stages, there is damage to the neural pathways associated with liver failure, pneumonia, hallucinations, delusions, or head trauma. One collateral disease, Korsakoff’s syndrome, arises from a thiamin deficiency from injury to the thalamus. In withdrawal, sudden termination of high alcohol intake levels can be devastating, producing an acute state of depression, convulsive seizures, nausea, agitation, and DTs.

Amphetamine dependence is bioequivalent to cocaine dependence. Both activate the CNS and generate intense but temporary anxiety, paranoia, and hypermania concurrent with aggressive, violent, or psychotic behaviors. Reversal of neuralgic arousal occurs in withdrawal phases, including lassitude, depression, weight loss, nightmares, and psychomotor retardation. Similarly, caffeine—the central ingredient in coffee, tea, colas, and most over-the-counter analgesics—ubiquitously tempts physical and psychological dependence. Substance-induced effects range from autonomic arousal and appetite suppression to inexhaustibility. Heavy doses of caffeine in excess of two or three cups of coffee daily result in restlessness, anxiety, insomnia, diuresis,

rambling speech, agitation, and possibly gastrointestinal disturbance. Eliminating caffeine “cold turkey” initiates symptoms of fatigue, anxiety, depression, or even nausea.

Unlike caffeine, hallucinogens atypically develop into physical dependence, whereas users may psychologically depend on hallucinogens for euphoric and psychedelic effects; consequently, cross-tolerance frequently exists with LSD, psilocybin, and mescaline. After stopping use, the long-acting effects of hallucinogens produce residual cravings and even persistent perception disorders or “flashbacks,” a symptom commonly found during overuse or intoxication. Inhalants, too, are mood-altering, psychoactive compounds. These volatile substances, which contain esters, ketones, glycols, and halogenated hydrocarbons (found, for example, in spray-can propellants and cleaner solutions), produce inhalable vapors absorbed orally or by olfaction, reaching the lungs, bloodstream, and neural sites very quickly. Opposite effects manifest during withdrawal, including irritability, sleep disturbances, tremors, and diaphoresis.

### 1.3. Medication Mismanagement

Drug dependence is not always a result of teen or adult street-addictive habits. Medicines prescribed by family physicians and psychiatrists for mental health illnesses, and used appropriately at first by consumers, may be mismanaged accidentally. For example, the cerebral stimulant Ritalin, given to children with attention deficit hyperactivity disorder, requires parents to be regimen conscience. Self-administering of pain killers (Darvocet and Vicodin) also requires adherence to a rigid timetable to prevent lethal intoxication. Despite copious compliance, mistakes do occur in caretaker-administered or self-administered doses, which cause physical tolerance. The five common miscalculation mistakes observed in clients are overuse, underuse, erratic use, contraindicated use, and abuse. Overuse is mistakenly taking several doses of a medication or taking the medication when it is not needed. Underuse includes both failures to fill prescriptions and forgetting to take the medication. Signs of underuse are the clients taking fewer doses than instructed to make the medication last longer or discontinuing use earlier than directed. Erratic use refers to the failure to follow physician or pharmacy instructions. This includes missing doses, taking double doses, taking doses at the wrong time, and confusion regarding which drug to take at which interval. Contraindicated use

occurs due to incorrect storage of medications, using outdated drugs, or not monitoring side effects. Errors in judgment pose higher physiological risks when clients unknowingly mix psychotropics with nonpsychotropic medicaments, such as anti-inflammatories, antihypertensives, or diuretics. The last miscalculation, that of abuse, is the intentional misuse of a drug for its secondary or tertiary properties either to hurt or help oneself. Suicidal patients overusing Tylenol with codeine may exceed the recommended dose and abuse the purpose of its analgesic properties.

### 1.4. Routes of Administration Accelerating Addiction

Substances ingested into the body pass through three stages before they are excreted: absorption, transmission, and metabolism. All these stages enable drugs to accelerate or decelerate through the bloodstream to the site of action. The speed with which drugs reach the site of action, called the fate of drugs, depends on how drugs enter the body or the route of administration. The three routes pertinent to drug dependence are oral, inhalation (nasal), and intramuscular. Orally taken drugs dissolve in the stomach and pass into the bloodstream through the stomach or small intestine. Solid drugs (pills and capsules) diffuse slower than liquid drugs but trigger desired effects rapidly. Inhaled drugs such as (powdered) cocaine enter the bloodstream through membranes lining the nose. The magnitude of peak effect is very fast but short in duration before excretion occurs. Consequently, abusers quickly establish a high tolerance, vigorously trying to replicate euphoric sensations they had during their first inhalant experience. By far the most accelerated drug passage to the bloodstream is intramuscularly (intravenously). Despite the obvious perils of unsanitary needles, abusers prefer injection because it bypasses the filtering process through different organs, which could retard onset of effect. Instead, the stimulated site of action is powerfully immediate, and equally brief, triggering higher and repeated injection doses for sustained pleasure.

## 2. CONTINGENCIES OF DEPENDENCE

The genesis of drug dependence is only partly physiological. Although drug tolerance builds body

resistance, tempting larger doses for desired effects, abusers shape their ritual habits through predictable phases of environment–behavior interactions. These phases, called contingencies, describe a unique learning relationship that the drug-using person has with his or her surroundings. Early teenage recreational use of alcohol or drugs, for example, may occur alone or with friends; this is called the setting. The teenager's disposition during these occasions is known as the set. Set and setting qualify abstractly as the place and individual emotions predisposing mild, moderate, or heavy substance use.

However, in every setting lies events. Events describe more operationally the context of settings, such as who is present, what they are doing, the time of day, and whether they are influential on the abuser. Events also name behaviors endemic to the abuser, such as deprivation and satiation levels (e.g., thirsty, hungry, full, high, and fatigued), sensory status (aroused, depressed, in pain, angry, and happy), and thoughts or expectations about the setting (plans to drink, abuse drugs, and sell drugs). Connections between setting events and behaviors create a special interdependence whereby as setting events change, they covary with changes in behaviors. In this way, settings events function like stimuli triggering not just one response but many responses in a response class. For example, meta-amphetamine (crystal) abusers may buy and use the drug at a nightclub and dance with friends who also use. Drug use and dancing with friends constitute the response class. However, setting events also demarcate drug use from drug nonuse behaviors. The crystal abuser may only engage in the drug when dancing but not when he or she is alone at home and not dancing.

### 2.1. Types of Drug Response Classes Conditioned

Differentially learning drug habits under different settings is the result of conditioning. Conditioning means that the outcome of drug use under one setting is reinforcing and under another setting is either neutral or aversive. Simply stated, crystal abusers enjoy themselves more when they are dancing at bars rather than getting high at home. At home, the effects of crystal are benign or stagnant (neutral), or withdrawal effects may cause depression (punishment). Condition outcomes vary by another property—the response class. Responses from buying the drug or alcohol to multiple consumptions of it span a series of separate

response units held together in one of two ways. Think of a constellation of stars strung together by a common link. As one star changes, it produces specific or broad effects on surrounding stars. First, response units can be concurrent. This means two or more response units occur simultaneously rather than in a linear fashion. Crystal use may accompany respiratory elevation, muscular contractions, glandular secretions, heightened sexual arousal, speech amplification, and accelerated dancing movements. Second, response units may be sequential. Here, responses occur logically in a row like a domino effect. As one response appears, it predictably evokes the serialization of several more responses coordinated in order. After using crystal, abusers know they will perform better dance steps or feel heroically uninhibited to flirt with the opposite sex. Response classes, either sequential or concurrent, are powerful in building drug dependence when they produce outcomes or “consequences” immediately reinforcing the final (terminal) response in a response class. When terminal responses produce aversive consequences, such as when the crystal abuser vomits violently or is robbed, drug dependence may diminish only for a short time. Drug habits regenerate rapidly because the physical cravings for the drug reactivate the same response class or a variation of the response class.

### 2.2. Topography and Function of Drug Response Classes

Response classes include the topography and function of behaviors. These indices measure behavior, purpose, and configurations predictive of whether drug abusers will increase or decrease consumption rates. Topography describes (i) how often the person abuses drugs (frequency), (ii) the response class (simultaneous, concurrent, etc.), (iii) how long the response class occurs (duration), and (iv) how the effects of the responses on others (magnitude). Function describes the consequences produced by the response class (reinforcing, punishing, or neutral) and whether other setting event variables interact with the response class to accelerate or decelerate occurrences. For example, topography indicates that a crystal abuser goes to a bar twice weekly, stays there for 4 h, dances with five partners, and gets irritable and aggressive toward uncooperative partners later in the evening. A functional explanation of the response class indicates that the crystal abuser dances elegantly by midevening, receives countless praise from admirers including the

partners toward whom he later makes sexual advances, and usually generates business from those wanting to learn to dance from him. In other words, functionally, drug dependence for this crystal abuser recurs partly from physical tolerance and largely from conditioned reinforcement outcomes of his drug response class.

### **2.3. Avoidance and Escape Effects of Drug Response Classes**

Reinforcement contingencies may increase a response class by enhancing the pleasure for drug abusers, as in the case of the crystal user described previously. Reinforcement contingencies may also be effective by removing displeasure (aversive consequences) for the user. Nicotine-dependent users, for example, may smoke cigarettes both to prevent or delay fatigue onset when staying up late and to eliminate anxiety produced by daily stressors. In the first case, nicotine functionally serves to avoid aversive consequences, whereas in the second case nicotine acts as a catalyst to remove, terminate, or escape from aversive consequences. Avoidance and escape patterns become viciously intertwined in the matrix of drug response classes and account for exacerbation of physical tolerance and resistance to withdrawal effects. For example, a three-pack-a-day cigarette smoker quit (escaped) a smoking-cessation program because the withdrawal effects reminded him of the chronic anxiety suffered during his traumatic, physically abusive childhood. He used to feel nervous when battered by his mother, and the same body sensations reemerged (without him being battered) during nicotine withdrawal. Another smoker sharply increased the number of cigarettes used from 10 to 20 per day to delay (avoid) confronting his spouse about divorce. In both cases, elevated nicotine consumption proportionally correlated with avoidance and escape patterns.

### **2.4. Susceptibility to Drug-Dependent Response Patterns**

Response classes are like an immune system. Tightly connected or interdependent responses are resilient against delays of desired drug effects, withdrawal effects, or disruptions in the response class, such as when an alcoholic runs out of beer or cannot buy more because the local stores are closed. Response classes resist extinction and the drug abuser can freely adapt without suffering withdrawal effects. In contrast, weak

responses that are poorly interlinked become vulnerable to deterioration, which puts the abuser at risk of suffering withdrawal effects. Two properties of behavior causing susceptibility to degenerative drug response classes are deficits and excesses. Deficits consist of underdeveloped and immature responses falling below the social norm for adults or children and forcing abusers to expend extra effort to keep up with other people. Adults paralyzed by interpersonal deficits may be unassertive, shy, introversive, and awkward in social situations. They may drink several beers or smoke several joints (cannabis) to overcome their psychosocial deficits. However, deficiencies masked by polydrug dependence remain detrimental to abusers because their response class cannot withstand even the slightest disruptions. Two missed days of not smoking marijuana or one less beer consumed than normal may damage the response class and trigger physiological withdrawal. With excesses, the abuser overdoes responses. Excesses consist of abundant and obsessive behaviors aberrantly beyond the social norm and causing repulsion for onlookers. Compulsive exercisers, competitive athletes, academic overachievers, and perfectionists arouse constant excitement in their bodies. Excesses may be intensified or deintensified by drug dependence, but in either case the response class is fragile and susceptible to penetration. The mildest interruption of excessive behaviors (e.g., brought on by delayed access to the drug) can throw the drug abuser into a panic.

### **2.5. Effects of Titration on Drug Response Class**

One way that drug abusers with excessive and deficient response classes protract against involuntary detoxification is by a process called titration. Titration can be controlled through medical prescription or haphazardly induced by desperate abusers. Titration consists of phasing out one drug while replacing it systematically with another drug, bioequivalently calculated to avoid withdrawal effects. Heroin (opiate synthetic) addicts traditionally taper off addictive levels through a regulated dose of a physician-prescribed drug called methadone, which is also a synthetic opiate but chemically purified to ease the abuser through stages of drug reduction. Nonprescriptive practices of titration are dangerous but common. Cocaine-dependent users may consume alcohol to offset adverse side effects of “coming down from cocaine.” Consequently, the

ethanol supplement builds a tolerance in the abuser's body, eventually replacing the cocaine when cocaine supply becomes too costly or is inaccessible. Alcohol, functionally, offers risk-reduction effects for the cocaine abuser, but it does not replace the desired effects of cocaine.

When titration serves to transition one desired-effect drug to another desired-effect drug, similar to replacing heroin with methadone, the result is a substitutive effect. Substitutes occur intentionally and unintentionally. Intentional substitutes describe the cannabis abuser who abstains from marijuana to avoid THC detection on job-related random urinalyses; instead, he or she begins regular intake of two or three glasses of vodka per night to remain calm. Unintentional substitute abusers may overestimate their recovery confidence once they abstain from "bad" drugs for a period of time. Sober alcoholics, for example ("dry drunks"), may innocently consume 5–10 cups of coffee (caffeine elevation) per day, oblivious to the outcome. They may spontaneously regenerate the climax of sensory arousal previously achieved through alcohol. Renewal of these accelerated sensations (through coffee) rapidly restores body tolerance levels and makes the abuser unknowingly prone to alcohol relapse. Another common substitute misused by nondrinking alcoholics is ingestion of sweets (e.g., candy and pastries). Here, the sugar compound in sweets is bioequivalent to the sugar composition of ethanol, the deadliest antagonist to the body, causing cirrhosis. Reemergence of physical dependence on sugar, the powerful ingredient in ethanol, can swiftly sabotage sobriety by tempting nondrinkers to reconsume alcohol for satisfaction of the sugar craving.

### **3. FAMILY REASONS FOR RECIDIVISM**

Erosion of recovery from drug dependence occurs for many reasons. The causes of recidivism range from cue-induced cravings triggered by drug substitutes and drug-using settings to unsupportive family members. The latter represents an unusual paradox in substance abuse rehabilitation because, in the spirit of change, therapists assume the victimized family welcomes abstinent members. After all, strategic planning and confrontation of the drug-abusing member by key family members probably precipitated treatment. With disunity of the family threatened, abusers are told they

must receive therapy, submit to medication blockers (disulfiram, CB-1, naltrexone, methadone, etc.), or, if they refuse either of these choices, be legally or morally expelled from the family system. Ironically, compliant abusers who undergo the treatment and return to family members as caring, loving, and sober are greeted with consternation and resistance. Family members are unfamiliar with the reformed abuser's strangely alien personality and become dubious that the recovery will last. Healthy nondrug members also threaten the family equilibrium; a recovering member invokes positive and negative changes on other family members, challenging routines, rituals, traditions, and habits sanctioned for years and possibly underlying the main reason why the target member began abusing drugs. Uncertainty, skepticism, and fear become destructive agents in slowly unraveling the abstainer's confidence. Consequently, families become instigators of drug relapse in four ways, as discussed in the following sections.

#### **3.1. Family Distrust**

Distrust guards against repeated emotional and even physical injury in an unforgiving family environment. Family members who are uncertain when they will witness the next mood swing, aggressive outburst, or odd behavior of the abstainer are suspicious of every unexplained reaction. Self-defensively, family members such as children, a partner, or a spouse discredit proof of sobriety and question signs of progress. This interrogatory phase controls chaos in the family by preventing ignorance and blind faith—both of which were betrayed by abusers. Family members believe that by policing the abuser, they can monitor stability and instability of behavior and emotions and be alert to problems causing relapse. Abusers, however, object to this surveillance. They regard the distrusting family as coercive, suffocating, aggressive, and deliberately undermining their therapeutic program for recovery.

#### **3.2. Family Protectiveness**

Overtrusting may also produce untoward effects during recovery. As cohesion grows by abstainer and family members mending their wounds, the family supply of compassion goes into overdrive. Members feel compelled to help the abstainer reach recovery goals and be a productive contributor within the family. The intensity of help doubles when members fear drug recidivism and its intolerable effects on the family. They may spare the abstainer of major decision

making, financial responsibilities, and household chores and compare the drug recovery period to recuperation from surgery. Just as stitches may reopen, so it is that emotional stitches sewn together by a loose thread are perceived as fragile. The family caters to the abstainer's needs and places the recovery process on a pedestal. Caretaking, the term frequently applied to this altruism, mistakenly signals love and respect for the abstainer. The drawback of caretaking involves treating the abstainer as inactive. However, the abstainers must be active participants in altering family interactions using techniques that replace their drug abusive or dysfunctional behaviors. When families shower abstainers with unconditional nurturance, preventing natural opportunities for change, abstainers feel defeated, ineffective, and regress to pretreatment patterns.

### 3.3. Newly Surfaced Resentment

Family member anger kept repressed during years of the abstainer's drug dependence may surface unexpectedly when the abstainer first returns from inpatient rehabilitation or enters outpatient treatment. Previously inhibitive children, for example, may declare their resentment as victims of the abstainer's long-standing reign of terror. Embittered spouses may confess their infidelities, express thoughts of divorce, and lambaste abstainers for ruining precious years of their lives. Catharsis explodes by unloading a Pandora's box full of hatred against the abstainer with the hope of inciting debilitating guilt. Instead of shame, abstainers perceive this emotional release as a terrorist attack aimed intentionally at detonating aggression and inducing drug relapse. Two other effects on abstainers are equally disastrous. First, abstainers disenfranchise from the accusatory family harboring anger and fear. They attach onto family surrogate friends—many who abuse drugs—for emotional refuge. Second, the inescapable onslaught of allegations and criticism can result in abstainers spiraling into depression. Learned helplessness develops from suppression of their motivation and skill efforts.

### 3.4. Recovery as Threat to Family

Observed changes in the abstainer's behavior signal a necessary family adjustment in ways that are uncomfortable and thus avoided. Spouses, for example, are not accustomed to communicating with a healthy, sober spouse and thus resist talking. Children who never received verbal or nonverbal affection now rebuff hugs, kisses, money, or gifts given unconditionally by

abstainers. Displays of love and family involvement frighten members away from the abstainer, causing the family to aggressively resent this improvement. Resistance arises when the family must engage in behaviors about which they are ignorant, have deficits, or have incompatible behaviors.

Consider affection. Instructions of therapy may inspire a wife who is a drug recoverer to be more sensual and receptive to sex. Surfacely, her romantic advances toward her husband may be what her husband always wanted. However, after years of sexual denials, emasculative teasing, and erectile inhibitions, he finds his wife's sexual overtures repulsive. He also developed alternative sexual relief outlets, such as frequent masturbation, cybersex with chat-session partners, visitation to topless bars, and flirtation at work. Consequently, his ritual of incompatible sexual behaviors is so reinforcing that the husband refuses his abstainer wife's passion and enrages her into feeling rejected and abandoned. Angered abstainers surmising their efforts are in vein may rebel against the uncooperative family members and gradually cease the recovery goals.

## 4. CONCLUSION

Conceptually, drug dependence is higher on the physiological and psychological continuum of pathology than drug abuse. Tolerance forms when continuous drug ingestion increases body absorption, digestion, and excretion, all directly connected to body weight, age, set and setting, gender, race, body disease, diet, personality, and history of cross-addiction. Abrupt cessation of the drug causes immediate cognitive, emotional, and biochemical changes in the body defining withdrawal symptoms. Because drug use and nonuse habits are muticausal, variables influencing high-rate or low-rate use include routes of administration and the unique relationship drug behavior shares with the environment. Under reinforcing environments, for example, drug consumption rates increase depending on the pattern or response class of drug habits. Topographically, patterns composed of sequential or concurrent responses bond together tightly and resist extinction even when the drug abuser stops misusing the drug. Patterns of eating, sleeping, avoidance (and escape) rituals, and social routines that persist can still trigger cues inducing cravings, especially if the abuser replaces the drug with substitute compounds or titration or returns to settings that were conducive to drug use. Altering this response class thus represents the first goal of therapy, combined

with controlling the withdrawal symptoms. Moreover, conscientious therapists recognize that abuser recovery is not unilateral, wherein only the abstainer gets healthy. Instead, abuser recovery is bilateral, affecting a two-way exchange between family and abstainer. For this reason, barriers of family resistance against a reformed abuser, including distrust, protectiveness, resentment, and change, all must be addressed early in treatment to prevent hazards of drug recidivism.

### See Also the Following Articles

Alcohol Dependence ■ Diagnostic and Statistical Manual of Mental Disorders ■ Drug Abuse ■ Nicotine Addiction

### Further Reading

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: American Psychiatric Association.
- Ray, O. (1978). *Drugs, society, and human behavior*. St. Louis, MO: Mosby.
- Ruben, D. H. (1990). *The aging and drug effects: A planning manual for medication and alcohol abuse treatment of the elderly*. Jefferson, NC: McFarland.
- Ruben, D. H. (2001). *Treating adult children of alcoholics: A behavioral approach*. New York: Academic Press.
- Stout, C. E., Levitt, J. L., & Ruben, D. H. (Eds.). (1992). *Handbook for assessing and treating addictive disorders*. Westport, CT: Greenwood Press.



# Eating, Psychology of

Jennifer S. Mills and Julie Coleman

York University, Toronto, Ontario, Canada

1. Introduction
  2. Cultural and Environmental Influences on Food Selection
  3. Social Influences on Eating Behavior
  4. The Role of Personality in Eating Behavior
  5. Set Point Model of Weight Regulation
  6. Obesity and the Problem of Eating Too Much
  7. Overcoming Obstacles to Weight Control
- Further Reading

## GLOSSARY

**basal metabolic rate** The daily energy requirements for the body to perform its basic functions, expressed in terms of number of calories.

**body mass index (BMI)** An estimate of body fat or degree of obesity ( $BMI = kg/m^2$ ).

**dieting** The attempt to limit food intake in terms of amount or type of food for the purpose of weight control or restrained eating.

**food neophobia** A fear of trying new or novel foods.

**satiety** The feeling of fullness after eating.

**set point** An internal sensor of body fat to which metabolic processes will respond to achieve a balance of energy input and expenditure so as to defend a specific weight range for an individual.

Eating is a complex human behavior that is influenced by cultural, social, personality, and biological factors. Dieting, obesity, and overeating all are important social and health problems within applied psychology. The

absence of demonstrably successful weight loss programs, as well as controversy over the supposed health benefits of weight loss in the long term, casts doubt on the popular notion that dieting makes people healthier.

## 1. INTRODUCTION

One might think that a behavior so essential to our survival as eating would be fully explained through biology. It might seem that, like other essential behaviors such as sleeping and breathing, one could neatly explain the physiological mechanisms at work just before we put food into our mouths. However, decades of research into the biological mechanisms involved in eating have produced rather disappointing results. It seems that humans just do not do a very good job of eating in accordance with verifiable hunger and satiety cues. The biological perspective of eating posits that we should eat when our metabolic energy fuels are expended. However, one problem with this explanation of eating behavior is that humans often report feeling "hunger" despite ample energy stores in the form of body fat. Furthermore, most of us eat much more food than we require to remain alive. Another problem with the biological explanation of eating is that overall blood glucose levels (another index of metabolic energy) do not correspond well with reported hunger or (again) eating behavior. Finally, the hormone and neurotransmitter involvement in

hunger and eating is especially complex and not yet completely understood, with several hormones potentially involved (e.g., leptin, insulin, ghrelin, cholecystokinin). Rather than being determined by iatrogenic biological markers, humans' eating behavior seems to be largely determined by factors outside the individual. This article examines the psychology of eating and attempts to answer the question of why we generally tend not to eat according to hunger and fullness. Culture and environment, sensory factors, social factors, and even personality can influence food choice and eating behavior, and all of these factors take us further away from eating according to our internal hunger and satiety cues. The article also examines the problems of eating too much and obesity as well as what applied psychology can tell us about the solutions to these complex health problems.

## **2. CULTURAL AND ENVIRONMENTAL INFLUENCES ON FOOD SELECTION**

In general, our diets are made up of what is readily available for us to eat. To a large extent, our culture decides for us the type of food we can eat as well as the portion sizes we consider to be normative. In brief, we tend to eat what is put in front of us. For instance, the North American diet, replete with an overabundance of saturated fats and larger portion sizes as compared with its international counterparts, can at least partially explain North America's relatively high obesity rates. Experimental studies have confirmed that the more we are served, the more we eat.

Humans and animals alike experience food neophobia, that is, an intrinsic fear of trying new or novel foods. This fear is believed to serve a protective function in that it dissuades us from eating foods with which we are not familiar and that might not be safe. Factors that have been shown to increase food neophobia include knowledge that the food is of animal origin as well as trait and state anxiety. Factors that decrease our unwillingness to try new foods include modeling by others trying the new food, information about the nutritional value of the food, information about the good taste of the food, and mere exposure to the food.

In addition to the role of culture in determining our diets, personal learning experience dictates which foods we will and will not tend to include in our diets. Negative food associations are formed quite

readily. For example, a single episode of upper gastrointestinal illness (i.e., vomiting) after eating a particular food can be enough to produce a strong and resilient conditioned taste aversion. Presumably, this aversion has a protective function by keeping us away from a food that may make us sick again and endanger our health. Interestingly, illness after ingesting foods of animal origin is more likely to lead to the development of a conditioned taste aversion than is illness after eating fruits, grains, or vegetables.

Sensory qualities of foods themselves may override our bodies' hunger and satiety cues, resulting in our eating either more or less of the foods. For instance, we generally eat more as a function of the variety of foods we are offered, with total food intake increasing with greater variety. Even variety in the texture of a single type of food we are offered (e.g., different pasta shapes, ice cream vs milkshake) can increase our total food intake as compared with the presentation of a single texture of that food alone. The taste of a food also moderates our desire to eat more or less of it. Foods that are high in sugar and fat are intrinsically more appealing to the human palate, so it is not surprising that we are more likely to consume more of these foods when we are given the opportunity. According to the evolutionary perspective, this preference for sweet and fatty foods is due to those foods' scarcity during the hunting-and-gathering phase of human evolution and their energy content per gram of food as compared with foods high in protein. We can also lose our appetite for eating a specific food, a phenomenon known as sensory specific satiety, but then begin eating again when a new food is introduced. Most of us have experienced a resurgence of appetite for dessert, despite feeling full after finishing dinner. It is as though our bodies "get bored" of a particular food or taste after a while, prompting us to seek out more variety in our diets. This effect would presumably have a positive impact on our health by increasing the number of nutrients in our system.

Despite the cultural variations in cuisine and food selection, research has shown that humans are generally good at self-selecting a nutritionally balanced diet—at least they are as children. In a classic series of studies by Davis, human toddlers were offered a "buffet-style" meal plan and were allowed to choose the type and amount of food they ate. Her results showed that the toddlers self-selected foods that together made up a nutritionally balanced diet and resulted in a normal and stable pattern of weight regulation.

### 3. SOCIAL INFLUENCES ON EATING BEHAVIOR

Although we may start out with an intrinsic appetite for foods that will maximize our chances of being healthy as well as a general reliance on internal hunger and fullness cues, as we grow, our eating is more likely to be influenced by non-nutrient-related variables. One social factor affecting how much we eat is the number of people who are present in a given situation. We generally eat more as a power function of the number of people present. One proposed explanation for this social facilitation of eating behavior is related to competition for food resources; that is, we do not want to lose out on the opportunity to eat by letting our companions get to the food first. Another explanation is that the mere presence of people may provide distraction from our satiety cues. Of course, in real life, the presence of eating companions is often confounded by other factors that stimulate eating such as the variety and quantity of foods available (e.g., the “dinner party” effect). There is some evidence that the relationship between people who are eating together also moderates this social facilitation effect; that is, we are more likely to eat dessert when our eating companions are friends than when they are strangers. This suggests that the presence of others, particularly others with whom we are socially close, reduces inhibitions around eating. Interestingly, the idea that we eat more with more people present does not appear to hold true for obese and dieting individuals, who actually eat less when eating with others and are prone to splurge when eating alone. Presumably, for these individuals, an impression management motive—the desire to present themselves to others as being small eaters—overrides the effect of social facilitation of eating.

In addition to the mere presence of other individuals, the eating behavior of those individuals affects our own eating behavior. Modeling by others has a strong influence on eating behavior in that individuals tend to match the amount eaten by those with whom they are eating. For instance, even when we are very hungry, we will tend to match a small amount eaten by our companions.

### 4. THE ROLE OF PERSONALITY IN EATING BEHAVIOR

Personality also influences diet and eating behavior. The goal of weight loss or weight maintenance is the most common reason for efforts to control one’s eating.

Herman and Polivy’s “restraint theory” of eating behavior emphasizes the importance of chronic attempts at dieting as a predictor of food intake and has been very influential in the study of eating behavior. In research studies, restrained eaters (or chronic dieters) have been shown to demonstrate several differences with respect to their eating as compared with their unrestrained eating counterparts. Psychologically, restrained eaters exhibit higher body dissatisfaction, lower self-esteem, and a higher drive for thinness. Although dieters’ eating can be characterized essentially as an attempt to reduce energy intake (calories, fat, and/or portion size), they are especially prone to overeating. Laboratory studies of eating often employ a taste test methodology to demonstrate this effect. Participants are allowed to ad lib eat as part of an ostensible taste test during which they rate the different taste properties of the food (e.g., sweetness, saltiness). Their total food intake is measured by subtracting the weight of the plate(s) (or, in some cases, counting the number of food items eaten such as candies) after the experiment from the preexperiment weight. Sometimes, participants are given a “preload” or snack to eat before the taste test, and they are typically exposed to an independent variable of interest before the eating task. Disinhibited eating or overeating among dieters has been demonstrated to occur in response to ingestion of a high-calorie preload (e.g., milkshake, chocolate bar), stress, perceived weight gain, food cues, lowered mood, positive mood, exposure to thin media images, and even anticipated future overeating. With all of these and more potentially disinhibiting factors, it is not surprising that studies have shown that dieters tend to eat more overall, and also weigh more, than do nondieters. As noted previously, humans are inclined to select foods based on their sensory qualities. Obese and dieting individuals are particularly dependent on external and sensory cues in deciding how much of a food to eat, showing a heightened preference for good-tasting food over that of their normal-weight or nondieting counterparts.

Herman and Polivy’s “boundary model” of overeating posits that restrained eaters have a cognitively determined “diet boundary” that exists before the experience of satiety. This boundary represents what a dieter believes he or she should eat. Under normal conditions, food intake is expected to remain within the limits set by the diet boundary. However, once eating has surpassed the boundary limit (e.g., by ingesting a high-calorie preload such as a milkshake), eating will continue until eventual fullness is achieved. Herman

and Polivy labeled this event the “what the hell” effect; given that the dieter’s self-imposed diet boundary has already been surpassed, the dieter continues to eat until satiety and sometimes even beyond. Therefore, intended eating behavior for restrained eaters is cognitively determined by perceptions of what should be eaten in terms of type or quantity of food. In contrast, unrestrained eating is relatively more determined by an awareness of hunger or satiety cues.

## **5. SET POINT MODEL OF WEIGHT REGULATION**

There is solid evidence for a genetic factor in body weight determination. In terms of weight, adopted children tend to resemble their biological parents more than they do their adoptive parents. Identical twins, even when reared apart, are more similar in terms of weight than are fraternal twins or siblings. These examples suggest that genetics plays a role in determining one’s “set point,” that is, a type of internal sensor around which weight is regulated. When fat levels fall below this set point, the body responds by slowing metabolic processes so as to require fewer calories, thereby making the body more efficient in its energy expenditures and fighting against the depletion of fat stores. Keys and colleagues demonstrated this scenario in an experimental study of human semistarvation during World War II. Participants in the study were young men who agreed to lose 25% of their body weights for the study for the purpose of determining the best way in which to refeed starving people in war-torn Europe. On rations of 50% of their original daily caloric intake, the men initially lost weight quickly. However, over the 6-month starvation phase of the experiment, the men’s rate of weight loss reached a plateau, with many participants’ rations needing to be cut even further to achieve the desired weight loss. Keys’s starving participants exhibited a number of symptoms characteristic of the body’s corrective action of falling below its normal weight (and, presumably, below its set point in this initially nondieting sample). The men became irritable, miserable, hungry, and obsessed with thoughts of food. Their basal metabolic rates decreased significantly, indicating their bodies’ desire to conserve as many calories as possible. They became apathetic and lethargic, avoiding as much physical activity and energy expenditure as possible. They had very strong urges to overeat, and when they

were allowed to eat during the refeeding phase of the experiment, they overate and demonstrated a preference for high-calorie foods that would increase their fat stores most quickly, a situation that is consistent with the predictions of set point theory.

In another study of the effects of overeating that also found evidence supporting the set point theory, Sims and colleagues recruited prison inmate volunteers to gain 25% of their body weights as part of an experiment. The men’s food intake was increased, and their physical activity was decreased and controlled. At first, the volunteers gained weight easily, but the rate of weight gain soon slowed and their daily caloric intake had to be increased for them to continue gaining. One man did not reach his goal weight even though he consumed more than 10,000 calories per day. Once the men reached their goal weights, the men continued to require extra calories to maintain their higher than normal weights. During the experiment, these volunteers began to be repulsed by food and had to force themselves to eat. Many considered dropping out of the study due to the discomfort of overeating. Physical responses included an increased basal metabolic rate and profuse sweating, indications that their bodies were eliminating the excess calories they were ingesting. After the experiment, there were differences in how quickly the men returned to their preexperiment weights, with two participants (both with family histories of obesity) never reaching their original weights.

In sum, there is evidence that our body weights and shapes have a genetic component, but changes in eating behavior, hormones, or other variables can also produce weight gain or loss. Because of this apparent contradiction, some researchers have abandoned the set point theory of weight regulation in favor of a “settling zone” alternative. In 2002, Levitsky argued that within the biologically determined range or zone of body weight, actual body composition may “settle” at a value determined by behavior. Furthermore, the width of one’s settling zone may also depend on individual factors such that some individuals will naturally experience greater weight fluctuations over time than will others.

## **6. OBESITY AND THE PROBLEM OF EATING TOO MUCH**

Although genetics can predispose individuals to higher than average weights, there is no doubt that overeating can contribute to obesity. Because of societal pressure

to be thin and the stigma against overweight persons in many parts of the world, the obese and dieting populations overlap considerably. Therefore, many of the disinhibiting factors discussed in the preceding section operate regularly on the eating behavior of obese persons, leading to frequent episodes of overeating. In the preceding section, it was pointed out that overeating can occur beyond the point of satiety and until stomach capacity is eventually reached. Binge eating consists of eating episodes that involve consuming large quantities of food (e.g., twice a normal serving) combined with a feeling of being out of control. For binge eaters, food intake may continue until they reach capacity, and there is never really a sense of feeling as if they have eaten enough to satisfy their hunger. Binge eating combined with compensatory behaviors, such as self-induced vomiting, excessive exercise, extreme food restriction, and laxative abuse, is the hallmark of bulimia nervosa, an eating disorder. "Binge eating disorder," although not currently a formal diagnostic category, is characterized by regular binge eating episodes without attempts at compensation and is estimated to be highly prevalent in the obese population.

There is little consensus on the technical definition of obesity. Traditionally, charts that provide normal weight ranges for various heights and body frame sizes were the standard tool for determining normal weight and overweight. These charts were originally published by the Metropolitan Life Insurance Company in 1959, were based on associated mortality rates, and have undergone revisions during more recent years. Another method that gained popularity during the 1990s is the body mass index (BMI), which is defined as  $BMI = \text{kg}/\text{m}^2$ . A BMI of 20 to 25 is considered to be medically ideal, and obesity is defined as a BMI of 30 or more. However, unlike the Metropolitan Life Insurance charts, BMI does not take into account gender, age, or body frame. Body fat distribution can be estimated through one's hip-to-waist ratio. A low hip-to-waist ratio indicates that the person is carrying fat around their midsection, whereas a high hip-to-waist ratio indicates a relatively small waist and larger hips. The skinfold technique measures the thickness of a pinch of skin, usually under the upper arm. Despite the convenience and low cost of all these techniques, none of them is very accurate at measuring body fat. The most accurate method for assessing body fat is the water displacement method, whereby the individual is immersed in a tank of water and the displaced amount of water is measured. Obviously, this method has its practical drawbacks.

Obesity is unquestionably socially undesirable in many cultures today. But is it actually unhealthy? Unfortunately, the research is unclear as to how much weight gain is associated with increased health risks such as hypertension and cardiovascular disease. Being obese ( $BMI > 30$ ) has reliably been shown to put people at a significantly elevated risk for all-cause morbidity, premature death, type 2 diabetes, gallbladder disease, high blood pressure, sleep apnea, respiratory problems, liver disease, osteoarthritis, reproductive problems (in women), and colon cancer. However, those in the overweight category ( $25 < BMI < 30$ ) might not experience much more risk than do those in the normal weight range. Furthermore, weight loss itself can be a health risk for some. Most experts agree that there is a U-shaped relationship between weight and mortality and that a person's optimal weight range is uniquely determined by his or her own biology, varying with age and even ethnic background.

## 7. OVERCOMING OBSTACLES TO WEIGHT CONTROL

Of course, the most common method for weight loss is caloric restriction through dieting. However, most normal-weight dieters would be healthier if they did not diet at all. Dieting has psychological costs (e.g., irritability, hunger, food obsession), may be ineffective in improving health, and can worsen body dissatisfaction. Anorexia nervosa, another eating disorder, is characterized by extreme food restriction despite very low weight. The dieting of anorexics typically has a compulsive quality to it, and anorexia nervosa is routinely a difficult disorder to treat.

There are several obstacles to weight loss through dieting. A person's genetically influenced set point may be higher than his or her desired weight, and efforts at reducing caloric intake may be counterproductive and result in decreased metabolism and increased fat stores. Dieting individuals show an increased preference for restricted or forbidden foods and are more sensitive to sensory food cues, eating more good-tasting foods—foods that are also likely to be higher in sugar and/or fat content. Emotional states may lead to overeating or binge eating. Finally, calorically dense foods are all too available and convenient, leading to unbalanced food choices.

With an increasingly thin and lean physical ideal for both men and women, it is not surprising that there has been a rise in more drastic weight loss methods over recent years. All of the drastic weight loss methods are

**TABLE I**  
**Suggestions for Successful Weight Control**

- 
- Spread daily caloric intake evenly throughout the day. Do not skip meals.
  - Try to eat only in a regular place and at regular times.
  - Observe suggested serving sizes.
  - Avoid buffets and second helpings.
  - Eat slowly.
  - Eat in the company of others.
  - Do not read or watch television while eating.
  - Limit intake of high-calorie foods.
  - Allow for variety. Do not label any foods as “forbidden.”
  - Do not try to lose weight too quickly.
  - Find non-food-related ways in which to cope with negative emotional states.
  - Combine diet with exercise.
- 

unhealthy. Extreme restriction and elimination of a food category from an individual’s diet are nutritionally ill advised. Fasting, surgery, and diet drugs have significant and sometimes fatal health risks associated with them. Even when weight loss is achieved (and regardless of the method used), weight loss is very difficult to maintain in the absence of permanent lifestyle changes and good nutritional habits.

However, despite the low success rates of weight loss methods in general, most experts agree that the way in which individuals select foods and eat can have a positive impact on their weight and health. **Table I** displays suggestions for achieving a healthy and stable weight. The goal of these behaviors is to bring one’s eating more in line with bodily cues of hunger and satiety and to try to override many of the psychological factors that can lead to overeating.

In addition to these practical ways in which to make eating more in tune with the body’s needs, cognitive factors can be modified to reduce a reliance on dieting. Specifically, there is evidence that many dieters demonstrate unrealistic expectations for weight loss. They believe that profound and positive changes will result from losing weight in terms of their success in

personal, social, and/or professional realms. In reality, weight loss rarely brings about the type of happiness that dieters anticipate. In addition, many chronic dieters demonstrate “false hope beliefs” around dieting. That is, despite numerous failed attempts in the past, these dieters believe that the next diet will surely work. Previous diet failures are frequently attributed to lack of will power or perseverance on the part of the dieters rather than to the intrinsically difficult nature of successful long-term weight loss. Not only is this belief incorrect, but it can also be demoralizing, often exacerbating the poor self-image that precedes dieting efforts.

### See Also the Following Articles

Health and Culture ■ Health Psychology, Cross-Cultural

### Further Reading

- De Castro, J. M., & Plunkett, S. (2002). A general model of intake regulation. *Neuroscience and Biobehavioral Reviews*, 26, 581–595.
- Drewnowski, A. (1995). Energy intake and sensory properties of food. *American Journal of Clinical Nutrition*, 62, S1081–S1085.
- Fairburn, C. G., & Brownell, K. D. (2003). *Eating disorders and obesity: A comprehensive handbook* (2nd ed.). New York: Guilford.
- Hetherington, M. M. (1996). Sensory-specific satiety and its importance in meal termination. *Neuroscience and Biobehavioral Reviews*, 20, 113–117.
- Levitsky, D. A. (2002). Putting behavior back into feeding behavior. *Appetite*, 38, 143–148.
- Martins, Y., Pelchat, M. L., & Pliner, P. (1997). “Try it; it’s good and it’s good for you”: Effects of taste and nutrition information on willingness to try novel foods. *Appetite*, 28, 89–102.
- Ogden, J. (2003). *The psychology of eating*. Oxford, UK: Blackwell.
- Polivy, J., & Herman, C. P. (2002). Causes of eating disorders. *Annual Review of Psychology*, 53, 187–214.
- Stunkard, A. J., Sorensen, T. I. A., Hanis, C., Teasdale, T. W., Chakraborty, R., Schull, W. J., & Schulsinger, F. (1986). An adoption study of human obesity. *New England Journal of Medicine*, 314, 193–198.



# Economic Behavior

Michael W. Allen

Griffith University, Brisbane, Australia

Sik Hung Ng

City University of Hong Kong, Kowloon,  
Hong Kong, China

1. The Nature of Humankind in Economic Theory and the Need for Psychology
  2. Microeconomics and Psychology: The Behavior of Consumers, Investors, and Workers
  3. Macroeconomics and Psychology: Perspectives on Inflation, Unemployment, and Economic Growth
  4. Conclusions and Future Directions
- Further Reading

reviews some of the main areas of interest to economic psychology, including those at the microeconomic level (e.g., consumption, investing, working) and macroeconomic level (e.g., rates of inflation, unemployment, economic growth). The convergence and deviation of economics and psychology are also discussed.

## GLOSSARY

**economic behavior** The behavior of individuals that involves economic decisions and the causes and effects of those decisions.

**macroeconomics** The study of the overall levels and changes in societal economic activity.

**microeconomics** The study of the economic behavior of individuals and small groups.

**neoclassical economics** Traditional economic models that assume that agents are rational and act out of self-interest (narrowly defined).

**rationality assumption** The position that individuals choose alternatives that maximize expected utility.

**self-interest assumption** The view that individuals choose options of greatest benefit to themselves.

Economic psychology is the study of the behavior of individuals that involves economic decisions and the causes and effects of those decisions. This article

## 1. THE NATURE OF HUMANKIND IN ECONOMIC THEORY AND THE NEED FOR PSYCHOLOGY

One might expect that the disciplines of economics and psychology are integrated because both study behavior. Historically, however, economics and psychology have been distinct, and the few intellectual contacts between them up to 1920 were unhappy ones for economists who, like Jevons, tried to exploit psychological principles (e.g., pleasure, pain) but failed to add much to economic theories. Apart from the unhappy experiences, another reason for the disciplinary separation is that, in the interest of precision, economics has chosen mathematics as its methodology and metaphor. Notwithstanding the separation, economics has made crucial assumptions about the nature of human decision making, which is also the subject matter of psychology.

One such assumption in neoclassical economics is “rationality,” in which individuals are said to choose alternatives that maximize expected utilities. In

particular, the neoclassical view is that individuals rank all possible alternatives according to how much satisfaction they will bring and then choose the alternative that will bring the most satisfaction or utility (provided that they believe the alternative is attainable). However, psychological research has shown that people often do not choose the alternative with the greatest utility due to limited information processing capacity, motivational biases, and other psychological processes.

Examples of biases and errors in judgment constraining rationality are numerous. For instance, Simon argued that people often use cognitive heuristics, or shortcuts, to reduce the amount of information processing required for making a decision, but this may lead them to select an alternative that is merely satisfactory rather than the alternative with the highest utility (a process Simon described as “satisficing” and a form of bounded rationality). People are also known to give different weights to costs and rewards that are of the same magnitude (i.e., the opportunity cost effect). They are sensitive to framing and context, misjudge prior probabilities, and ignore sample size (i.e., the representativeness bias). They have an intransitive choice set (e.g., they might prefer  $A > B$  and  $B > C$  and yet they prefer  $C > A$ ) and assign a higher value to an item once it becomes their possession (i.e., the endowment effect). Their preferences are present biased (i.e., a slight delay in gratification is perceived to be worse if the delay occurs today than if a delay of the same magnitude occurs in the distant future), as are their personal experiences (i.e., greater weight is assigned to recent experiences). Finally, individuals are sensitive to social comparisons and often judge their economic well-being in social terms rather than purely economic terms.

A second assumption in neoclassical economics, closely allied to the first, is that individuals act out of self-interest (narrowly defined). That is, people will choose the option that will give them the most. However, psychological research challenges this assumption as well. For instance, people are affected by satiation and personal considerations of fairness (e.g., when asked to split a pile of money between themselves and a stranger, many people choose a 50–50 split, whereas neoclassical economics would predict a 100–0 split).

Psychological research suggests that many individuals are unlikely to meet both the rationality and self-interest assumptions of neoclassical economics (although psychology and economics are not as far apart as is implied herein). Some economists suggest that the rationality and self-interest violations in neoclassical economics can be dealt with by redefining rationality

and interest. Other economists argue that these breaches can be ignored because individual differences and biases “wash out” at the macro level of analysis.

An alternative strategy pursued by economic psychology is to model economic behavior in a way that is psychologically realistic. In particular, economic psychology is the study of the behavior of individuals that involves economic decisions and the causes and effects of those decisions. Most decisions are economic ones in some form, making economic psychology a broad field. Many of the areas of interest to economic psychology are also key topics of research in other fields such as applied psychology, economics, sociology, and political science. This article provides a brief overview of the micro- and macroeconomic issues of interest to economic psychology.

## **2. MICROECONOMICS AND PSYCHOLOGY: THE BEHAVIOR OF CONSUMERS, INVESTORS, AND WORKERS**

The psychological approach to microeconomics examines the behavior of individuals and small groups with regard to diverse areas such as saving, buying, investing, gambling, working, giving, and paying taxes. For instance, theories of saving suggest that people choose how much to save by considering the amount saved by a reference group, choosing a fixed percentage of their disposable income, or varying their saving and spending of savings at different stages of their life cycle. This section reviews the behavior of individuals as consumers, investors, and workers.

### **2.1. Consumers**

Consumer behavior is one of the most extensively researched areas in microeconomics. Initially, the field was dominated by approaches based on neoclassical economics. The most fundamental of these early approaches was expected utility theory, which argued that consumers have complete information about each product, evaluate that information in a deliberate and exhaustive manner, and ultimately choose the product that has the greatest utility (subject to constraints of money, availability, etc.). Critics argued, however, that it is unrealistic to assume that consumers choose the brand with the maximum utility because, as noted previously, individuals have limited information

processing and make errors in judgment. Consumers are also unlikely to have all of the information about all brands, and the information they do have is subject to perceptual and motivation biases. For instance, one well-documented effect is that consumers place higher value on products that appear to be in short supply (a phenomenon that Brehm in 1966 explained as psychological reactance to the loss of freedom). Similarly, consumers' reference point for deciding whether the price for a product is fair is not only the absolute price, as neoclassical economics posits, but also the change in price and frame of reference (i.e., prospect theory).

Consequently, the field of consumer behavior now largely draws on psychological insights. That is, although some recent consumer decision-making models do leave room for extended rationality (e.g., expectancy value theory), other models recognize that consumers do not maximize expected utility and might simply compare brands on a single attribute (e.g., the lexicographic model). The neoclassical economic approach to consumer behavior also assumes that consumer preferences are stable and makes no mention of where consumers derive their preferences in the first place. Hence, consumer socialization and social influence are major areas of study by economic psychologists. Psychological approaches to understanding consumer behavior also investigate the roles of emotions, motivations, lifestyles, and the self-concept that have largely been absent from the neoclassical view of the consumer.

## 2.2. Investors

A common model in neoclassical economics for predicting how individuals make investment decisions involving risk and uncertainty is the expected utility theory, according to which the probability of an outcome is multiplied by the utility of the outcome (and is added across all outcomes). For instance, an individual deciding whether to invest in a new venture would gauge the probability of the venture's success times the financial payoff of that success. Hence, expected utility theory assumes that people objectively and rationally judge risk. In contrast, psychological research has shown that decision makers consistently over- or underestimate probabilities, interpret independent outcomes as dependent ones, and weight gains differently from losses of the same magnitude (i.e., prospect theory). Moreover, personality differences mean that individuals differ in their perceptions of

risk, in their willingness to take risks, and in their ability to delay gratification (which investing necessitates).

Social psychological processes also underpin risk taking. For instance, the "risky shift" effect, in which group discussion shifts individuals toward taking greater risk in decisions where the initial preference already favors risk, indicates clearly the transformational power of group dynamics in decisions that involve risk taking. This may help us to understand why overvalued markets continue to rise (and why undervalued markets continue to fall).

## 2.3. Workers

Another area of microeconomics studied by economic psychologists is how individuals behave at work. In the traditional economic view, the goal of an organization is profit maximization. On the one hand, this maximization might be achieved through a rational mathematical calculation involving the selling price, the output level, and costs such as labor, materials, and transportation. On the other hand, labor is about people and so is inherently psychological. That is, although the number of hours an employee works is prescribed by the firm, how much effort the worker exerts is dependent on psychological factors such as motivation, conditioning and reinforcement, social context, and ability.

For instance, needs-based approaches to work motivation suggest that humans have certain deeply felt needs and that each individual's motivation to perform depends on whether the specific features of the job fulfill his or her needs. A second approach to understanding the amount of effort an individual exerts at work is reinforcement theory (and other learning and conditioning principles). According to this view, effort at work is a result of reward for positive performance and/or punishment for negative performance. A third approach examines how an individual's work behavior and effort are influenced by the social context. For instance, research has demonstrated a "social loafing" effect whereby individuals exert less effort when they work in a group than when they work alone (although the effect varies by task difficulty and cultural values).

In addition to the issue of whether people are motivated to perform at work, the effectiveness of the workforce depends on workers' job-related abilities. Here, psychological research finds scientific ways in which to match the skills, education, and training of individuals with the requirements of the job. Economic psychology also studies the effect of unemployment on the

individual. In particular, it examines the negative and positive sides of unemployment, the psychological stages an individual goes through when losing a job, and the link between unemployment and health.

### **3. MACROECONOMICS AND PSYCHOLOGY: PERSPECTIVES ON INFLATION, UNEMPLOYMENT, AND ECONOMIC GROWTH**

Macroeconomics studies the overall levels and changes in societal economic activity. Topics include economic growth, inflation, recession, unemployment, and income distribution. Often, economists explain the overall levels and changes in economic activity as issues involving the demand for goods and services, the money supply, and/or (more recently) the supply of goods and services. For instance, macroeconomic approaches to solving inflation are to decrease the demand for goods and services, decrease the money supply, and/or increase the supply of goods and services. However, each of these approaches overlooks, to varying degrees, the psychological aspects of economic activity and the position of that action in a larger sociocultural environment.

#### **3.1. Inflation**

The neoclassical economic view asserts that individuals objectively gauge the rate of inflation. This “expected inflation” allows individuals to make decisions about saving, purchasing now, and going into debt. However, psychological research has revealed that people often do not accurately remember past prices and make poor guesses about the future rate of inflation. Furthermore, wage increases during times of high inflation may be perceived as positive by workers because “money illusion” leads them to evaluate wages on their nominal value rather than their real value. On the other hand, people assess their satisfaction with wages not only by comparing wages with an absolute standard but also by comparing them with a reference group. Hence, people may demand increases in wages simply to keep up with their neighbors. This illustrates that inflation is, in part, a social psychological phenomenon.

In 1984, Maital and Maital also suggested that inflation can be profoundly psychosocial and can be researched using the prisoner’s dilemma experimental procedure. In this case, an individual can gain

financially if he or she demands a large wage increase and raises the prices of any products the individual may sell. However, if everyone in society does the same, the result is inflation. In addition to seeking to contribute to an understanding of the causes of inflation, economic psychology examines the psychosocial effects of inflation on the individual such as helplessness and perceived lack of personal control.

#### **3.2. Unemployment**

Whereas the effects of unemployment on the individual are an area of concern in microeconomic psychology, how society explains unemployment is a focus in macroeconomic psychology. Three broad types of social explanations have been identified: individualistic (e.g., laziness or lack of effort on the part of unemployed people, fussiness of unemployed people when it comes to taking jobs, poor education, qualifications of unemployed people), societal (e.g., policies of the government, less competitive industries that go bankrupt, prior overmanning in industry, weak trade unions), and fatalistic (sickness and physical handicaps among unemployed people, bad luck). Individualistic explanations are most common among conservative (right-wing) voters, whereas societal explanations are stronger among labor (left-wing) voters. This underscores that individuals’ beliefs about how the economy functions is less about economic models and more about morality and political persuasion.

#### **3.3. Economic Growth**

One final macroeconomic area of current interest is how cultural values and personal attitudes are linked to economic growth. The strict economic view is that the economic growth of a nation stems from conventional economic factors such as labor force and capital. However, Reynaud argued in 1981 that these factors per se cannot fully account for economic growth and that psychosocial factors must also be considered. In support of Reynaud’s position, some theorists claim that cultural values such as the Protestant work ethic and individualism produced the economic growth seen in Western nations (i.e., a cultural determinism claim). On the other hand, other theorists suggest that economic development leads to changes in cultural values (i.e., economic determinism).

In addition to the issue of economic determinism versus cultural determinism, researchers are divided regarding the extent to which human values are

aligned with economic growth and development. One form of modernization theory suggests that development proceeds in a series of linear stages; hence, as nations develop, their cultural values will become more similar to those of developed nations. However, the extent that the values of nations are converging is unclear because the values that predict economic growth are shifting as Asian economies expand.

#### 4. CONCLUSIONS AND FUTURE DIRECTIONS

Despite the polarized way in which economics and psychology have been presented here, the two disciplines are in fact deeply interconnected. Central and historically significant branches in psychology also assume that individuals are rational and act out of self-interest. For instance, attribution theory assumes that people are naive scientists who seek rational explanations for their own and others' behaviors, behaviorism assumes that organisms act out of self-interest by seeking rewards and avoiding punishments, and social exchange theory suggests that individuals compare their inputs and outputs with those of others to gauge whether the relationships are satisfactory.

Similarly, economics is more psychological than is reported here, as evidenced by the 2002 Nobel Prize in economics being awarded to Daniel Kahneman for his work on human judgment (which integrates economics and psychology). Another example of psychological constructs being central to economic analysis is Katona's use of the attitude construct in his consumer optimism–pessimism index.

Notwithstanding that some models incorporate both psychology and economics, the two fields need greater integration. For instance, Lea and colleagues argued in 1987 that economic psychology should examine how the individual affects the economy and vice versa (i.e., dual causation). More recently, in 1995, Lewis and colleagues called for a better understanding of post-modern economic behavior and how economic behavior has communicative and identity functions.

One way in which to integrate psychology and economics is to propose models that include the neoclassical economic view and the psychological perspective within the same conceptual framework. These models would accept that the rationality and self-interest assumptions of economics and the

alternative explanations offered by psychology both are applicable, depending on the context and the individual. For instance, Allen and Ng's dual-process model of how human values guide consumers' product decisions, as described in their 1999 article, suggests that consumers make both rational and "nonrational"/emotional choices, depending on the type of product or on the motivation driving the consumer. That is, consumers motivated by a need for instrumentality will make a rational, attribute-by-attribute judgment of the product's utility, whereas those motivated by a need for self-expression will make a holistic, emotion-laden, and spontaneous judgment of the product's image. Similarly, hybrid approaches were proposed in the important area of social welfare by Taylor-Gooby in 2000 and in how people's views about the economy affect their political party choices by Allen and Ng in 2000.

#### See Also the Following Articles

Advertising and Culture ■ Advertising Psychology  
■ Consumer Psychology ■ Values and Culture

#### Further Reading

- Allen, M. W., & Ng, S. H. (1999). The direct and indirect influences of human values on product ownership. *Journal of Economic Psychology*, 20, 5–39.
- Allen, M. W., & Ng, S. H. (2000). Self-interest, economic beliefs, and political party preference in New Zealand. *Political Psychology*, 21, 323–346.
- Antonides, G. (1996). *Psychology in economics and business: An introduction to economic psychology*. Dordrecht, Netherlands: Kluwer Academic.
- Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.
- Bromley, P., & Curley, S. P. (1992). Individual differences in risk taking. In J. F. Yates (Ed.), *Risk-taking behavior* (pp. 87–132). New York: John Wiley.
- Earl, P. E. (1990). Economics and psychology: A survey. *Economic Journal*, 100, 718–755.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Katona, G. (1960). *The powerful consumer*. New York: McGraw-Hill.
- Kemp, S. (1987). Estimation of past prices. *Journal of Economic Psychology*, 8, 181–189.
- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social

- loading. *Journal of Personality and Social Psychology*, 37, 822–832.
- Lea, S. E., Tarpy, R. M., & Webley, P. (1987). *The individual in the economy: A survey of economic psychology*. Cambridge, UK: Cambridge University Press.
- Lewis, A., Webley, P., & Furnham, A. (1995). *The new economic mind: The social psychology of economic behavior*. Hertfordshire, UK: Harvester Wheatsheaf.
- Maital, S., & Maital, S. (1984). Psychology and economics. In M. L. Bornstein (Ed.), *Crosscurrents in contemporary psychology*, Vol. 3: *Psychology and its allied disciplines* (pp. 55–88). Hillsdale, NJ: Lawrence Erlbaum.
- Mitchell, W. C. (1967–1969). *Types of economic theory: From mercantilism to institutionalism* (Vols. 1–2, J. Dorfman, Ed.). New York: Augustus M. Kelley.
- Ng, S. H. (1984). Social psychology and political economy. In H. Tajfel (Ed.), *Social dimension: European contributions to social psychology* (Vol. 2, pp. 624–645). Cambridge, UK: Cambridge University Press.
- Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review*, 46, 657–685.
- Reynaud, P. L. (1981). *Economic psychology: Overview of methods, applications, and theory* (S. E. G. Lea, Trans.). New York: Praeger.
- Simon, H. A. (1982). *Models of bounded rationality* (Vols. 1–2). Cambridge, MA: MIT Press.
- Taylor-Gooby, P. (2000). *Risk, trust, and welfare*. London: Macmillan.
- MacFadyen, A. J., & MacFadyen, H. W. (Eds.). (1986). *Economic Psychology: Intersections in Theory and Application*. Amsterdam: North-Holland.
- Webley, P., Burgoyne, C. B., Lea, S. G., & Young, B. (2001). *The economic psychology of everyday life*. East Sussex, UK: Psychology Press.



# Educational Achievement and Culture

David S. Goh

City University of New York, Queens College, Flushing,  
New York, USA

Harry W. Gardiner

University of Wisconsin, La Crosse, Wisconsin, USA

1. Introduction
2. Overview of Cultural Approaches to Teaching, Learning, and Achievement
3. Cross-Cultural Differences in Mathematics Achievement
4. Cultural Explanations for Cross-Ethnic Differences in Achievement in the United States
5. Ethnic Differences in Cultural Factors Affecting Achievement
6. Directions for Future Research on Cross-Cultural Achievement  
Further Reading

## GLOSSARY

**cultural difference theory** Theory that maintains that schools in the United States are based on the majority (European American) culture, which differs in many ways from the minority cultures; these cultural differences interfere with the learning of these students and affect their motivation to do well.

**deficit theory** The ethnocentric idea that students who do poorly in school are considered to be deficient in some way compared with those who do well in school; most social scientists now consider this idea to be out of date.

**direct participation** An educational environment in which parents are actively involved in helping their children with their schoolwork, maintaining good communications with the school, participating in school activities, and the like.

**ethnocentrism** The belief that one's own cultural ways are superior to those of other people.

**expectation theory** A belief that some children do poorly in school because their teachers have a low expectation for their success.

**genetic epistemology** The study of the development of knowledge.

**indirect support** An educational situation in which there is not direct participation of the parents in their children's education; rather, parents work to maintain a home environment that is highly supportive of their children's learning (e.g., arranging for ample time to study, paying for private lessons).

**individualized assessment** Evaluation of individual student needs, skills, and performance.

**involuntary minority groups** Those who are incorporated into the U.S. society against their will (e.g., African Americans).

**pedagogical model** A design of systematized learning or instruction with an emphasis on the goals, principles, and methods of teaching.

**standardized curricula and evaluations** A set of formally designed and accepted teaching materials along with methods for measuring and comparing performance on them.

**student-centered teaching** Nontraditional instruction in which primary responsibility for learning is on students (rather than on teachers); students work independently or in groups on specific assignments or projects.

**teacher-centered instruction** Traditional approach to learning in which teachers provide information, typically through lectures along with textbooks and/or assigned readings.

*triarchic model* A holistic approach to understanding differences in educational achievement that looks at multiple factors (e.g., school, family, student) that contribute to ethnic differences in educational achievement.

*voluntary minority groups* Those who come to the United States by their own choice (e.g., early Europeans, Asian Americans).

Culture plays a crucial role in educational achievement. Pedagogical models and practices tend to be culturally embedded. A significant portion of cross-cultural and cross-ethnic differences in achievement may be attributed to cultural factors.

## 1. INTRODUCTION

Educational achievement is an important concern for parents, educators, psychologists, policymakers, and others in cultures throughout the world. Many factors (e.g., pedagogical practices, educational resources, socioeconomic status [SES], parental education level) contribute to differences in the learning outcome of students. Culture has been found to play a critical role in educational achievement in cross-cultural and cross-ethnic studies. Significant differences in educational achievement have been reported among students from different countries as well as among different ethnic groups within the United States. For example, compared with many other countries, the United States has a wide range of ethnic groups that share within-group backgrounds, languages, values, beliefs, thinking, and norms of behavior. Available research has shown that some ethnic minority groups of students (e.g., African Americans, Hispanic Americans) achieve below the national average of all students. One exception is the high educational achievements demonstrated by Asian American students. These ethnicity-related differences have been documented in standardized achievement test scores, grades, high school graduation rates, dropout rates, and other measures of educational achievement. However, in understanding cultural or ethnic differences in educational achievement, it is important to note that there are vast variations among students within each cultural or ethnic group. Research findings that reflect between-group patterns in achievement may or may not apply to individual students within the groups. In other words, despite reported patterns of group differences, there are high-achieving students and low-achieving students within each cultural or

ethnic group. Therefore, it is important to avoid stereotyping in understanding educational achievement and culture. One should be wary of prejudging students based on their cultural or ethnic status.

This article first presents an overview of cultural approaches to teaching, learning, and achievement. Then, it discusses research on cross-cultural differences in mathematics achievement and the explanations for the differences. This is followed by a review of different theories that have been proposed to explain cross-ethnic differences in academic achievement in the United States. In addition, the article discusses ethnic differences on several cultural factors that are important to educational success. Finally, it recommends directions for future research on cross-cultural and cross-ethnic achievement.

## 2. OVERVIEW OF CULTURAL APPROACHES TO TEACHING, LEARNING, AND ACHIEVEMENT

North and South America represent an extremely diverse hemisphere characterized by struggles for quality and equality in education. In 2000, Mazurek and colleagues discussed educational practices and reforms in the United States, Canada, Brazil, Mexico, and Chile. They pointed out that in the United States and Canada, demands for reforms come from interest groups, the media, politicians, and parents organizations. In Brazil and Chile, national social/political movements and a return to democracy are behind educational reforms. In Mexico, reforms appear to be a response to a combination of social and political turmoil within the country.

In Western Europe, education has taken on a new priority involving technology and business orientation at all levels. In some countries, this includes a particular focus on vocational education to prepare students for entry into the workforce. Some countries (e.g., Switzerland, Sweden) have introduced a stronger emphasis on vocational training in their schools, whereas others (e.g., France, Spain) continue to focus on the teaching of traditional academic subjects. At the same time, a number of Middle Eastern and African nations (e.g., Israel, Palestine, Turkey, Pakistan, South Africa), characterized by some as “nations in turmoil,” are struggling to solve serious political, social, and economic problems while modernizing their countries and reforming their educational systems—a task that each country takes very seriously. At the postsecondary

level, most countries throughout the world are trying to determine what should be the proper direction and focus of their educational efforts.

### 2.1. Cultural Models of Schooling

Schools throughout the world are experiencing a period of rapid change and, in many cases, are finding it extremely difficult to achieve a balance among a number of critical concerns. Some of the issues that educators and schools are facing include uncertainty about what academic and cultural knowledge and skills will be needed by students in the future, wholesale revisions of curricula, experimentation in teaching strategies, the need for teachers and students to become aware of and competent in using new technologies, dramatic changes in bureaucratic and legislative policies and regulations, and increased demands on teachers.

With the exception of the education system in the United States, perhaps no other education system has been studied more intensively than that of Japan. In 2001, in a well-balanced presentation of the Japanese model of schooling, including its similarities to and differences with that in the United States, Tsuneyoshi characterized the American approach to education as one that places an emphasis on competitiveness, individual attention from teachers along with individual accomplishment on the part of students, development of cognitive abilities, and separation of teachers in terms of their disciplines. In contrast, the Japanese approach (particularly at the elementary school level) focuses on the “whole child”; close interactions between teachers and pupils for long periods of time in cooperative settings with attention to collective goals, tasks, and rewards; and efforts to provide the same or very similar treatment for all students. One advantage of the American approach that is seriously missing in the Japanese approach is the former’s attention to diversity and a sensitivity and concern for minority rights. On the other hand, the Japanese approach has a clear advantage in its ability to focus on the whole child within a community structure designed to provide close, cooperative, face-to-face student–teacher relationships.

### 2.2. Approaches to Teacher Training

There appear to be as many approaches to teacher training as there are teacher training programs. As several recent studies show, a majority of the world’s cultures have made substantive changes in teacher training and certification, whereas others are just beginning to do so. For example, in Brazil, a secondary education teacher

must pass a course of study corresponding to a teaching degree. In Russia, teachers are generally trained to teach one or two specific subjects, although many rural teachers can, and frequently do, teach between three and five subjects. An early unique feature of teacher training in Turkey was the establishment of “village institutes” where elementary teachers were trained to work with selected bright village children. Today, teachers are required to be university graduates, preferably from education faculties; however, they can receive teaching certificates if they graduate from other 4-year programs.

Yet in 1998, Murray strongly suggested that teacher education reform has largely failed, and he argued for the position that teacher education, developmental theory, and research are important and interlocking elements in bringing about a reform in school practices as well as the modification of school outcomes. It is for these reasons that Murray recommended that efforts to establish or change teacher education curricula should seriously consider the use of developmental psychology, especially pedagogical content knowledge and genetic epistemology, as a foundation.

In closing, research shows that pedagogical models tend to be embedded within cultures and that attempts to transplant them from one society to another, no matter how beneficial some may believe this will be, might not always be possible and, if attempted, will require serious planning, preparation, and attention to indigenous concerns.

Mazurek and colleagues, in discussing the debate over pedagogy, pointed out that while some countries are employing so-called progressive approaches such as student-centered teaching and individualized assessment, others are continuing to use traditional approaches such as teacher-centered instruction and standardized curricula and evaluations. Still others are experimenting with a wide variety of alternative and combined approaches. At the same time, these authors pointed to a common theme, namely that countries appear to be moving away from pedagogical strategies that are uniform and system-wide in their approach and toward those that are represented by greater diversity.

## 3. CROSS-CULTURAL DIFFERENCES IN MATHEMATICS ACHIEVEMENT

Research literature reports very few well-designed and carefully conducted cross-cultural studies in the area of mathematics achievement. Most notable are those carried out by Stevenson and colleagues, nearly all of which

showed American students lagging behind their counterparts in other industrialized countries. In 1990, in a classic study of math achievement among American, Chinese, and Japanese children, Stevenson and Lee attributed American children's poor performance to several factors, including a lack of emphasis on academic activities, the absence of a shared goal in academic achievement, an overestimation of children's accomplishments by parents and children themselves, low parental standards for achievement and little direct involvement by parents in their children's schoolwork, and a belief that all children, except those who may be seriously disabled, should be able to learn and do well in elementary school.

In 1998, in another study of differences in mathematical achievement among students in China, Taiwan, Japan, and the United States, Lee demonstrated that achievement differences are closely related to cultural expectations about teaching and learning. For example, in general, Asians believe and expect that everyone is capable of learning, whereas Americans believe that individual differences are inherent and, as a result, that people are expected to behave and perform differently. Lee concluded that it is this American philosophy that prevents the development of common standards and makes it difficult to institute new teaching strategies. Lee suggested that this view may seriously impair efforts to improve teaching and learning among students and teachers in the American culture for quite some time into the future. In fact, evidence from several sources shows little improvement in American mathematical performance over the past decade or so, with proficiency being even lower than it was during the 1970s.

In 2005, Gardiner and Kosmitzki reported trends similar to those just cited. In 2001, in an earlier summary of the themes and results of cross-cultural studies of mathematical achievement, Gardiner pointed first to the dominant role in this process played by mothers and second to cross-cultural evidence strongly suggesting that informal learning styles, such as those found in many Asian cultures that focusing on building interest, are more effective in teaching children a variety of skills, including math.

#### **4. CULTURAL EXPLANATIONS FOR CROSS-ETHNIC DIFFERENCES IN ACHIEVEMENT IN THE UNITED STATES**

The relationship between educational achievement and culture is also shown in research in the United States. As

indicated earlier, cross-ethnic comparisons of educational achievement have shown that Asian American students outperform European American students, who in turn outperform African American and Hispanic American students. A significant portion of these differences may be attributed to cultural factors in values, beliefs, expectations, cognitive styles, and the like. These factors directly or indirectly influence students' motivation for learning and their approaches to school achievement. Over the years, several theories have been advanced to explain ethnic differences in educational achievement.

An early explanation was the so-called deficit theory. According to this theory, students who do poorly in school are considered to be deficient in some way compared with those who do well in school. The deficit may take the forms of inferior ability, culture deficiencies, and/or language problems. Proponents use these deficits to explain why some minority group students underachieve in education. These students are viewed as incapable or inferior because they exhibit cultural backgrounds, thinking, learning styles, and communication patterns that are different from those of the dominant culture. The deficit explanation stems from ethnocentrism, that is, the belief that one's own cultural ways are superior to those of other people. It is also known as the "blaming the victims" model. When children are not doing well in school, it is assumed that something is wrong with the students or with their particular culture. Most social scientists now consider the deficit explanation to be out of date.

A second explanation comes from the expectation theory, which was first made popular by Rosenthal and Jacobson in 1968. This theory posits that some children do poorly in school because their teachers have low expectations for their success. Subsequently, research has found that racial and ethnic backgrounds are factors that affect teachers' expectations of students. Teachers often have low estimations of the academic potential of students from certain racial and ethnic minority groups and do not expect much of these students. Because of the low expectations, teachers may unknowingly teach and interact with minority students in ways that are different from the ways in which they teach and interact with other students. This may, in turn, affect the students' own expectations and learning and result in low academic performance. The entire cycle becomes a self-fulfilling prophecy. The expectation theory holds that academic performance of certain minority group students can be improved if teachers can reduce their subjectivity and modify their behavior in working with these students.

The third explanation is the cultural difference theory. This theory maintains that schools in the United States are based on the majority (European American) culture, which differs in many ways from the minority cultures. Therefore, there is a mismatch or incompatibility between the school culture and the home cultures of minority students. The values, languages, cognitive styles, interaction patterns, and behaviors required in the school are often not the same as those that minority students have learned at home. These cultural differences interfere with the learning of these students and affect their motivation to do well. As a result, schools generally fail to tap the competence of minority students. Proponents of the cultural difference theory contend that it is important for teachers to develop better cross-cultural understanding and to adopt instructional practices that are more responsive to minority cultures. A learning environment that is more congruent with the minority cultures is necessary to foster effective learning of minority students and to enhance their academic achievement.

However, not all cultural differences lead to conflicts and underachievement among minority students. For example, Asian American students have cultural and linguistic backgrounds that are different from those of the majority culture in significant ways, yet these students consistently do well in school. In 1991, Ogbu proposed an explanation to account for this phenomenon. He differentiated between minority groups as involuntary immigrants and voluntary immigrants. Involuntary minority groups are those whose members are incorporated into the U.S. society against their will (e.g., African Americans), whereas voluntary minority groups are those who came to the United States by their own choice (e.g., Asian Americans). Involuntary minority students view maintaining their own cultural identities as very important and are not inclined to become enculturated to the majority culture. In addition, sometimes there is also social pressure from parents and peers discouraging these students from conforming to the school culture. For these reasons, it is not uncommon for involuntary minority students to develop a resistance to schooling and to reject behaviors that would make them successful in school. Consequently, their academic performance suffers. On the other hand, voluntary minority students have less difficulty in crossing cultural boundaries and are less affected by identity conflict. They are able to adjust to the school culture and do well in school.

Finally, a more recent explanation was proposed by Okagaki in 2001. Okagaki developed a triarchic model to explain ethnic differences in achievement. Unlike

previous explanations that focus on a single perspective, this triarchic model takes a holistic approach and looks at multiple factors that contribute to ethnic differences in educational achievement. The three major factors identified in the model are characteristics of (a) the school (e.g., the structure of school and perceived function of education by various groups within a society), (b) the family (e.g., cultural beliefs and practices of parents and families), and (c) the student (e.g., ethnic and academic identities). Okagaki maintained that all three factors shape the ways in which students approach and perform in school. Therefore, all three factors need to be examined for a thorough understanding of cultural influences on minority students' achievement. Because of its comprehensive nature, the triarchic theory is able to explain some of the variations in minority students' achievement that cannot be explained satisfactorily by the earlier theories.

## **5. ETHNIC DIFFERENCES IN CULTURAL FACTORS AFFECTING ACHIEVEMENT**

Research on ethnic differences in educational achievement over the past 20 years or so has identified a number of cultural factors that contribute to school success. Five major factors are considered in this section: educational effort, cultural values and beliefs, parental educational expectations, parental involvement, and peer influence. These factors are important because they can enhance the educational achievement of all students regardless of race or ethnicity.

### **5.1. Educational Effort**

Educational effort has a direct impact on school performance. Students who spend more time learning and doing homework perform better than do students who spend less time studying. Educational effort has also been found to be associated with culture and ethnicity. For example, a number of studies have found that Asian American students, and to some extent European American students, are more task oriented, pay more attention in class, exert more effort, spend more time doing homework, and participate more in academic-related activities compared with other ethnic minority students. Asian Americans are also more likely to believe that high achievement is the result of motivation and hard work. On the other hand, students from other ethnic

groups are more likely to attribute academic success to ability or talent—factors that are outside of their control.

### **5.2. Cultural Values and Beliefs**

Another factor that influences educational achievement is cultural values and beliefs related to education. All ethnic groups value education. However, in 1990, Mickelson indicated that it is not the abstract belief that education is important; rather, it is the more concrete pragmatic belief about the benefit of education that separates high-achieving students from low-achieving students. Specifically, students who do well in school believe that education serves an important function for them (e.g., getting better paying jobs, improving family financial or social conditions, overcoming occupational discrimination). Conversely, students who do not see education as critical to serving a particular pragmatic function show less motivation to do well in school. In a study conducted by Steinberg and colleagues in 1992, it was found that African American and Hispanic American students did not believe that doing poorly in school would hurt their chances for future success. Asian American and European American students, on the other hand, were more likely to believe that academic success would have a significant payoff. These different beliefs among various ethnic groups affect their behaviors related to education.

### **5.3. Parental Educational Expectations**

Research has shown that parental educational expectations is another contributing factor to success in school and is closely associated with students' academic performance. Differences in parental expectations exist among ethnic groups. For example, compared with parents in other ethnic groups, Asian American parents place greater emphasis on educational accomplishments and have higher expectations for achievement for their children. They expect their children to receive better grades in school and to earn higher educational degrees than those of their counterparts in other ethnic groups.

### **5.4. Parental Involvement**

Parental involvement is empirically linked to academic performance. Research has shown that parental involvement can improve children's school performance. In general, parental involvement may take place in two

broad forms: direct participation and indirect support. In direct participation, parents are actively involved in helping their children with their schoolwork, maintaining good communications with the school, participating in school activities, and the like. On the other hand, indirect support does not involve direct participation of the parents in their children's education. Rather, parents work to maintain a home environment that is highly supportive of their children's learning (e.g., arranging for ample time to study, paying for private lessons). In terms of direct participation, research has shown that European American parents show the highest level of involvement in their children's education. In comparison, Asian American parents show the least direct participation in their children's education. They seem to perceive direct involvement with the school as less important due to cultural beliefs and/or language barriers. Instead, Asian American parents provide more indirect support to their children's education than do parents in other ethnic groups. Asian American parents are willing to make sacrifices for their children's educational pursuits to make sure that their children have sufficient time to do schoolwork and to create a climate in which their children's job is to study and do well in school.

### **5.5. Peer Influence**

In addition to parental influences, studies have shown that students tend to do better with peer support for academic achievement. In some cases, peer influence is even greater than parent influence. In a 1992 study with adolescents, Steinberg and colleagues found that European American and Asian American adolescents were more likely to have friends who place a great deal of emphasis on academic achievement and, as a result, worked hard to keep up with their friends. In comparison, Hispanic and African American adolescents were less likely to receive peer support for academic achievement.

## **6. DIRECTIONS FOR FUTURE RESEARCH ON CROSS-CULTURAL ACHIEVEMENT**

Making reliable and valid comparisons across cultures is often a bewildering and difficult task if not an impossible one. Conducting cross-cultural and cross-ethnic studies in educational achievement is fraught with a variety of methodological difficulties. Therefore, one

must exercise caution in drawing definitive conclusions from such studies. Also, as mentioned earlier, vast variation exists within any cultural or ethnic group. It is critical that research findings on achievement differences among cultural and/or ethnic groups be understood in the context of differences within groups.

Several researchers have recommended possible solutions that are worthy of further consideration. For example, in 1998, Stigler outlined the benefits and problems associated with his large-scale video survey approach to teaching and learning used in the Third International Mathematics and Science Study (TIMSS) Videotape Classroom Study. Among the advantages is the fact that video data are concrete, are grounded in practice, allow for merging of quantitative and qualitative analyses, and enable study of complex processes. However, if these data are to be used successfully, a number of disadvantages must be addressed, including sampling of only a few instructional lessons, standardization of camera procedures, problems of observer effects, sampling and validity, and confidentiality.

In another approach, in 1998, Paris and van Kraayenoord demonstrated the usefulness of specially designed tasks for assessing young children's literacy and development (e.g., asking students to examine and select books they find personally interesting and to construct meaning and stories from pictures or wordless picture books). They suggested that this approach can provide a blueprint that could lead to consistent and sensible instruction and assessment practices that are more easily understood by children, parents, and teachers.

The authors of the current article recommend that researchers seriously consider the development of a variety of new approaches, including large-scale, interdisciplinary, multimethod, longitudinal studies that focus on important aspects of educational achievement such as cooperative learning environments, multicultural literacy, and pedagogical changes in teaching and learning.

### See Also the Following Articles

Academic Failure, Prevention of ■ Academic Interventions ■ Achievement Motivation in Academics ■ Educational and Child Assessment ■ Gender and Education ■ Home-School Collaboration ■ Learning ■ School-Community Partnerships

### Further Reading

- Alexander, R. J. (2001). *Culture and pedagogy: International comparisons in primary education*. Malden, MA: Blackwell.
- Beauchamp, E. R. (Ed.). (2003). *Comparative education reader*. New York: Routledge Falmer.
- Mazurek, K., Winzer, M. A., & Majorek, C. (2000). *Education in a global society: A comparative perspective*. Boston: Allyn & Bacon.
- Okagaki, L. (2001). Triarchic model of minority children's school achievement. *Educational Psychologist*, 36, 9-20.
- Steinberg, L. (1996). Ethnicity and adolescent achievement. *American Educator*, 2, 2-48.
- Tsuneyoshi, R. K. (2001). *The Japanese model of schooling: Comparisons with the United States*. New York: Routledge Falmer.



# Educational and Child Assessment

Jan ter Laak and Martijn de Goede

Utrecht University, Utrecht, The Netherlands

1. Introduction
  2. Four Goals of Assessment
  3. A Comprehensive Definition of Assessment
  4. Possibilities and Limitations in Assessing Children
  5. Conclusion
  6. Case Study: an Example of Educational and Child Assessment
- Further Reading

description, diagnosis, prediction, or explanation of the child's behaviors containing three components: implicit and explicit psychological theories and constructs, test theory representing the constructs in a quantitative model, and instruments, tests, or procedures constructed in accordance with theories and their representation in the model.

## GLOSSARY

**assessment** A procedure whereby the assessor's reformulation of the client's problem or question is answered (or solved) using scientifically sound methods with relevant and reliable information about the child's behaviors and environment.

**educational and child assessment** An assessment that is focused on "maximum and typical performance," that is, on the child's intelligence, aptitude, school achievements, and typical behaviors.

**individual differences approach to achievement and typical performance** An approach that determines specific content domains of achievement (e.g., reading, arithmetic, biology, history, science) and of typical behaviors (e.g., socioemotional behaviors, personality); experts decide which content must be acquired as a result of a curriculum, and other experts design instruments that measure the level of acquisition of this knowledge and children's skills.

**psychological assessment** The professional process of investigating and judging a client's behaviors, resulting in a

The assessment process is the professional procedure whereby a client's question (or problem) is answered (or solved) using psychological theories and constructs that are modeled by test theoretical and statistical techniques and that result in reliable and valid instruments to measure the child's achievement and typical behaviors. The goal is a valid description, categorization, prediction, and/or explanation of the child's behaviors (e.g., cognitions, emotions, observable behaviors). These are the points of departure for answering the client's question or deciding on the treatment for his or her behavioral problems.

## 1. INTRODUCTION

What is educational and child assessment? How do we assess the educational level of the child and his or her social behaviors and personality? In this article, first the child as the subject of an assessment is described. Second, some specific issues in assessing children are

reported, four goals of assessment are distinguished, and assessment is provisionally defined. Third, assessment is comprehensively defined with the following three components:

- Implicit and explicit behavioral theories and constructs
- Test theory and statistical models to represent these theories and constructs
- Instruments for measuring these constructs

Fourth, the assessment process, which consists of compiling these three components and contains several steps from the client's question or problem to the answer or solution, is described. This process includes the integration of diagnostic information and is subject to judgmental errors. Fifth, the possibilities and limitations of assessing achievement in children are described. Finally, the case of "Bryan" is reported using this article's definition of assessment.

Assessment in children is considered if cognitive, emotional, or behavioral problems are experienced by parents, teachers, the social environment, the institution, and/or the child himself or herself. The help of a professional assessor is sought when the implicit assessment of the parents, teachers, or child does not result in a satisfying solution. Assessment is the procedure whereby the assessor reformulates the client's question (or problem) in psychological terms and answers the question (or solves the problem) using scientifically sound methods and information about the child and his or her environment.

In assessing children, one is confronted by special issues. First, parents, teachers, and/or the school or institution—but rarely the child—ask for an answer or for help. They may or may not agree about the presence and nature of the problem. For example, Lewis reported in 1999 that the manifestation of child depression is not shared by children, parents, and teachers. Second, given that children develop rapidly and are "moving targets," how do we know whether the current behaviors are advanced or behind or whether they are normal or abnormal? An accurate picture of behavioral development is needed to make that determination. Third, assessing a child is a social process. The child must consider the assessment as a task and not as a game. Fourth, in (young) children, the nature and source of diagnostic information deviate from procedures used with adults, for example, because questionnaires cannot be used with children. Their behaviors must be observed, and parents, teachers, and peers must be interviewed.

## 2. FOUR GOALS OF ASSESSMENT

Although assessment is not a psychological subdiscipline such as educational and developmental psychology, it depends largely on psychology's knowledge base and methodology. Assessment refers to gathering and integrating relevant diagnostic information about the child's behaviors so as to help him or her. In their definition of developmental assessment in 1999, Johnson and Sheeber stressed the child's levels of cognitive, sensorimotor, language, and socioemotional behaviors in presenting the weak and strong sides of the child's functioning. So, assessment contains deciding, problem solving, information gathering and integrating, helping, constructing a whole picture of the child's behaviors, and using teachers, parents, peers, and the child as sources of information.

Assessment has four practical and scientific goals. The first goal is a valid description of the child's cognitions, emotions, and behaviors (hereafter referred to as "behaviors") and of his or her social environment. The practitioner needs valid descriptions that enable him or her compare children, to describe the "natural" course of development of behaviors, and to establish the effects of social contexts on behaviors. This description is also needed to estimate the relationships between behaviors, to determine context conditions that influence the behaviors, and to describe and explain their course in time. The second goal is diagnosis, which refers to how experts assign the child to a category (e.g., the attention deficit/hyperactivity disorder category). This is relevant for the practitioner and the scientist. The quality of these assignments depends on the probability of false positives and false negatives and on differential diagnoses and comorbidity of two or more behavioral syndromes. The third goal is prediction, which allows the scientist to test a nomological network and enables the practitioner to predict the child's future school achievement and social adaptation. The fourth goal is explanation, which is expressed in the hypotheses testing model of assessment. Explanation fits the dominating empirical-analytical approach in psychology.

Educational assessment is focused primarily on maximum performance. "Child assessment" in the title of this article is viewed as measuring the child's the socioemotional behavior and personality (or "typical performance"). Intelligence, aptitude, and school achievements are called "achievements."

Intelligence refers to a general ability that includes reasoning, planning, problem solving, abstract thinking,

understanding complex ideas, and learning from experience. It is not rote learning or school achievement, and it differs from creativity, character, and personality. Closely related to intelligence is aptitude, which refers to the ability to achieve in a professional field after specific training.

School achievement is related to both of these concepts but refers to what a child has achieved in a specific domain as a result of specific training or of attending a curriculum (e.g., reading or arithmetic at the primary school level, achievements at the high school or university level).

Socioemotional adaptation and personality refer to the maturity level, adaptation, lack of pathology, and individual differences in temperament and personality.

### 3. A COMPREHENSIVE DEFINITION OF ASSESSMENT

Psychological assessment is a process of investigation and judging that results in a description, diagnosis, prediction, or explanation of the child's (problem) "behaviors." This process involves the following three components:

- Implicit and explicit psychological theories and constructs
- Test theory representing the constructs in a quantitative model
- Test construction using theories and their representation in the model

The three components are combined in the assessment process, with the assessor acting as the "composer." The components and the process constitute the comprehensive definition of assessment used in this article and are illustrated in the following subsections. The process consists of the composition of the three components. In this process, the assessor must integrate the diagnostic information and formulate advice. Information integration is known to be subject to errors.

#### 3.1. Psychological Theories and Constructs in Assessing Children's Achievements and Social Adaptation and Personality

Assessment involves content (i.e., behaviors) that is inferred from observations and that fits in a theory or network of constructs. The many theories and constructs

for children's achievements and personality are subsumed under the "three central orientations" or approaches used in psychology:

- The individual differences orientation
- The developmental orientation
- The (social) context orientation

The prototype of the individual differences orientation is Spearman's approach to intelligence. Spearman insisted on the study of individual differences as being the main way in which to understand intelligence. He discovered or invented the "G" (general) factor because he assumed that one general dominating ability was present in all of the activities requiring mental effort. In 1998, Jensen characterized the G factor as a discovery that can be compared to Skinner's laws of reinforcement and Thorndike's law of effect.

Thurstone was the father of the multiple intelligence theories. He accepted the G factor but attached little significance to it. He viewed it as an abstract, "second-order factor" that "floated above" more interesting specific factors. Thurstone discovered or invented the seven "primary mental abilities": verbal comprehension, word fluency, number, space, associative memory, perceptual speed, and reasoning. Later, Guilford distinguished 120 factors, and Vernon created order among the many factors by formulating a hierarchical model with G on top and two broad groups of factors below: verbal-educational (e.g., English, history) and spatial-mechanical (e.g., automotive repair). Below these, he defined all kinds of specific factors.

The orientation of Spearman and the other authors is known as the psychometric approach to intelligence. This approach combines intelligence and aptitude because Thurstone's multiple intelligence resulted in measuring aptitudes in the same way as the primary mental abilities.

The individual differences approach to school achievement does not aim to discover factors in school achievement; rather, it aims to determine specific content domains such as reading, arithmetic, biology, history, and science. Experts decide what knowledge must be acquired as a result of a curriculum, and other experts (e.g., from the Educational Testing Service [ETS]) design instruments that measure the level of acquisition of this knowledge and the children's skills. The results can be used to describe individual differences in children's achievements, the level reached by a class or school, and/or the degree of success of the curriculum and teachers. In addition, the individual differences approach has had a great impact on child

assessment in measuring dimensions of personality, temperament, and social adaptation.

This approach served the goal of providing a reliable and valid description of the ranking order of a child's achievement and personality. It helped to categorize children, for example, distinguishing cognitively retarded children from nonretarded children and distinguishing well-adapted children from not well-adapted children. The individual differences were used for predicting the future levels of achievements, school and vocational success, and levels of adaptation.

The prototype of the developmental orientation is Piaget's cognitive developmental theory. Piaget was convinced that to understand human cognition, one must study its historical and ontogenetic development (i.e., within the life course). His original question was an epistemological one: How could humankind change from biological beings to logical scientific ones? Piaget was not interested in individual differences or in fostering cognitive development by curricula or training. His subject was not a specific pupil but rather the epistemic subject, and the environmental stimuli were only an "aliment" (nourishment) to existing cognitive structures. Piaget's theory was also applied to social adaptation (e.g., the development of levels of perspective taking). This developmental orientation implied that children's stages of cognitive and social development had to be assessed. The stage concept hypothesized the existence of qualitatively different stages, ordered in time, that resulted in an inevitable and equilibrated final stage and that were characterized by sudden transitions. Empirical support for these claims was scarce and circumstantial.

This developmental orientation could be readily applied because many achievements at school (e.g., understanding simple scientific laws, understanding the concept of living beings, understanding economic laws) can be ordered along the Piagetian developmental dimension from prelogical thought to formal operational thought.

The impact of the developmental orientation on educational and child assessment is limited. Instruments for measuring stages were scarce, and measurements were more difficult to conduct and interpret than were the traitlike individual differences measurements.

The prototype of the contextual orientation to educational and child achievements is the attempt to influence achievements and socioemotional adaptation by curricula and programs. The latter are the independent variables that should change the dependent variables of intelligence, aptitude, and school achievements and

socioemotional adaptation. This orientation is also considered to be the only road to describing and explaining behaviors by efficient causes, and it fits well the dominating empirical-analytical approach in psychology. The methodology was important, because the effect of independent variables could be revealed. It served the goal of determining the effect of training, intervention, and situational change (i.e., the context) on achievements and on social adaptation and personality. These independent variables of training and the like could vary from small manipulations in content and form of curricula to big programs such as Head Start. The dependent variables were either borrowed from the individual differences orientation (e.g., intelligence, aptitude, and school achievements; level of socioemotional adaptation) or constructed for a specific question (e.g., change in a memory retrieval strategy).

The context orientation was not aimed at constructing instruments for measuring children's achievements and personality. This is possibly one reason why assessors underestimate the role of the context in understanding and explaining achievements and typical performance. However, the effects of schooling on intelligence and the effects of the cultural influence on generational growth of intelligence are present (e.g., in the Flynn effect).

Educational and child assessment is dominated by the individual differences approach, partly because this approach contained the simplest road to developing instruments for measuring maximum and typical performance in children.

### **3.2. Modeling Children's Achievements and Socioemotional Adaptation**

#### **3.2.1. Modeling Individual Differences in Intelligence and Aptitudes**

Spearman combined conceptual analysis and test-theoretical and statistical modeling. His two-factor theory implied that individual differences in true scores could be attributed to two factors: general (present in all tasks) and specific (present in tasks of the same type, e.g., verbal, spatial). His type of factor analysis was suited to discover a first strong factor (e.g., with the now obsolete centroid method).

Thurstone's multiple intelligence construct required a search for a simple structure, that is, where items loaded high on one intended factor and low on all other factors. Orthogonal rotation and a specified number of factors

were appropriate. This simple structure concept is usually applied to the measuring of aptitudes (e.g., in the differential aptitude test). Moreover, simple structure analysis is often applied in an exploratory way to reduce many items to a few dimensions.

Guilford needed a factor analysis that discovered many factors. He “forced” the items in these factors by using the Procrustes technique. The hierarchical model was developed to rank the many factors.

The analysis of dimensionality and of the quality of individual items is now supported by the many possibilities of modern test theory or item response theory (IRT). These theories contain models to test the unidimensionality of a scale and to weight the individual items according to their contributions to the total test score.

The analysis of school achievement was a matter of experts deciding what to measure and how to measure it. There was no effort to find any dimensionality for school achievement. Modern test theory (IRT) is however, used more and more to construct one-dimensional achievement tests.

### **3.2.2. Modeling Development of Achievements**

Developmental psychology was characterized by Cronbach as an individual differences discipline, whereas Wohlwill tried to define a real developmental orientation. By considering the developmental orientation as an individual differences orientation, developmental research was biased toward investigating stability with the help of tests, that is, rank order and normative stability. Because researchers were looking for stability, they found it. However, there was never complete stability, even after correction for measurement error. The dominating message from this research is that achievements are relatively stable, predictable, and foreseeable. Moreover, the message is that achievements increase over time in the individual and sample because children grow older and because of the Flynn effect. Rarely is a developmental construct used to model the development of achievements in the course of time.

### **3.2.3. Modeling the Context of Achievements**

The effects of manipulations of independent variables, interventions, programs, events, stimuli, and situations on achievements can readily be modeled with (multiple) analysis of variance (ANOVA). True and quasi-experimental designs have been well elaborated. This technique is appropriate for determining the existence of significant differences at the group level, and

computing effect sizes, that give a fair estimate of gain. However, these differences do not apply to every element of the sample, and the differences do not necessarily last forever. Both are partly reasons for the disappointing results seen in programs to improve children’s achievements. In addition, behaviorists’ optimism for the extent to which social change can be effected by educational programs has proven to be misguided.

The modeling of social adaptation and personality does not deviate from this pattern in modeling intelligence, aptitude, and school achievements. The individual differences approach and factor analysis also rule these domains, and IRT is rarely applied in measuring personality and socioemotional adaptation.

## **3.3. Instruments for Assessing Children’s Achievements and Socioemotional Adaptation**

Theories and constructs, on the one hand, and test-theoretical and statistical modeling, on the other, can be balanced and produce theory-driven instruments for educational and child assessment. However, practical questions have dominated test construction, resulting in the gathering of items that could predict practical criteria such as school and vocational success and adaptation and pathology. Examples of instruments measuring the following eight domains are presented next:

- Screening of developmental backwardness
- Cognitive and sensorimotor development
- School achievement
- Socioemotional adaptation
- Childhood pathology
- Personality and temperament
- Social status in a peer group
- Children’s social environments

### **3.3.1. Screening of Developmental Backwardness**

Screening instruments attempt to measure developmental backwardness or disturbance. An example is the Denver Developmental Scale. Children from 6 days to 6 years 6 months of age are investigated with tasks, and the parents are interviewed. Despite its name, this instrument is an individual differences instrument and considers backwardness to be a one-dimensional scale. The McCarthy Developmental Scales presuppose

orderly development in several domains. However, this presupposed “natural” development has not been tested empirically. These instruments contain different scales for cognitive and sensorimotor behaviors, and the children (2–9 years of age) are investigated using tasks and games to elicit problem-solving behaviors. Some items must be answered by the parents. The Gesell Motor Scales measure motor, speech, and social behaviors in children from 4 weeks to 6 years of age. Gesell presupposed a developmental scale in these behaviors.

### **3.3.2. Cognitive and Sensorimotor Motor Development**

Cognitive and motor development can be investigated in young children using the Bayley Developmental Scales. Tasks are given to babies and toddlers, and their parents are interviewed. Two scores result: one for mental development and one for motor development. This is an individual differences test, and the two factors are rationally defined. The Wechsler scales for preschool children (Wechsler Preschool and Primary Scales of Intelligence [WPPSI]) and for school children (Wechsler Intelligence Scale for Children [WISC]) are the most widely used intelligence tests. Wechsler defined intelligence as the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his or her environment. Intelligence was considered to be an individual differences phenomenon. The performance and verbal intelligence factors were rationally distinguished and, in fact, usually correlate at approximately .70. Many factor analyses were conducted. The “three Kaufman factors” (verbal comprehension, perceptual organization, and freedom from distractibility) interpretation of the WISC is popular. The existence of these factors is supported by factor-analytical results. The Stanford–Binet scale is the oldest intelligence test. The intelligence quotient (IQ) is defined as a comparison between what is normal for a certain age and what is observed in a specific child.

### **3.3.3. School Achievement**

School achievement tests (SATs) are found in every Western country and many other countries. These were developed to cover a content domain and are treated as individual differences variables. It is not the primary goal of these measures to distinguish underlying factors that generate or cause these individual differences. SATs and grade point averages are useful in selecting and placing children in various

educational levels, and they do predict future school achievements fairly well. The Differential Aptitude Test contains nine subtests (e.g., sentences, analogies, technical ability, arithmetic skill, words) that resemble Thurstone’s multiple intelligence factors.

### **3.3.4. Socioemotional Adaptation**

Social emotional adaptation can be measured using many instruments, but all are based on the individual differences orientation, for example, Harter’s scales for measuring the aspects of self (e.g., academic, social, and motor skills; being socially accepted).

### **3.3.5. Childhood Pathology**

Childhood pathology can be measured by Achenbach’s Child Behavior Check List (CBCL). It is based on the individual differences approach, and its factors are partly rational and empirically defined (e.g., the distinction between internalizing and externalizing behaviors). In 2000, Wenar and Kerig described several pathological syndromes in children using the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Although this system was originally developed for persons 18 years of age or over, it can be used for childhood pathology as well. In addition, many lists are available for teachers to judge the social and task behaviors of pupils. Kovacs’s questionnaire to measure depression is a downward extension of the Beck Depression Scales and allows one to define clinical cases of depression in children. The Attention Deficit/Hyperactivity Disorder (ADHD) is a short questionnaire to measure impulsivity, lack of attention, and the level of undirected activity in children 4 to 18 years of age. The Bender consists of prescribed figures that children must draw and helps to screen neurologically based deficits in perception and expression of simple and complex “Gestalts.”

### **3.3.6. Personality and Temperament**

Personality characteristics are considered to be individual differences variables. In addition to the “Big Five,” there are the “Small Five” for children measuring extraversion, agreeableness, conscientiousness, emotional stability and openness. The California Q sort of Block’s Berkeley group distinguished three types of children: ego resilient, overcontrolled, and undercontrolled children 2½ to 18 years of age. These can easily be connected to the five personality factors.

Temperament is often considered to be a precursor to the personality traits. Several distinctions can be made: easy, slow to warm up, and difficult children.

### 3.3.7. Social Status in a Peer Group

Social status is an important variable because children have peers who influence their behaviors. In this procedure, children nominate peers with whom they like to play or work. This results in three types of children: popular, unpopular, and neglected children.

### 3.3.8. Children's Social Environments

Family and school relationship instruments measure individual differences in families and schools. Moos's social climate scales usually consist of approximately 100 items and result in three to nine dimensions underlying these items. The three most general dimensions are relationship (e.g., cohesion, conflict), personal growth (e.g., autonomy, task oriented), and system continuity/change (e.g., control, order, leadership). There are several questionnaires to measure the parental styles: authoritative, authoritarian, or laissez-faire. The school environment is also measured, that is, how this environment fosters achievement motivation, well-being, and self-confidence in pupils. The mental measurement yearbooks offer results from research on all of these instruments. The *Standards for Educational and Psychological Tests*, published in 1999, contains guidelines to judge the theoretical and psychometric qualities of these tests and questionnaires. Recent meta-analyses are a good source for judging the construct and criterion validity of many instruments.

In conclusion, all of the preceding instruments are based on the individual differences orientation. There are only a few instruments that use a developmental construct (e.g., Loevinger's sentence completion test to measure the level of ego development, Keressies and colleagues' scales for sensorimotor development). However, these latter instruments are not frequently used by assessors. In the contextual orientation, the dependent variables usually are the well-known individual differences variables.

## 3.4. The Assessment Process in Children

The assessment process starts with the client's question (or problem) and ends with the advice (or solution). It

is modeled according to the procedure of empirically testing hypotheses in a true experiment and contains the following:

- The problem/question as formulated by the client
- The assessor's reformulation of the problem in psychological constructs
- The assessor's formulation of several hypotheses containing different causes of the (problem) behaviors
- The choice of instruments to "test" the hypotheses
- Integration of all the information from dossiers, anamnesis, test results, and interviews with parents, teachers, and institutions into a coherent picture
- The professional advice or solution, including recommendations for a specific intervention, training, therapy, or program

Assessment differs from testing in that the latter is more oriented to measuring variables, whereas assessment is more oriented to answering a client's question or solving his or her problem. Testing is limited to the use of standardized instruments, whereas in assessment the judgmental processes and information integration play a significant role. Testing requires skills to conduct and score tests, whereas the assessor must also be familiar with the three components involved in assessing children. Assessment involves choosing, conducting, and interpreting tests and their results, the analysis of each specific case, and the weighing and integration of information. Finally, the assessor should be able and willing to account for all of the steps taken in this process.

Traditionally, pitfalls threaten the assessor. This issue concerns the integration of all the information from the various sources and is known as the clinical prediction versus statistical prediction controversy. In 1999, Garb reviewed all of these pitfalls and concluded that clinical prediction usually performs less well than statistical prediction. Pitfalls include the halo effect, fundamental attribution error, ignoring the base rate of phenomena, overestimating a salient detail and neglecting available statistical information, and giving information a wrong differential weighting.

It is recommended that the assessor use decision aids, such as the multi-attribute utility theory and the Bayesian rule, as much as possible to integrate information. A recent meta-analysis showed that clinical prediction is weaker than statistical prediction, but not to such a large extent as has been suggested. In addition, tests that traditionally demand much of the assessor's interpretive abilities, such as projective tests and interviews, are not as unreliable and poorly valid as is often

claimed. So, clinical prediction and controlled information gathering, using interviews and projective tests, and disciplined interpretation, using formulas whenever possible, can contribute to sound assessment of children's achievements and social adaptation.

#### 4. POSSIBILITIES AND LIMITATIONS IN ASSESSING CHILDREN

Assessment is performed using the composition of three components in the assessment process. This section addresses the possibilities and limitations. First, theories and constructs offer the possibility of defining the core elements of children's achievements and adaptation but are often not specific enough for testing in a model. Second, the individual differences approach dominates the measuring of achievements and social adaptation, but more developmental theorizing that allows instrument construction is necessary to gain a more complete picture of children's achievements and socioemotional adaptation in the course of time. Even more constructs may be necessary to model the diverse and real developmental pathways. Third, the quality of test construction can be fostered using the *Standards for Educational and Psychological Tests*, especially the standards for reliability and validity of tests. Fourth, achievement test construction was usually controlled by practical questions and contained much practical wisdom, but it can be enriched by clear theoretical considerations about the structure of achievements and adaptations. Fifth, the assessment process was structured after the empirical cycle for conducting scientific research, but this process can be supported by models and decision aids to minimize errors. Sixth, the assessor should be sensitive to the pitfalls in choosing, integrating, and interpreting the diagnostic information.

The assessment process requires disciplined assessors, whereas clinical assessors sometimes act freely and creatively to do justice to their "unique" clients.

Educational and child assessment poses specific problems.

#### 5. CONCLUSION

The assessor must be aware of the relevant three components and integrate these into his or her assessment

process. The first component is biased toward measuring individual differences on one or more dimensions, and this may detract attention away from real development in the child's achievements and typical behaviors and from the influence of the context of these achievements and behaviors (e.g., events, stimuli, programs).

The assessor can be biased. Using formulas to integrate diagnostic information and standardized procedures in tests that usually allow for subjective interpretations helps to reduce such biases.

The assessor must operate in the complex environment of the child, parents, teachers, the institution, and the wider society, on the one hand, and in the psychological discipline, on the other. The latter offers much for accomplishing assessment but is so broad that the assessor must specialize in-depth. Recently, many meta-analyses have appeared and offer an equilibrated summary of the qualities of instruments and interventions. An assessment is best conducted by a team representing the knowledge and skills of several disciplines, thereby preventing individual biases.

#### 6. CASE STUDY: AN EXAMPLE OF EDUCATIONAL AND CHILD ASSESSMENT

"Bryan" is a 10-year-old boy who is in the fifth grade. He has a younger sister who is in the fourth grade. Bryan's mother has no paid employment, and his father is a small businessman. The parents went to a school psychologist for advice because the teacher reported problems with Bryan's behavior in the classroom, and they wondered what was causing this. The teacher recommended that the parents seek professional advice for Bryan. The school psychologist listened to the parents and formulated their question as follows: "They want to know more about Bryan's lack of concentration and ask whether he himself experiences this lack of concentration. They would appreciate advice on how the teacher could help him." The psychologist reformulated the parents' question into the following diagnostic question: "How much are concentration and behavioral problems present, and how can these be explained?" This case was submitted to a professional institution, and it was decided to assess Bryan so as to give a professional and scientifically sound answer to the diagnostic question that

reflected the perceived problem of the parents, the teacher, and Bryan.

Bryan's problem was discussed with his parents and teacher. The parents had had no early problems with Bryan in starting to talk and walk. He cried a lot because he was eager to do more than he could. The preschool noted nothing special except that Bryan was inaccurate in fine motor tasks. He had to take the third grade of primary school twice because his reading and calculating were behind those of his classmates. Bryan could not concentrate well, and he quarreled a lot with the teacher. Some improvement was seen in the fourth and fifth grades, but concentration remained a problem and his writing was slovenly. In the eighth grade, Bryan said that he hated school and liked to stay at home. He often interrupted his tasks when sitting in front of the teacher, so he was placed at the rear of the class where he was quieter. The parents themselves said that Bryan needed order and structure, that he was willing to do his utmost, and that he was pleased with good results. He has broad interests. He has no steady friend, but this does not seem to bother him. He gets on very well with his father. Bryan is a scout, and the troop leaders consider him to be a "model child." The parents do not understand why he should need special help.

The teacher, on the other hand, called Bryan a chaotic boy who is not motivated to achieve. The teacher punished him, but this did not lead to any noticeable change in Bryan's behavior. His reading is average compared with his classmates, but it is actually 1 year behind the national average. His handwriting is bad, and he is forced to do many of his writing tasks twice. He makes too many errors in arithmetic because he is careless. He tries to finish his tasks as quickly as possible and then bothers other pupils. Sometimes he is very introverted. He is involved in conflicts in the schoolyard and when he works in a small group. He is not willing to admit that he makes mistakes. The teacher has suggested that Bryan be given professional help.

These complaints were clustered by the assessor into the following: lack of concentration, chaotic behavior, demanding attention, poor arithmetic, low reading level, bad reading and writing, low achievement motivation, conflicts with peers, and lack of social relationships at school. A first guess at the possible causes resulted in the following: inadequate visual organization, no planning, insufficient fine motor development, bad relationship with teacher, needs more structure, inefficient compulsion and punishment, low self-esteem, avoids contacts with peers at school.

Hypotheses as definite answers to the question were formulated, including criteria to test the answers:

1. Lack of perceptual organization, reduced concentration (WISC, observation, concentration test)
2. ADHD (DSM, 16-item ADHD questionnaire, observation)
3. Low self-esteem (observation, interview, two questionnaires)
4. Lack of achievement motivation (questionnaire)
5. Bad relationship with the teacher (interview, observation in the classroom)

The hypotheses were tested using a criterion set prior to the start of testing. The answers were as follows: 1, yes; 2, no; 3, no; 4, no; and 5, yes.

### **6.1. Summary of the Results for the Client (Bryan, Teacher, School, and Parents)**

Bryan has an average IQ, his perceptual organization is more than 1.5 *SD* below average, his attention and concentration are more than 1 *SD* below average, his relationship with the teacher is very tense, and Bryan avoids going to school. His self-esteem and achievement motivation are average.

Advice for treatment was concrete, and its implementation was feasible. It was considered that it would be better if Bryan could change class for the next year. The relationship with the teacher had to be improved. A daily planning of tasks was made and agreed on by both Bryan and the teacher, and the tasks were divided into small units. Instructions were short and clear, a time limit was given for each task, and he could ask questions at fixed times. The teacher was informed about the assessment results.

### **6.2. Three Months Later**

Bryan had started the new school year with renewed vigor. The relationship with his new teacher was normal. He still showed easily "off task" behavior, but this was simple to correct, and he ended his tasks within the time limits if supervision was available. His results were average, with only his progress in arithmetic still being slow; he had made progress compared with the past year. Bryan was more involved in peer activities and had a steady friend now. His parents said that Bryan was somewhat pleased about going to school.

### 6.3. The Psychological Investigation of Bryan and the Assessment Components

Concerning the instruments, all of the tests were constructed according to classic test theory, and results of IRT-based educational tests were used. The choice of these instruments was accounted for by the hypotheses; that is, definite testable answers to the questions were needed. The criteria for acceptance and rejection of the hypotheses were formulated in advance. There were several “rival” hypotheses to explain Bryan’s problem behaviors; two of these were not rejected, and three were rejected. Although the rival hypotheses were not really independent, the process did contain steps for trying to discover the “real cause” of the problem. There was no ordering in the hypotheses; that is, there was an educated guess as to the most or least successful. The cause of the problem was sought mainly “in the pupil,” although one hypothesis considered the relationship between Bryan and the teacher. There were more tests and questionnaires available to test the hypotheses. There was no argument for the specific choices (e.g., the quality according to *Standards for Educational and Psychological Tests*). Every instrument was about individual differences, and Bryan’s results are compared with those of a norm group or a criterion. Projective tests, often used in socioemotional problems, were not used.

There was no concept of development involved in the assessment; however, there was a follow-up interview with the parents and Bryan’s new teacher. There were nonspecific ideas about the effect of changing the classroom environment for Bryan. What was the exact condition changed (independent variable), and what was the precise behavior (dependent variable)? The process followed the prescriptions of the hypotheses testing model of assessment. Theory, as present in the instruments used, was scarce. Assessment was focused on answering concrete questions, but the theoretical base could have been stronger.

It is clear that this assessment protocol depended on the components of classic test theory and that IRT was used only because of the available educational tests. The individual difference orientation dominated, but implicitly a contextual orientation was present (changing the structure of the tasks with some effect on Bryan’s behaviors), whereas there was very little developmental orientation. Theoretical possibilities for explanation of the behaviors were not structured and possibly were insufficiently exploited. The hypotheses were not

ordered, and no real rival explanations of the problem behaviors were stated. The instruments were chosen because of the questions posed, and there was an attempt to use various methods to assess the same behaviors (e.g., tests, observation, interview). The assessment was discussed by a team, and the parents, the teacher, and Bryan himself were informed of the results.

The preceding analysis, using the comprehensive definition of assessment used in this article, helps to pinpoint the stronger and weaker elements in a concrete psychological investigation of a child. It is recommended that these components be used in conducting such investigations of clients.

### See Also the Following Articles

Assessment and Evaluation, Overview ■ Behavioral Assessment in Schools ■ Child Development and Culture ■ Educational Achievement and Culture ■ Intelligence Assessment ■ Intelligence in Humans ■ Neuropsychological Assessment in Schools ■ Personality Assessment ■ Psychological Assessment, Standards and Guidelines for ■ Psychophysiological Assessment

### Further Reading

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological tests* (7th ed.). Washington, DC: Author.
- Caspi, A., and Roberts, B. W. (1999). Personality continuity and change across the life course. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 300–327). New York: Guilford.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.
- Fernández-Ballesteros, R., De Bruyn, E., Godoy, A., Hornke, L., ter Laak, J., Vizcarro, C., Westhoff, K., Westmayer, H., and Zaccagnini, J. (2001). Guidelines for the assessment process. *European Journal of Psychological Assessment, 17*, 178–191.
- Garb, H. N. (1999). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.
- Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.
- Jensen, A. J. (1998). *The G factor: The science of mental ability*. London: Praeger.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M.

- (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128–165.
- Neisser, U. (Ed.). (1998). *The rising curve: Long term gains in IQ and related measures*. Washington, DC: American Psychological Association.
- Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (pp. 703–733). New York: John Wiley.
- Ter Laak, J., De Goede, M., & Brugman, G. (2001). Teacher judgments of pupils: Agreement and accuracy. *Social Behavior and Personality*, 29, 257–270.



# Effective Classroom Instruction

Margo A. Mastropieri and Thomas E. Scruggs

George Mason University, Fairfax, Virginia, USA

1. Introduction
2. Overview of Effective Classroom Instruction
3. Planning for Content Coverage
4. Using Effective Teaching Strategies
5. Promoting Effective Inclusive Instruction  
Further Reading

Effective classroom instruction refers to the application of the “teacher effectiveness” variables, that is, those variables that have been demonstrated to bear the strongest relation to student achievement. These variables include time on task, content coverage, pacing, scope and sequence, questioning, feedback, and praise. Systematic application of these elements has been demonstrated to increase academic achievement.

## GLOSSARY

- allocated time on task* The amount of time that teachers plan for instruction.
- alterable variables* The variables that teachers can actually alter (unlike age, gender, and personality) to improve student achievement.
- engaged time on task* The amount of time that students are actively engaged in instruction.
- guided practice* Student practice activities that are undertaken with direct teacher supervision.
- independent practice* Student practice activities that are undertaken independently.
- pace* The rate at which teachers proceed through instructional content; it also refers to the rate of presentation of information.
- teacher effectiveness* Teacher behaviors, identified through empirical research, that bear the strongest and most consistent relation with student achievement.
- time on task* The amount of time that students are working on instructional materials or attending actively to instructional presentations and discussions.

## 1. INTRODUCTION

Researchers over the years have been interested in what makes a “good” teacher. Early investigations focused on nonalterable behaviors such as personality, gender, intelligence, and even style of dress. Later, research focused on curriculum comparisons (e.g., phonics vs look–say methods of teaching reading) to determine effective teaching. Over the past several decades, researchers began to examine alterable variables, that is, those variables that teachers could alter in themselves to improve student achievement. This “teacher effectiveness” literature, although not without its critics, has been instrumental in identifying how teachers can best engage students in instruction to maximize their learning. Some important variables identified by teacher effectiveness research are provided in Table I.

Teacher effectiveness research has sometimes been criticized for its supposed emphasis on

**TABLE 1**  
Some Teacher Effectiveness Variables

<i>Variable</i>	<i>Description</i>
Time on task (engaged)	Direct student engagement with instructional presentations or materials
Planning for content coverage	Content presented to students during a given time period
Scope and sequence	Content to be covered and order of presentation
Objectives	Behavioral outcomes of instruction
Curriculum	Instructional materials and course of study
Pacing	Rate at which objectives are met
Teacher presentation	Teacher delivery of information, including elements such as structure, clarity, redundancy, and enthusiasm
Feedback and praise	Teacher acknowledgment and provision of information regarding performance and positive consequence of student responding
Guided and independent practice	Opportunities for student practice of acquired content or skills in supervised or unsupervised circumstances
Weekly and monthly review	Activities to promote long-term retention of academic content or skills

teacher-determined (rather than student-generated) curricula, transmission of knowledge by direct instruction (rather than construction of knowledge by student discovery and inquiry), and pointed emphasis on achievement test scores as outcomes (rather than measures of affect, classroom tests, performance or portfolio assessments, etc.). However, it has also been suggested that the “effective instruction” variables are important regardless of the type of teaching, learning, or assessment that is being undertaken. For example, even when curricula are student generated, learning is inquiry based, and outcomes are measured with classroom performance assessments, it is important that students be actively engaged in relevant activities and instructional materials. Likewise, regardless of content or approach to instruction, it is important that students receive relevant feedback and praise for their positive efforts.

## 2. OVERVIEW OF EFFECTIVE CLASSROOM INSTRUCTION

Effective teaching variables include those that involve planning for content coverage and that use effective teaching strategies. Planning for content coverage includes scope and sequence, objectives, curriculum, and pacing, whereas variables relevant to delivering instruction include maximization of academic engaged time, effective teacher presentations, appropriate use of practice activities, review, and formative evaluation. Making effective use of these variables can help to maximize the performance of general education students as well as that of students with disabilities or other special needs.

## 3. PLANNING FOR CONTENT COVERAGE

The importance of content coverage is obvious in that students almost certainly will not learn content or skills that have not been covered. Nevertheless, teachers vary greatly in planning for content coverage, and these differences are strongly related to student achievement. An important consideration in content coverage is that the amount and type of content covered must match the preskills and abilities of the students learning the content. Variables relevant to the general area of content coverage include objectives, scope and sequence, curriculum, and pacing.

Academic content is learned better if it is based on instructional objectives. Objectives state the desired outcomes of instruction so that the effectiveness of the instruction can be determined. Typically, an objective specifies (a) the content being addressed by the objective, (b) the conditions under which the student’s performance is to be assessed, and (c) the criteria established for acceptable performance. For example, consider the following objective: “The student will identify 5 homonyms from a list of 10 word pairs with 100% accuracy.” The content of the objective is homonyms, the conditions specify that students will be presented with a list of 10 word pairs, and the criterion for acceptable performance is identification (rather than production) with 100% accuracy. Another example of an objective is as follows: “The student will read four pages from the grade-level reading materials at a rate of 150 words per minute and will answer five comprehension questions with 100% accuracy.” This

objective also specifies content, conditions, and criteria to be achieved.

Scope and sequence refer to the breadth and depth of content to be presented and the order in which the content will be presented. Scope and sequence considerations allow for long-term planning of instruction and can provide implications for time allocations and the overall pace of instruction through the curriculum. Scope and sequence of instruction can help to prevent instructional redundancy and can help to ensure that students have mastered prerequisite content or skills before learning new content. Most states have developed standards for learning across the curriculum areas as well as grade levels that are used to define the scope and sequences selected for use at the school and classroom levels. These state standards present the basic learning that is expected at each grade level in every subject area. For example, the state of Virginia has standards for learning from kindergarten through 12th grade in mathematics, science, English, history, and social sciences, and these subjects are then assessed on statewide achievement tests. These standards can be accessed on each state's education department Web site.

The curriculum includes the instructional materials used for learning as well as the course of study and the scope and sequence within each grade level. The curriculum serves as an interface between the student and the learning objectives. Curriculum materials do not determine what instructional objectives will be; however, when they are used appropriately, they support and enhance the meeting of instructional objectives. For example, a teacher stating "Today we'll cover Chapter 12" is not specifying an instructional objective but merely stating which curriculum will be employed. However, curricula do carry implications for how instructional content will be presented. For example, a "hands-on" science curriculum that includes materials for student construction of electromagnets and telegraphs implies a model of instruction that is very different from that of a curriculum that includes a textbook with a chapter on electromagnetism.

Curricula are most effective for a variety of learning needs when they are directly relevant to the learning objectives, are carefully sequenced, provide sufficient redundancy of key concepts, and contain relevant activities for teacher evaluation of student progress toward meeting objectives. The most effective curriculum materials are meaningful and relevant for all students, including male and female students and students from a variety of cultural backgrounds, and should

include adaptations (or suggestions for adaptations) for students with disabilities or other special needs.

#### 4. USING EFFECTIVE TEACHING STRATEGIES

The most consistently identified teacher effectiveness variable is time on task. That is, the more time that students spend learning specific academic content, the better they will learn it. Distinction is made between allocated time on task (the time that teachers allocate to particular instructional activities) and engaged time on task (the time that students actually spend engaged in instruction). Although no consistent relation has been seen between allocated time on task and academic achievement, research has demonstrated that engaged time on task is positively related to academic achievement. Strategies for maximizing engaged time on task have been identified and include effectively managing transitions, avoiding digressions and other irrelevant verbalizations, and managing classroom behavior. Sometimes, teachers can improve student-engaged time on task by examining the amount of allocated time for class and monitoring their own teacher behaviors more closely. For example, teachers may find that they start classes 5 minutes late, end classes 5 minutes early, stop during the middle of class to search for relevant instructional materials, and stop to manage inappropriate classroom behaviors. Once such target areas are identified, teachers can design strategies to improve each area that will result in increasing the amount of student-engaged time on task during classes.

Effective teacher presentations have also been identified. Teachers maximize achievement when their presentations contain elements such as structure of lesson, clarity of teacher presentation, redundancy in emphasizing important concepts, and enthusiasm. Structure includes enlisting students' attention, providing a lesson overview that includes lesson objectives, providing outlines of the lesson and indicating when transitional points occur, and summarizing and reviewing key points as the lesson proceeds. It is important that students be made aware of the structure and objectives of the lesson so that they will know what is to be accomplished and how it will be accomplished. For example, teachers who begin with statements indicating the order of the learning activities (e.g., "First, we will . . .," "Second, we will be . . .," "Finally, we

will . . .”) provide the structure for the lesson for all students. Clarity includes presenting clearly and directly to the point of the lesson, avoiding vague or unfamiliar terminology, and providing concrete understandable examples. All teacher dialogue should be directly pertinent to the objective of the lesson. Redundancy does not refer to unnecessary repetition; rather, it refers to reemphasis and restating of the key elements of a lesson, particularly significant concepts and rules. Redundancy refers to the provision of multiple opportunities to practice learning newly presented content. Additional practice opportunities can vary in format and length given that the goal is to provide students with many chances to practice learning new information. Many students not only benefit from redundancy but also require redundancy prior to mastering new content. Teacher enthusiasm helps to maintain students’ attention and helps to model a positive attitude toward learning. Components of teacher enthusiasm include appropriate body movements and gestures, animated facial expressions, vocal inflections, positive acceptance of relevant student contributions, and a high overall display of positive energy.

Teachers also can increase achievement when they teach at an appropriate rate of presentation and when they maximize active student engagement with instruction or instructional materials. Selecting an appropriate rate of instruction not only is very important but also can be very challenging for teachers during the current era of statewide high-stakes testing. For example, if it is apparent that students have not learned a concept, it becomes necessary to back up and reteach that concept. This may require taking an additional class period or more depending on the students and the knowledge level required. The challenge for teachers is to ensure that all necessary content is covered before the end of the school year.

During teacher presentations, appropriate use of feedback and verbal praise is also associated with higher achievement. For example, praise that acknowledges student effort, specifies what the student did to merit praise, and demonstrates the association between effort and achievement is positively related to student achievement. Another important characteristic of praise is that it be genuine.

Following teacher presentation, guided and independent practice activities are associated with long-term learning and academic achievement. Guided practice activities are conducted under teacher supervision until it is certain that students are able to proceed independently. Students undertake independent

practice activities when the students have demonstrated acquisition of skills or concepts being taught and need independent practice to further reinforce learning. With both guided and independent practice, instructional tasks and materials must be directly relevant to instructional objectives.

Overall, a model of effective instruction for an individual lesson includes daily review, statement of objective, teacher presentation, guided practice, independent practice, and evaluation of learning objectives. Weekly and monthly review activities are also related to long-term learning and retention of important concepts.

## **5. PROMOTING EFFECTIVE INCLUSIVE INSTRUCTION**

Research has supported the use of teacher effectiveness variables with students with special learning needs as well as with normally achieving students. In inclusive classrooms (i.e., classrooms in which students with special needs are included with normally achieving students), use of teacher effectiveness variables has been seen to increase learning for all students. In inclusive classes, awareness of the range of student needs is even more important, and consideration of the effective classroom instruction variables can guide teachers’ instructional decision making.

Many of the teacher effectiveness variables have a direct impact on the characteristics of students with special learning needs. For example, because time is an important mediator of achievement, maximized time on task can directly address students who are having difficulty in assimilating academic content. Structure, clarity, and redundancy in teacher presentations can be directly beneficial to students who have difficulty in sustaining attention as well as to those who do not grasp new concepts quickly. Enthusiasm is helpful for students who are less motivated to succeed in school, and appropriate rate (or pace) of presentation of information and maximized engagement can be critical variables for those who must maximize efficiency to make adequate progress. Frequent review and evaluation can help to determine whether instruction was successful and when further intervention is needed.

Teacher effectiveness variables have also proven to be useful as a model for addressing specific learning problems. For example, if a student is having particular difficulty in mastering specific academic content,

teachers can consider whether certain variables can be adjusted. That is, teachers can increase the amount of engaged time on task, or they can adjust the rate of presentation or overall pace of instruction for meeting instructional objectives. Likewise, they can adjust their teacher presentations, the rate of presentation, or student engagement. Practice and review activities can also be adjusted. By attending to these alterable variables, student learning in inclusive settings can be addressed.

### See Also the Following Articles

Academic Failure, Prevention of ■ Mathematics, Teaching of ■ Reading, Teaching of ■ Teaching Effectiveness ■ Writing, Teaching of

### Further Reading

- Algozzine, B., Ysseldyke, J. E., & Elliott, I. E. (1997). *Strategies and tactics for effective instruction*. Longmont, CO: Sopris West.
- Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, 51, 5–32.
- Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook on research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.
- Carnine, D. (1976). Effects of two teacher presentation rates on off-task behavior, answering correctly, and participation. *Journal of Applied Behavior Analysis*, 9, 199–206.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Larrivee, B. (1985). *Effective teaching for successful mainstreaming*. New York: Longman.
- Mastropieri, M. A., & Scruggs, T. E. (1997). What's special about special education? A cautious view toward full inclusion. *Educational Forum*, 61, 206–211.
- Mastropieri, M. A., & Scruggs, T. E. (2002). *Effective instruction for special education* (3rd ed.). Austin, TX: ProEd.
- Mastropieri, M. A., & Scruggs, T. E. (2003). *The inclusive classroom: Strategies for effective instruction* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Pressley, M., & McCormick, C. B. (1997). *Advanced educational psychology for educators, researchers, and policy-makers*. Reading, MA: Addison-Wesley.
- Rosenshine, B. (1997). Advances in research in instruction. In J. W. Lloyd, E. Kameenui, & D. Chard (Eds.), *Issues in educating students with disabilities* (pp. 197–220). Mahwah, NJ: Lawrence Erlbaum.
- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York: Macmillan.
- Scruggs, T. E., & Mastropieri, M. A. (1995). What makes special education special? An analysis of the PASS variables in inclusion settings. *Journal of Special Education*, 29, 224–233.
- Sindelar, P. T., Smith, M. A., Harriman, N. E., Hale, R. L., & Wilson, R. J. (1986). Teacher effectiveness in special education programs. *Journal of Special Education*, 20, 195–207.



# Elder Abuse

Michael A. Rodríguez

University of California, Los Angeles, California, USA

Donna Benton

Los Angeles Caregiver Resource Center, University of Southern California, Los Angeles, California, USA

1. Introduction
  2. Law
  3. Theories of Abuse: Risk Factor Profiles
  4. Clinical Management of Elder Abuse
  5. Ethical Issues
  6. Future Directions
  7. Summary
- Further Reading

## GLOSSARY

**biopsychosocial model** Model that describes how the interaction between the social, physical, and psychological characteristics of individuals results in a physical, mental, and emotional health outcome that influences their future behavior.

**elder abuse** (1) Intentional actions that cause harm or create a serious risk of harm (whether or not harm is intended) to a vulnerable elder by a caregiver or other person who has a relationship of trust with the elder. (2) Failure by a caregiver to satisfy the elder's basic needs or to protect the elder from harm.

**feedback loop** Scheme by which outcomes are continually processed back to the variables that determined the outcome. The new outcomes result in new interpretations of the variables, which results in new outcomes.

**financial abuse** Misuse of another person's funds or theft of money, property, or possessions.

**neglect** A type of physical abuse that can be either intentional or unintentional. In terms of elder abuse, neglect is

the failure to provide the basic needs of the elder or to provide goods or services that are necessary to avoid or prevent physical harm, mental anguish, or mental illness.

**physical abuse** An act that may result in pain, injury, and/or impairment. Physical abuse includes bodily harm, neglect by others, and medical mismanagement.

**psychological abuse** A deliberate act inflicted on another to cause mental anguish.

**sexual abuse** Forcing another to take part in any unwanted sexual activity.

**stakeholders** Those who are invested in the lives of a particular individual. For the elderly, stakeholders can be health care personnel, social service agencies, friends, and families.

**transgenerational violence** Abuse that is "passed down" or continually generated throughout generations of a family.

Elder abuse is a social and public health problem that affects over half a million abused elderly victims in the United States as well as their caregivers, family, and community institutions. Abuse results in physical, emotional, and mental angst for victims, some of whom are cognitively and/or functionally impaired. There are also societal economic repercussions of elder abuse due to the increased need and demand for greater health care and social services to assist the victims. Although health care and law enforcement participation have helped assuage the situation, elder abuse detection, assessment, and treatment remain challenging. Adding to the difficulty of combating elder abuse is the booming

geriatric population, which is predicted to increase the strain on caregivers and societal institutions. This stress may perpetuate greater incidences of elder abuse in the future. Currently, multidisciplinary teams of psychologists, physicians, social workers, and public health advocates work together to educate, prevent, and intervene earlier in potential elder abuse cases. This article discusses the ways in which professionals are attempting to ameliorate elder abuse, in addition to providing a brief background of the causes and effects of this problem. The discussion includes methods of identifying victims and abusers, assessment of abuse, methods of intervention, and factors to account for in treatment.

## 1. INTRODUCTION

### 1.1. Background

Elder abuse has been a phenomenon recognized in medical and social practice since the 1970s, when the terms “granny battering” and “granny bashing” were mentioned in literature from the United Kingdom. In the same decade, work on aging populations established the presence of similar incidents in the United States. Terms such as elder mistreatment, elder abuse, and battered elders syndrome have variously attempted to describe abuse against elderly individuals.

### 1.2. Definition

The Institute of Medicine defines elder abuse (or elder mistreatment) as (1) intentional actions that cause harm or create a serious risk of harm (whether or not harm is intended) to a vulnerable elder by a caregiver or other person who stands in a trust relationship to the elder, or (2) failure by a caregiver to satisfy the elder’s basic needs or to protect the elder from harm. While the terms elder mistreatment and elder abuse are frequently used to describe this behavior, the term elder abuse will be used in this article.

### 1.3. Types of Abuse

Elder abuse occurring either in the home or in an institution can be described as one or more of the following types: physical abuse, neglect, psychological abuse, financial abuse, and sexual abuse. This article focuses on the issues of elder abuse as it occurs in the home.

#### 1.3.1. Physical

Physical abuse is an act that may result in pain, injury, and/or impairment. Physical abuse includes bodily harm, neglect by others, medication misuse, and medical mismanagement. Bodily harm or physical assault can take many forms, such as beating, shaking, tripping, punching, burning, pulling of hair, slapping, gripping, pushing, pinching, kicking, and the use of physical restraints.

**1.3.1.1. Neglect and Self-Neglect** Neglect, which is considered a type of physical abuse, can either be intentional or unintentional. Intentional neglect is the deliberate failure to provide the basic needs of the elder. It also includes the failure to provide goods or services that are necessary to avoid or prevent physical harm, mental anguish, or mental illness. Unintentional neglect occurs when the caregiver is not knowledgeable about the elder’s needs or when the caregiver is restricted in the care they can provide due to his or her own infirmities.

Self-neglect is a failure to engage in activities that a culture deems necessary to maintain socially accepted standards of personal or household hygiene and to perform activities needed to maintain health status. Self-neglect has been categorized as a type of physical abuse by some authors. The highest national percentage of reported cases of abuse are those of self-neglect.

#### 1.3.2. Psychological/Emotional Abuse

Psychological and/or emotional abuse can occur independently or can be related to physical abuse such as neglect or other forms of abuse. It is a deliberate act inflicted on the elder that is intended to cause mental anguish. Psychological abuse may include isolation, verbal assault (name calling), threats that induce fear (intimidation), humiliation, harassment, ignoring, infantilization, and emotional deprivation.

#### 1.3.3. Financial Abuse

Financial abuse is the misuse of an older person’s funds or theft of money, property, or possessions.

#### 1.3.4. Sexual Abuse

Sexual abuse involves forcing the elder to take part in any unwanted sexual activity, such as touching that makes the elder feel uncomfortable or photographing the elder person while he or she is changing clothes or bathing.

### 1.4. Epidemiology of Abuse

The National Center on Elder Abuse (NCEA) reports that in 1996 an estimated 551,011 persons aged 60 and over experienced abuse, neglect, and/or self-neglect in a 1-year period. The report also adds that there were four times as many new, unreported cases of elder abuse. Another study noted that more than 105,000 elderly Americans were victims of non-fatal violent crimes in 2001.

An anticipated increasing geriatric population is predicted to expand the elder abuse problem. According to the 2000 U.S. census, the number of individuals over the age of 65 has increased from 31.2 million in 1990 to 35 million in 2000. Population specialists predict that with these continuing trends, by the mid-21st century there will be more elderly people than young people in the United States. The increasing number of senior citizens creates greater dependency on the immediate and extended family and demand for health and social services to accommodate to the increased longevity and chronic medical care of the elderly. Higher costs of health care, inadequate support for caregiving, and personality conflicts between the elder and the family or caregiver can strain relationships and may contribute to the possibility of abuse.

## 2. LAW

Federal policymakers first addressed the issue of elder abuse through the Older Americans Act of 1965, which was followed by the Vulnerable Elder Rights Protection Program in 1992. Since then, states have developed their own laws governing elder abuse management. State elder abuse statutes have primarily been based on laws developed to address child abuse and spousal/intimate partner abuse. The primary focus of elder abuse laws in most states has been mandatory reporting statutes and follow-up investigational procedures. Forty-two states have mandatory reporting laws, and eight states have voluntary reporting requirements. Psychologists are mandated reporters in 29 states and are encouraged to report elder abuse in five states. The investigating agency and the scope of authority of the agencies varies from state to state, but the Adult Protective Services (APS), the Long-Term Care Ombudsman, and the local law enforcement agency are the most commonly recognized organizations to whom a report can be made about suspected or confirmed elder abuse. Penalties for not reporting elder

abuse, the time frame for emergency reports, and the maximum length of investigation also depend on the state. It is recommended that interested parties become familiar with the reporting requirements specific to each state (<http://www.elderabusecenter.org>).

## 3. THEORIES OF ABUSE: RISK FACTOR PROFILES

The literature suggests that all older adults may be at risk of abuse. Reviews of the literature have found conflicting profiles of both victims and possible abusers, indicating that it is not fitting to eliminate any particular profile; nor is it appropriate to say that a typical profile exists. The following sections discuss factors clinicians should be aware of when working with older adults.

### 3.1. Impairment of the Elder

Alcohol abuse by the elder is one of the most common risk factors for elder abuse. Although research shows variations in the impact of alcohol abuse on elder abuse, it has been found that the victim's risk for elder abuse could increase up to 10-fold. Other commonly established risk factors for elder abuse include low cognitive abilities and physical functional impairment. Those with depression, dementia and psychiatric illness are more likely to be abused. Elders who have difficulty eating are also at greater risk of abuse. These impairments may lead elders to live with family members. Elders are more likely to be abused if they live with family but do not have a living spouse. This places women, who are more likely to be widowed, at a greater risk for abuse. However, women may also be abused when they care for a spouse who may have been or continues to be abusive.

### 3.2. Perpetrator Characteristics

Other risk factors of elder abuse have also been proposed. Anetzberger suggested that elder abuse is primarily a function of the perpetrator's characteristics. This means that the caregiver's problems, pathologies, and perceptions may be the key to understanding who becomes an abuser. Per G. Anetzberger, life stresses, financial problems, mental disabilities, and lack of empathy for older people with disabilities may render some caregivers "ill-suited for caregiving and, given the potential dynamics associated with

caregiving, can make them prone to inflict abuse” (2000, p. 48). Research from the Three Model Projects on Elder Abuse found that in many cases the abuser was emotionally or financially dependent on the victim. Substance abuse by the caregiver was also related to risk for elder abuse.

### **3.3. The Interaction between Caregiver and Care Receiver**

The caregiver–elder relationship also affects the potential for abuse. Risk may increase if the caregiver perceives the care recipient as difficult, combative, excessively dependent, or any combination of these. Homer and Gilleard found that socially disruptive behavior by the care receiver was related to elder abuse. However, others have found no such relationship. The duration, type, and intensity of care needed, cultural values, and individual family expectations also contribute to the possibility of abuse. Among ethnic minority caregivers, burden may be less likely to be an issue expressed by caregivers. Some studies have found risk factors associated with non-white or ethnic minority populations.

### **3.4. Multi-Factor Clinical Assessments**

Given the conflicts that have emerged in research, the prudent clinician will continue to use all of these factors in his or her assessments. Many clinicians often overlook the importance of understanding the caregivers of older adults. Clinicians should recognize that caregiving is a risk factor for abuse, and paying close attention to the caregivers’ attitude toward their roles may be useful in preventing abuse from occurring or being overlooked. Therefore, they need to observe the dynamics of the family relationship when meeting either the caregiver or the care receiver.

## **4. CLINICAL MANAGEMENT OF ELDER ABUSE**

Elder patients can self-refer or be referred for a psychological review by their primary care physician or other health care provider. Whether the health care provider refers elders for suspected elder abuse or for changes in emotional and/or psychological and cognitive status, it is necessary that psychologists always screen elder patients for incidences of abuse.

The theoretical model that guides clinical management is described more fully in a 2003 Institute of Medicine Report. This model is adapted from Engel’s biopsychosocial model. According to the elder abuse model, the social, physical, and psychological characteristics of the elder interact with those of the caregiver or trusted other. The interaction between the elder and the caregiver or trusted other person occurs in an environment governed by the socioeconomic conditions of the involved parties, the level of economic dependency, and the normative expectation of other stakeholders (health care personnel, social service agencies, friends, and relatives). Also important to the interaction between the caregiver and the elder is the sociocultural context in which they live, which encompasses the institutional or organizational locus (such as nursing home, private household), race or ethnic group, and social network of the elder and caregiver. The presence of social ties of the stakeholders to the elder and the caregiver may serve as a monitoring control on their interaction. Literature suggests that absence of this social network increases the vulnerability of the elder to the risk of abuse.

The physical, mental, and emotional health outcomes of the interaction between the elder and the caregiver impact their future behavior. These outcomes continually feedback to and remold the elder’s and caregiver’s social, physical, and psychological characteristics that will produce new outcomes. The feedback loop may sometimes result in the occurrence of abuse, either as a one-time incident or as a progression of violent episodes. This process-oriented model can help expand understanding of the etiology of specific types of elder abuse, help develop suitable interventions, and eventually lead to a reversal of the process.

### **4.1. General Evaluation and Analysis**

#### **4.1.1. Approach**

The person who elicits the elder’s history during a clinical visit should be patient and tolerant. Elders who arrive at a clinic for assessment may speak slowly. It is important to talk clearly and slowly to an elder patient, as he or she may have hearing impairments. Care should be taken not to infantilize the patient and not to subscribe to ageist attitudes and myths about the elderly such as forgetfulness, senility, dependency, ineptness, unproductivity, and unattractiveness. Cultural and ethnic differences between the physician and patient must be respected and considered. Tact, belief,

and discretion are paramount to developing a trusting relationship with the patient.

#### **4.1.2. Observation**

Anyone who assesses an elder for signs of abuse needs to pay close attention to patient–caregiver interactions. Increased discomfort, silence or monosyllabic responses, and fear exhibited by the patient when the caregiver is present may be indicators of elder abuse. In such situations, it is advisable to interview the patient when the caregiver is not present. Patients' responses and body language need to be critically observed. Fear, anger, infantile behavior, agitation, rocking (in the absence of other motor diseases), and sucking are some behavioral responses that may indicate possible abusive situations in the patient's life. Other indicators include, but are not limited to, confusion or disorientation while providing responses, withdrawal, denial of events and happenings, failing to talk openly, and providing implausible stories. Physical signs provide vital clues to possible elder abuse incidences. Cuts, lacerations, bruises, welts, dehydration, loss of weight, and burns should raise suspicion of abuse.

#### **4.1.3. History Taking**

**4.1.3.1. Past History** Theories about elder abuse indicate the role of transgenerational violence as a high risk factor for violence against a senior adult. Past history of domestic violence in the life of the victim (either as the perpetrator or the victim) needs to be elicited with a nonjudgmental and sensitive approach. A history of past relationships, family dynamics, the number of household members, education levels, and available societal resources such as visits from friends, other relatives, and neighbors provide clues to detect isolation and incidences of neglect. History of substance abuse (alcohol and or chemical dependency), employment status, housing, and financial status provide insights into possible cases of dependency and potential abuse. Sexual history should also be considered.

**4.1.3.2. Family History** Death of a spouse or partner and history of alcoholism among family members are areas that need to be explored. Serious psychiatric disorders in the family might provide a clue to the level of caretaking and dependency that the elder might be responsible for in his or her home.

#### **4.1.4. Screening**

Detection of elder abuse can be accomplished by routinely screening for abuse. The American Medical Association urges every clinical setting to use a routine protocol for the detection and assessment of elder mistreatment. There are several instruments available for screening the patient and/or the caregiver. **Table I** contains a list of screening methods and instruments along with information on their measurement properties. Once a case is identified as suspected positive, an intervention plan has to be developed and applied with the consent of the patient. The following sections describe other assessments that aid in identifying incidences of elder abuse.

### **4.2. Assessment of Cognitive Status and Personality of the Patient**

Psychological assessment of an elder should include a comprehensive assessment of age-related conditions such as anxiety disorders, depression, mood disorders, sleep disorders, sexual dysfunction, substance abuse, and personality. Especially important are those tests that assess for risk factors of elder abuse, such as alcohol abuse and depression. The American Psychological Association recommends the following measures for assessing the psychological and cognitive status of the elder patient: The Beck Anxiety Scale, the Hamilton Depression Rating Scale, and other oral and written forms of depression-assessing instruments such as the Geriatric Depression Scale (GDS). Some authors also recommend the Beck Depression Inventory or the Yesavage Depression Scales. The CAGE instrument assesses alcohol usage. Additionally, the geriatric version of the Michigan Alcoholism Screening Test can be specifically used with older adults. The Mattis Dementia Rating Scale or Cummings Inventory for Alzheimer's Dementia helps evaluate dementia.

Health care providers need to look for overt and covert clinical signs of the patient's psychological and cognitive status during evaluations. Overt signs such as crying, silence, and irritability indicate a possible depressive state or social withdrawal; they could also be learned passive behavior. Prior to performing any assessments, the elder patient should be familiarized with the tests and procedures. Elders should have appropriate assistive devices such as eyeglasses and hearing devices. For those who are non-English speakers, the tester should be able to converse in the language spoken by the elder or seek the assistance of a professional interpreter.

TABLE I  
Elder Mistreatment Measures<sup>a</sup>

Measure	Summary	Characteristics	Properties
The QUALCARE Scale (Philips <i>et al.</i> , 1990a, 1990b)	Assessment of six areas: physical, medical management, psychosocial, environmental, human rights, and financial	Fifty-three-item observational rating scale designed to quantify and qualify family caregiving	Extensive psychometrics reported: Interrater agreement range: 0.79–0.88. Chronbach's alpha: 0.81–0.95 on six subscales
Hwalek-Sengstock Elder Abuse Screening Test	Assessment of physical, financial, psychological, and neglectful situations	Fifteen-item assessment screen for detecting suspected elder abuse and neglect	Discriminant function analysis: nine items identified 94% of cases. Three conceptual domains: violation of personal rights, characteristics of vulnerability, and potentially abusive situations
Fulmer Restriction Scale (Fulmer and Gurland, 1996)	Assessment of physical, psychological, and financial restriction of older adults	Thirty-four-item scale designed to elicit information regarding unnecessary restriction of the older adult	Chronbach's alpha: 0.78 Interrater agreement: 0.93 Available in Spanish

<sup>a</sup>Adapted from Bonnie, R. J., & Wallace, R. B. (Eds.) (2003). Appendix A: Elder mistreatment measures and studies. In *Elder mistreatment, abuse, neglect, and exploitation in aging America*. Washington DC: The National Academies Press.

### 4.3. Risk Factor Assessment

An assessment of possible risk factors by a home care team or social worker can help identify needs and possible resources. Incidences of substance abuse, the availability of a caregiver and/or a support system from the patient's family or friends, financial dependency on the patient by other family members, and caregiver stress are factors that influence the plan of action. Any course of action that is recommended by the psychologist should be relayed to the elder's primary care physician. Available options should be discussed with the patient and the social worker if possible.

### 4.4. Treatment and Management

Once suspicion of elder abuse has been confirmed, the next step is to establish a plan of action to minimize the abuse or effects of abuse. A safety assessment is performed to ensure that preventive measures are adapted to the elder's needs. The elder's physical appearance, home environment, and personal surroundings should be considered when determining the immediacy of danger. A multidisciplinary team of professionals

should be involved in the diagnosis, management, and prevention of suspected elder mistreatment.

If the elder is in imminent danger, immediate action should be taken to remove the elder from the situation. One possible method of intervention is to admit the elder to the hospital. In the hospital, the elder will receive medical treatment and not be endangered. Less serious cases of suspected elder abuse can be dealt with through continuous monitoring of the elder. A safety plan may also be developed to ensure that the elder knows what to do if placed in a compromising situation. This may include packing a bag of clothes and keeping copies of important papers, keys, money, and other necessary articles to take if the elder needed to quickly leave his or her home.

If the danger involves the caregiver's burden, an intervention should be designed to lessen the stress of the caregiver through services including home health aides, adult day care, and other types of respite programs. Community resources should also be consulted for possible safety, educational, caretaking, and social support services.

When determining a plan of action, it is important to consider the capacity of the elder and the caregiver, due to the need for cooperation of both parties and the

desire to accommodate the individual to whom services are directed. One such method of intervention is psychotherapy. However, no specific psychological intervention is preferred for elders. Instead, treatment options are individualized, led by the nature of the problem, therapeutic goals, preferences of the involved elder, and convenience for the involved parties. During psychotherapy, the psychologist must remember to be culturally sensitive and respectful of the individual's race, gender, sexual orientation, and social class when assessing issues of mistreatment and formulating interventions. The psychologist should also be attentive to the sensory deficits of the elder, particularly hearing and vision loss, which may hinder communication. Last, psychologist should consider whether psychological symptoms are caused or exacerbated by underlying medical problems. This will enhance the patient-psychologist relationship and enable more effective and productive therapy sessions.

Another method of intervention is education. Elders can learn how to protect themselves and gain empowerment through education. Education may be used to assist psychological interventions by providing additional information regarding the rationale, structure, and goals of psychotherapy. Psychoeducation can also help families caring for cognitively impaired elders. It can familiarize the family with the nature of cognitive loss, problem solving for practical problems, and providing emotional support to cognitively impaired elders who have experienced abuse. Additionally, education teaches other professionals about aging and how to investigate suspected elder abuse. These professional can advocate for the safety of elders while providing psychological services to this population.

Physicians can also prescribe medications to elders to cope with the involved abuse. These prescriptions may enable elders to function well within society or address the repercussions of the abuse. If an elder is abusing his caregiver, medication may be used to normalize the elder's behavior and prevent him or her from doing further harm.

Patient/family cooperation should also be considered when creating a plan of action. For the plan to be useful, it must be implemented by the assisting family. Thus, interventions that consider abused elders' families and their shared values are needed. If the family of the abused elder disagrees with the chosen method of intervention, it is suggested that psychologists be sensitive to the family's opposition and work with them to create modifications to an intervention.

The key to success in each of these intervention methods is the follow-up process. To determine the

efficacy of the proposed intervention, the elder or caregiver must be contacted to ensure that the elder's needs are being met, that a safe environment is preserved, and that wellness of the abused individual is maintained. If interventions are ineffective, physicians must modify their plan of action by re-evaluating the abused elder's needs. One reason that physicians may need to re-evaluate action plans is a change in the elder's access to resources. Interventions must consider the elder's standard of living. This will guarantee the potential for intervention use by the abused elder.

An additional note of concern involves identifying potential challenges to suggested interventions. Challenges may include the resistance of elders to disclosing their personal experiences of neglect or abuse. This resistance may be associated with feelings of embarrassment or shame involving their mistreatment. Elders may also feel uncomfortable with the notion of receiving mental health services for their abuse. Their reservations require physicians to attend to their needs through reassurance of confidentiality during scheduled meetings, by participating in active listening with the elders, and by actively engaging with elders and their caregivers. This may require the enlistment of an individual who has a trusting relationship with the abused victim.

One specific way to counter sentiments of resistance is to validate both the positive and the negative feelings expressed by family members and victims. This validation will cause resistance to slowly diminish as elders begin to feel that their physician is trustworthy and wants to help them. This gradual progression toward an open relationship will foster an environment suitable for the disclosure of sensitive material.

Intervention efforts may be impaired if the elder is cognitively unaware. Should the abused elder suffer cognitive impairment, it is necessary for psychiatrists or geriatricians to ensure that the elder receive adequate care from a reliable caregiver. If the caregiver is a possible abuse perpetrator, or if he or she is unable to handle the elder's impairment, additional respite care or institutionalization should be considered.

## 5. ETHICAL ISSUES

When developing interventions in elder abuse cases, the clinician needs to be cognizant of the basic principles of autonomy, justice, beneficence, and nonmaleficence.

### 5.1. Autonomy

Autonomy is the right of self-determination. It is the right to choose one's actions or course in life as long as they do not interfere unduly with the lives and actions of others. A clinician must respect the options and choices made by the elder in regards to his or her living situation and caregiver. If the patient decides to continue living in the abusive situation, written information regarding emergency assistance, a follow-up plan, and a safety plan should be developed in consultation with the victim. However, it is imperative to assess the cognitive status and mental stability of the elder and caregiver.

### 5.2. Justice

The principle of justice allows everyone the right to that which he or she is due. Some elder patients may not receive adequate care because the caregiver and/or family is exploiting the finances of the patient. Although the caregiver may use the patient's funds to care for the elder, any unequal distribution of resources that jeopardizes the health and well-being of the elder and that which is done against the wishes of the elder is a violation of the rights of the patient.

### 5.3. Beneficence and Maleficence

Any intervention that is developed to address elder abuse should do no harm to the patient. The psychologist developing interventions should consult the patient, the patient's doctor, and the social worker before deciding on a course of action. Decisions that raise ethical dilemmas should be weighed for the merit of potential good in relation to the potential harm. This involves recognizing the limitations of the available resources for the patient, the caregiver, and the cognitive and general health of the patient and the caregiver.

## 6. FUTURE DIRECTIONS

Psychologists and other public health personnel should educate the public about elder abuse. Culturally relevant materials and messages tailored to suit the needs of minorities conveyed in English and other languages will help inform the general public about the issue. Information for caregivers on their role and the availability of resources in the community will improve

early detection and management of elder abuse. Advocacy for the rights of elders, especially in the area of mental health, to support their social and emotional well-being is necessary. Collaborative research with community organizations and academic institutions involved in research will help identify new areas for development.

## 7. SUMMARY

Forms of elder abuse include physical abuse, neglect, psychological abuse, financial abuse, and sexual abuse. These types of abuse may occur independently or coexist with another form of abuse and may be difficult to detect. When assessing for signs of abuse, a clinician should pay attention to visible marks as well as emotional and mental expressions of fear, anger, confusion, and sadness. The caregiver's circumstance, competency, and relationship to the elder should also be considered in the assessment. One difficulty in recognizing and diagnosing potential elder abuse is that its symptoms and risk factors may resemble otherwise normative age-related illnesses such as dementia and depression. Also, no typical profile exists for either the caregiver or the victim. When assessing for abuse, a range of psychological tests should be performed and extensive past histories and current circumstances should be elicited to gain a full understanding of the elder's situation. After diagnosis, possible methods of intervention include psychotherapy, psychoeducation, and separation of the elder from the caregiver. In formulating an intervention, the needs and abilities of the victim and caregiver should be taken into consideration, as must the rights of the victim. Plans of actions should be routinely re-evaluated to judge the intervention's effectiveness and should be adjusted according to changes in the circumstances of the victim and caregiver. The clinician should approach the process of assessment to intervention with patience for disabilities that the elder might have and without discriminating against factors such as age, culture, gender, and social class.

## Acknowledgments

The authors gratefully acknowledge Kirsten La and Hema Ramamurthi for their assistance with this manuscript.

### See Also the Following Articles

Aging and Competency ■ Elder Caregiving ■ Personality and Emotion in Late Life

### Further Reading

- Abeles, N., & The APA Working Group on the Older Adult Brochure. (1997). *What practitioners should know about working with older adults*. Washington, DC: American Psychological Association. Retrieved on December 3, 2003 from <http://www.apa.org/pi/aging/practitioners.pdf>.
- American Medical Association. (2002). *Diagnosis and treatment guidelines on elder abuse and neglect*. Chicago, IL: American Medical Association. Retrieved on December 9, 2003 from <http://www.ama-assn.org/ama1/pub/upload/mm/386/elderabuse.pdf>.
- Anetzberger, G. J. (2000). Caregiving: Primary cause of elder abuse? *Generations*, 24, 46–51.
- Beauchamp, T., & Childress, J. (1994). *Principles of biomedical ethics* (4th ed.). Oxford, UK: Oxford University Press.
- Benton, D., & Marshall, C. (1991). Elder abuse. *Geriatric Home Care*, 7, 831–845.
- Bonnie, R. J., & Wallace, R. B. (Eds.) (2003). *Elder mistreatment: Abuse, neglect, and exploitation in aging America*. Washington, DC: The National Academies Press.
- Dyer, C. B., Pavlik, V. N., et al. (2000). The high prevalence of depression and dementia in elder abuse or neglect. *Journal of the American Geriatrics Society*, 48, 205–208.
- Engels, G. L. (1977). The need for a new medical model: A challenge for bio-medicine. *Science*, 196, 129–136.
- House, J. S., Umbersome, D., & Landis, K. R. (1988). Structures and processes of social support. *Annual Review of Sociology*, 14, 293–318.
- Lachs, M. S., Berkman, L., et al. (1994). A prospective community-based pilot study of risk factors for the investigation of elder mistreatment. *Journal of the American Geriatrics Society*, 42, 169–173.
- Lachs, M. S., Williams, C., et al. (1997). Risk factors for reported elder abuse and neglect: A nine-year observational cohort study. *The Gerontologist*, 37, 469–474.
- Lauder, W., Scott, A. P., & Whyte, A. (2001). Nurses' judgments of self-neglect: A factorial survey. *International Journal of Nursing Studies*, 38, 601–608.
- Lin, N., Ye, X., & Ensel, W. E. (1999). Social support and depressed mood: A structural analysis. *Journal of Health Behavior*, 40, 344–359.
- The National Center on Elder Abuse at The American Public Human Services Association. (1998). *The National Elder Abuse Incidence Study Final Report*. Washington, DC: National Center on Elder Abuse. Retrieved on November 19, 2003 from [http://www.aoa.gov/eldfam/Elder\\_Rights/Elder\\_Abuse/ABuseReport\\_Full.pdf](http://www.aoa.gov/eldfam/Elder_Rights/Elder_Abuse/ABuseReport_Full.pdf).
- Paveza, G. J., Cohen, D., et al. (1992). Severe family violence and Alzheimer's disease: Prevalence and risk factors. *The Gerontologist*, 32, 493–497.
- Pay, D. S. (1993). *Ask the question: A resource manual on elder abuse for health care personnel*. Vancouver, Canada: BC Institute Against Family Violence. Retrieved on November 13, 2003 from [http://www.bcifv.org/pubs/ask\\_question.html](http://www.bcifv.org/pubs/ask_question.html).
- Quinn, M. J., & Tomita, S. K. (1997). *Elder abuse and neglect: Causes, diagnosis, and intervention strategies* (2nd ed.). New York: Springer Publishing Co.
- Rennison, C. (2002). *Criminal victimization 2001: Changes 2000–2001 with trends 1993–2001*. Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice. Retrieved on January 30, 2004 from <http://www.ojp.usdoj.gov/bjs/pub/pdf/cv01.pdf>.
- Sandefur, R., & Laumann, E. O. (1998). A paradigm for social capital. *Rationality and Society*, 10, 481–501.
- U.S. Census. (2002). *United States Census 2000*. Washington, DC: U. S. Census Bureau. Retrieved on November 15, 2003 from <http://www.census.gov>.



# Elder Caregiving

Allison M. Burton, Jung Kwak, and William E. Haley

University of South Florida, Tampa, Florida, USA

1. Introduction
  2. Overview of Elder Caregiving
  3. Stressors of Caregiving: What Do Caregivers Do?
  4. Individual Differences in Reaction to Caregiving
  5. Racial and Ethnic Diversity Issues
  6. Biopsychosocial Impact of Caregiving
  7. Interventions for Caregivers
  8. Concluding Comments
- Further Reading

## GLOSSARY

**activities of daily living (ADLs)** Basic activities necessary for a person to perform on a daily basis, specifically bathing, dressing, toileting, making transfers, continence control, and eating.

**culture** Worldview, values, norms, and behavior guidelines shared by a group of individuals that influence and form the group's way of life.

**ethnicity** A group's shared personhood based on a distinctive social and cultural heritage passed on from generation to generation.

**formal caregiver** Any person who provides assistance or care to another for financial compensation such as a nurse, a home health aide, or a social worker.

**informal caregiver** Any person who provides assistance or care to another without financial compensation such as a family member, a friend, or a church volunteer.

**instrumental activities of daily living (IADLs)** Basic activities necessary for a person to perform to be able to live independently, specifically managing money, using the telephone, grocery shopping, personal shopping, using

transportation, housekeeping, doing chores, and managing medications.

Caregiving for older adults is a prominent occurrence in the daily lives of many Americans. It is a topic that has received considerable attention from a variety of sources, including research, popular media, and society as a whole. This article offers an overview of the current body of knowledge in the field, including but not limited to specific topics such as the stressors associated with caregiving, the impact of caregiving, and interventions that may be beneficial.

## 1. INTRODUCTION

This article provides an overview of important issues faced by family caregivers, a topic that is of increasing interest in our society. Although illness and disability can strike at any time in the life cycle and lead to care by family members, this article focuses on caregivers for older adults. It reviews the importance and growth of family caregiving; stress process models used to guide research and clinical practice; common stressors faced by family caregivers; the impact of caregiving on psychological, social, and health functioning; individual differences in response to caregiving; and psychosocial interventions for caregivers.

## 2. OVERVIEW OF ELDER CAREGIVING

As the population ages and an increased number of individuals are faced with the responsibility of caring for aging loved ones, the topic of elder caregiving becomes increasingly prominent as an area of research, clinical practice, and public policy. Broadly defined, caregivers are individuals who provide assistance of varying levels to other people who are unable to perform these tasks for themselves. Often the term “caregiver” is associated with a group of individuals who are thought of as homogenous; however, the open nature of the definition allows a high degree of variability within this group. One important distinction is that caregivers can be either paid (formal) or unpaid (informal). Formal caregivers are those who receive financial compensation for the services they perform (e.g., nurses, home health aides), whereas informal caregivers are those who volunteer their time to assist loved ones (e.g., family members, friends). It is estimated that between 80 and 90% of individuals with physical impairments reside in the community and receive the vast majority of their care from informal sources.

The cost of informal caregiving in the United States, including things such as lost wages, keeping an individual out of a nursing home, and performing tasks that otherwise someone would have to be hired to do, was estimated to be \$196 billion in 1997, the year of the most recent national reliable estimates. This dollar figure is more than twice the amount estimated for nursing home care during that same year. Informal caregiving, although substantially less of an economic cost to the family members compared with institutional care, is often coupled with a number of other physical, psychological, and social costs to the individual (discussed in greater detail later).

Individuals choose to take on the responsibility of caring for a loved one for a variety of reasons, including economic conditions, perceived obligation, a cultural ethic that caregiving is expected, and a positive desire to provide care as a meaningful and worthy activity. Formal care for a chronically ill individual is very expensive, whether provided by professionals who come into the home, in an assisted living facility, or in a nursing home. Medicare and Medicaid rarely cover such services fully; thus, many families simply cannot afford such care. However even in situations where financial considerations are not the primary concern, many

families consider it their duty to provide the necessary care and keep their loved ones at home. It is also common for caregivers in these situations to report that they believe they can provide better care than can staff members who do not have personal relationships with their loved ones.

## 3. STRESSORS OF CAREGIVING: WHAT DO CAREGIVERS DO?

Depending on the nature of the illness, caregivers may face a variety of daily demands. These can include impairments in self-care and cognition as well as development of changes in personality and behavior. Other stressors can include physical care demands and issues specific to end-of-life caregiving.

Self-care impairments include higher level instrumental activities of daily living (IADLS) and more fundamental activities of daily living (ADLs). IADLS include tasks such as managing finances, housekeeping, using transportation, and taking medications independently. Loss of independence in IADLS occurs relatively early in the course of progressive dementias but can also occur due to less serious medical problems (e.g., arthritis) that may limit mobility. ADL impairment occurs during middle and later phases of dementia and includes loss of independence in fundamental skills such as bathing, dressing, feeding, and continence of bowel and bladder. Family caregivers who assist with self-care impairments may provide only minor intermittent assistance (e.g., assistance in balancing a checkbook). Assistance with personal care, such as bathing and managing incontinence, is more psychologically stressful because it may involve embarrassing personal care. With assistance with ADLs and IADLS, families may face issues such as role reversal.

Cognitive impairment produces its own unique stressors. In addition to producing decrements in ADL and IADL functioning, dementia can lead to problems such as becoming disorientation, repeating questions, and forgetting to do tasks. Dementia also leads to behavioral problems such as wandering, agitation, and delusional behavior. *Table 1* lists 10 patient problems reported to be the most stressful by a large sample of caregivers for older adults with dementia. Of particular note is that behavioral problems were reported to be much more stressful than self-care problems—even incontinence.

Caregiving for persons with terminal illness creates additional stressors. Beyond the actual tasks of caregiving, other issues include stressors that are more difficult

**TABLE I**  
**Patient Problems Reported as Stressful by Caregivers**

- 
- Dangerous behavior
  - Getting lost
  - Embarrassing behavior
  - Waking the caregiver
  - Agitation and restlessness
  - Anger
  - Suspiciousness
  - Depression
  - Incontinence
  - Repeated questions
- 

*Source.* Adapted from Haley, W. E., Wadley, V. G., West, C. C., & Vetzel, L. L. (1994). How caregiving stressors change with severity of dementia. *Seminars in Speech and Language, 15*, 195–205.

to define. One of these is witnessing the suffering of the care recipients. Caregivers may also be responsible for management of complex medical regimens.

Interaction with the formal care system can be another important stressor. Caregivers report that it is very difficult to negotiate care, manage in-home assistance, and deal with nursing homes. Research has also shown that when caregivers receive formal assistance, it does not decrease their involvement in or commitment to informal caregiving; instead, it simply supplements the informal caregiving activities.

A national survey of family caregivers conducted in 1996 found that the overall mean number of hours of care per week was 17.9; however, this number ranged from 3.6 hours for individuals whose relatives were mildly impaired to 56.5 hours for those whose relatives had significant physical and/or cognitive disabilities. In addition, the stage of the illness is a major factor in affecting how much care a person provides. Research examining spousal caregivers of persons with end-stage lung cancer in hospice has shown that an average of 120 hours per week of care is provided by these individuals. The actual services performed by family caregivers can vary from occasional shopping and household assistance for a frail elder; to assistance with bills and organization for an individual with early-stage dementia; to assistance with dressing, bathing, and ambulation for an individual who has suffered a moderate stroke; to complete responsibility for toileting, eating, and medications for an individual with end-stage cancer.

To get a better understanding of the broader picture of caregiving in the United States, the National Caregiver Survey was conducted in 1996. This study

included 1500 family caregivers of individuals with both dementia and non-dementia diagnoses and provided very valuable insight into the differences between the two. The participants in this study had been providing care for slightly longer than 5 years regardless of diagnosis; however, the dementia caregivers reported significantly more hours per week in the caregiving role. In addition, a significantly higher percentage of dementia caregivers were providing care for each of the ADLs surveyed than were non-dementia caregivers. These findings suggest that the dementia caregivers in this study experienced greater demands while in the caregiving role, which apparently carried over into other areas of their lives. The dementia caregiving group reported greater impacts in terms of family conflict, caregiver strain, time for leisure and other activities, and mental and physical health problems.

#### 4. INDIVIDUAL DIFFERENCES IN REACTION TO CAREGIVING

A number of studies have shown that high levels of stress experienced by caregivers of elders, in particular those with Alzheimer's disease or other types of dementia, often lead to a number of negative psychosocial and health outcomes such as feelings of burden, social isolation, depression, and health problems. However, although most caregivers endure immense stress, they vary in their ability to cope successfully. Researchers have found that it is not necessarily the objective level of stress experienced by caregivers that affects their mental and physical health; rather, it is actually a combination of factors, including context of the caregiving experience, individual appraisal of the caregiving role and competence in that role, coping style, and social support.

First, the context variables, such as the caregiver's gender, age, and health, the quality of the caregiver's relationship to the care recipient, and the length of time providing care, are believed to exert influence throughout the caregiving experience. Socioeconomic status, such as income and education, can affect caregiving outcomes, including morbidity and mortality. It is well established that there is a negative relationship between socioeconomic status and chronic morbidity and mortality, especially among older persons. Socioeconomic characteristics can exert an indirect effect via stress, social relations, and/or health status.

It is common for one individual to take on the role of primary caregiver, with women being the most likely to provide care either for themselves or for their own spouses. It is estimated that 70% of all family caregiving responsibilities are taken on by women and that more than 50% of all women provide this type of care at some point in their life courses. This is significant because studies that examine the effect of gender on the caregiving experience have consistently found that female caregivers may be at greater risk for negative mental health outcomes. Female caregivers tend to report higher levels of depressive symptomatology and anxiety, and lower levels of life satisfaction, than do male caregivers. Such gender effects seem to exist among both spousal and adult child caregivers.

The quality and nature of the relationship between caregivers and care recipients can also influence caregiving outcomes. It is generally recognized that positive social ties have a beneficial influence on general well-being. In caregiving situations, greater affection or intimacy with care recipients can reduce perceived burden or strain and can lower psychological distress among caregivers. Negative social interactions while caregiving, such as family conflict and criticism of caregivers, are a risk factor for greater caregiver depression. Other problematic situations include caregivers who suffer abuse from care recipients and estranged family members who are left with "unfinished business." Spousal caregivers generally spend a substantially greater number of hours per week in the caregiving role, do so for a longer period of time, and report more difficulty in sleeping, less energy, and more fatigue than do nonspousal caregivers. Spousal caregivers are also generally more willing to provide care even when care recipients need extensive care, and they are much less likely to place care recipients in nursing homes, than are adult children. In addition, spousal caregivers consistently have been shown to be in poorer health than noncaregivers. This finding is also shown for adult child caregivers; however, the increased age of spousal caregivers places this group at an even higher risk for personal physical health conditions. These findings, taken together, suggest that spousal caregivers are very committed to their roles and responsibilities, even at the expense of their own health. The vast majority of nonspousal caregivers are adult children for elderly individuals; however, African Americans are more likely to be cared for by members of their extended family than are their European American counterparts. In addition, the societal changes in marital status and number of

children per family have introduced an increasing number of friends, primarily older, never married, childless women, who care for each other, and this population of caregivers is likely to increase during the coming years as the baby boom generation ages.

A second important variable in the caregiving experience is appraisal. Caregivers vary greatly in their subjective reactions to the same circumstances. Caregivers who appraise the problems they face as highly stressful are at increased risk for depression, whereas caregivers who are able to find meaning and benefit from their caregiving experience, and who have high levels of self-efficacy in caregiving, are more resilient to the negative effects of caregiving. Caregiver appraisals are often influenced by caregivers' personalities, with a number of studies showing that neuroticism leads to negative appraisals and higher depression. Anger tends to be one of the most common negative affects among caregivers, especially the ones caring for Alzheimer's patients and those who have poor physical health. One common cause of anger in caregivers that can be addressed therapeutically is inappropriate attributions that care recipients' problems are intentional rather than being the effect of Alzheimer's disease, stroke, or other brain impairments. In contrast, caregivers with personality characteristics such as high levels of optimism may be protected from some of the negative consequences of stress.

Third, research has demonstrated that coping style and social support are critical in explaining the tremendous variability in physical and mental health outcomes for caregivers. The role of resources such as social support and coping strategy is to buffer the effect of stressors on the outcome. In general, informal support networks lessen the negative consequences of caregiving such as emotional distress, health concerns, and economic strain. It is important to note that social support can be viewed not only in terms of quantity but also in terms of subjective perceptions of the quality of support. Coping is considered another important buffer and is usually separated into two categories: problem-focused coping and emotion-focused coping. Caregivers may use problem-focused coping by seeking information about their relatives' conditions, pursuing medical advice or rehabilitative care, and/or arranging for assistance with caregiving duties. In contrast, emotion-focused coping by caregivers may include adaptive efforts such as finding meaning in adversity and distracting themselves from caregiving concerns as well as maladaptive efforts such as engaging in excessive rumination, denial, and/or substance abuse. Problem-focused coping is generally

adaptive but may reach limits when little or nothing can be done to improve the objective aspects of the care recipients' functioning. At such a point, the ability to accept limitations may be most adaptive.

## 5. RACIAL AND ETHNIC DIVERSITY ISSUES

The sociocultural adaptation of the stress and coping model suggests that the role of ethnicity is greater than merely being a status variable reflecting the disadvantaged socioeconomic status of members of ethnic minority groups. This view asserts that one's ethnicity and culture exert an influence on nearly all domains and stages of the stress process model, including who provides care, how caregiving demands are expected to be handled and shared among family members, how caregiving experience is interpreted and perceived, how people cope and deal with stress during caregiving, and what stress outcomes are expected. As to why caregiving experiences differ by race and/or ethnicity, some studies suggest that racial/ethnic differences in values and beliefs about aging and caregiving may be responsible for the observed differences. Although it is dangerous to generalize about racial/ethnic differences due to the risk of stereotyping, the existing literature suggests that there may be important cultural differences in caregiving; however, the individual must clearly be viewed in his or her totality and not just as a member of a racial/ethnic group.

The literature indicates that African Americans often hold different role expectations and attitudes about providing care and coping mechanisms compared with those held by European Americans. African Americans tend to hold more positive attitudes toward older adults and perceive cognitive and/or physical decline in health among elders as part of a normal process of aging. In terms of caregiving experience, African Americans report heavier workloads and more unmet needs than do European Americans; however, African American caregivers also report that they experience lower levels of stress and burden as well as higher levels of satisfaction. Moreover, African American caregivers report increased levels of religiosity in response to caregiving. They are more likely to use their religious faith and activities (e.g., prayer) as a coping strategy than are European American caregivers.

Although studies on Hispanic caregivers are limited and there is great diversity among members of this

ethnic group, traditional values of Hispanic families influence their caregiving experience. One of these traditional values is familism, which emphasizes the importance of the family connection and interaction as well as the family serving as the core of the social support system. Thus, the kinship network is considered to be a very important source of social support for Hispanic elders. Family loyalty and filial piety among Asian Americans, and importance of relational networks among Native Americans, are known to be other important cultural values of ethnic groups that help to create a strong support system among them.

## 6. BIOPSYCHOSOCIAL IMPACT OF CAREGIVING

The beginning of this article noted that although informal caregivers ease the financial burden of long-term care for society, they do so at the risk of experiencing a variety of psychological, physical, and social consequences of caregiving. Research at this point has clearly demonstrated diverse effects of caregiving, but questions remain as to what degree and in which areas such effects occur.

One of the most consistent and well-researched relationships regarding the effects of caring for chronically or terminally ill individuals is the increased risk of depression. The availability of validated quantitative tools for evaluating depression has led to comparability across studies and confidence in the strength of the results. The actual rates of depression vary; however, most studies find that caregivers who are involved in dementia care, or in other extensive caregiving, are two to three times more likely to be at risk for clinically significant depression than are their noncaregiver counterparts. In addition, many caregivers experience subsyndromal or subclinical levels of depression. Caregivers are also much more likely to use psychotropic medications for mental health problems, such as depression and anxiety, than are noncaregivers.

To better understand the relationship between caregiving and increased risk of mental health problems, it is necessary for researchers to identify risk factors. Documented risk factors for increased distress during caregiving include being female, experiencing other stressful life events (e.g., bereavement), lacking social support, experiencing family tension, being younger, being in poorer physical health, ruminative thinking, and having lower education. Research has

also identified a number of factors that appear to be protective against negative mental health outcomes, including high levels of self-esteem and mastery, high ratings of optimism, high ratings of life satisfaction, and low ratings of neuroticism. The findings regarding patient diagnosis have been inconsistent, with some reporting that caregivers for individuals with Alzheimer's disease experience higher rates of depression and others reporting that the rates are comparable between diagnoses such as cancer and Alzheimer's disease.

An important emphasis of research in this area seeks to determine how long these changes in mental health status last, with a fairly large number of studies finding that elevated levels of depression are seen in some individuals for as long as 2 to 5 years following the loss of loved ones, although there is also evidence that some caregivers experience improvements in mental health when their caregiving ends. Additional longitudinal data are needed in this area to better understand what places an individual at risk for this type of chronic bereavement reaction.

Research on the physical health effects of caregiving has varied greatly in its sophistication in assessing health outcomes. A common assessment of physical health status in caregiving studies is a single-item self-rated health question, and more of these caregivers rate their health as poorer than do noncaregiving controls or population norms. Many studies in this area have used this single item as a proxy for physical health in the participants, and although research has demonstrated that this method can be reliable, more sophisticated methods are needed for success in this research.

There have been few longitudinal studies of the impact of caregiving on physical health. One large study comparing caregivers with noncaregivers of similar ages indicated that caregivers showed increased physical health symptoms over a 2-year period, whereas noncaregivers showed no increase over time.

Other studies have used more sophisticated measures of physical health. These studies indicate that caregiving can lead to increased health risk behaviors (e.g., drinking alcohol, smoking cigarettes), lower levels of physical exercise, changes in sleep patterns, decreased immune system functioning, altered response to the influenza vaccination, slower wound healing, increased blood pressure, and altered lipid profiles. There has also been interest in whether or not these physical health changes lead to increased mortality in caregivers, and this question was examined in the Caregiver Health Effects Study. This prospective, population-based

cohort study followed caregivers for an average of 4½ years and reported that caregivers who experience strain while in the caregiving role showed a 63% increase in mortality as compared with caregivers who did not experience strain and noncaregiving controls. Additional work is needed in this area to identify the mechanisms underlying these physical health changes and increases in mortality; however, this research has been very valuable in demonstrating that caregiving can represent a major and clinically significant strain.

Caregivers commonly report that their responsibilities infringe on their personal and social lives. Caregivers often are unable to take part in leisure activities, vacations, and/or valued activities such as church attendance. Spousal caregivers become devoted to their roles and the responsibilities they entail, often at the expense of their own needs, including social interaction. They report spending less time engaging in social activities (e.g., having lunch with friends, spending time with grandchildren) and in personal hobbies (e.g., gardening, sewing). A small group of caregivers, particularly adult daughters, are sometimes described as the "sandwich generation"; in some cases, they are responsible for caring for their own children as well as for their parents. Some researchers believe that this may become an increasing phenomenon given the current trend toward women holding off on having children until later in life. These caregivers very commonly report significant changes in their employment status (e.g., decreased hours, extended leave, leaving the workforce altogether) as a result of time constraints due to the needs of their aging parents. The specific effects vary based on relationships to care recipients; however, the overall pattern appears to be consistent in that caregivers sacrifice their social and personal roles to accommodate the needs of their loved ones.

The cyclic relationship among mental health, physical health, and social engagement is a concern for health care workers and clinicians. Increases in depression and physical health problems can lead to decreased social interaction; however, the reverse is also true in that role constraints that inhibit social activities can lead to negative mental health outcomes, specifically increased depression. An additional concern is in the case of bereavement following a period of caregiving. The literature has demonstrated that low social support and mental and physical health problems during caregiving can lead to complications during bereavement. Therefore, if individuals have experienced these changes as a result of the caregiving experience, they are at risk. It is important to identify caregivers who are at risk for

these negative psychological, physical, and social outcomes so that they can be targeted for assistance before the complications progress.

Finally, it should be noted that caregiving can lead not only to negative impacts but also to benefits. Recent studies show that caregivers often report benefits from caregiving such as a sense of satisfaction, a feeling that they are repaying care they received when they were young, and feelings of mastery and accomplishment. These positive feelings can occur simultaneously with negative effects.

## 7. INTERVENTIONS FOR CAREGIVERS

Each individual in a caregiving situation can experience his or her responsibilities and role as a caregiver differently due to that person's unique characteristics and circumstances. Moreover, caring for loved ones can be a meaningful and positive experience. However, caregiving can also lead to negative psychosocial and physical outcomes such as depression, lowered level of physical health, and lowered sense of well-being. Thus, many caregivers are in need of psychosocial and instrumental support. With an increased knowledge of how various stressors and other caregiving characteristics interact with each other and lead to different outcomes, various intervention methods have been developed and provided to caregivers in different caregiving settings. These interventions have also been evaluated for their clinical significance and effectiveness in both statistical and practical ways. A number of studies that summarize, evaluate, and examine the strengths and weaknesses of currently available caregiving intervention research literature have been published. These review studies can provide further detailed knowledge and information on the progress of caregiving intervention research. However, this article focuses on providing an overview of the caregiving intervention research, including primary entities being targeted by the interventions, types of interventions, and effectiveness of these interventions.

In general, interventions need to be tailored to specific needs of caregivers who have diverse and unique characteristics and caregiving experiences. For example, a caregiver who is dealing with severe behavioral problems due to dementia might want to learn about the nature and course of disease progression and how to handle behavioral problems effectively. On the other hand, a caregiver who has recently placed a loved one in a

nursing home might seek an intervention that can help him or her to cope with feelings of guilt. Therefore, caregiving interventions are designed to reduce different types of stress outcomes so as to meet unique individual needs of caregivers and their care recipients.

There are a few major domains of stress outcomes for which various intervention methods aim to improve. These outcomes include how caregivers respond to stress in terms of mental and physical symptoms, improved general sense of well-being and quality of life, and delayed institutionalization of care recipients. Symptoms that caregivers often exhibit in reaction to caregiving stress include elevated levels of depression, anxiety, anger, and hostility as well as lowered health status (often measured by self-reported health and blood pressure). Also, quality of life of caregivers is often measured broadly by life satisfaction, sense of burden, mood, and social support. Finally, other outcomes targeted for intervention include care recipient outcomes such as functional status and delay of institutionalization of care recipients and use of formal and informal support as well as services available to caregivers.

Caregiver interventions are focused on either improving quality of life and coping skills of caregivers or reducing the objective stressors such as amount of care provided by caregivers. These interventions include psychoeducational, psychotherapeutic, supportive, and service-based interventions. The psychoeducational intervention is often offered with other types of interventions and aims to improve caregivers' knowledge about themselves, care recipients, and the environment. This type of intervention focuses on teaching caregivers to develop skills in dealing and coping with stressors through providing comprehensive information on characteristics of disease process of care recipients, resources and services that caregivers can use, and other specific coping skills. Psychotherapy is another form of intervention that is often used in conjunction with psychoeducation involving individual counseling to caregivers by trained professionals. This type of intervention is usually designed to teach coping skills and problem-solving techniques with the cognitive behavioral approach.

Supportive intervention is another type, and its main goal is to build support systems or networks among caregivers and to create an environment for them to discuss and share problems and feelings related to caregiving. Thus, supportive intervention is focused on allowing participants (i.e., caregivers) to exchange ideas and strategies for coping with stressors and

difficulties. Service-based intervention programs usually provide respite and/or adult day care services. Respite and adult day care programs are designed for caregivers to take some time off from caregiving responsibilities by providing assistance with ADLs and/or certain activities for care recipients. Other types of intervention methods include providing memory training, activities, and medications to care recipients as well as relaxation training to caregivers.

Most of these interventions are multifaceted in their design and method, and they usually are offered in conjunction with other methods. This makes it difficult to attribute any specific component of intervention to the success of intervention outcomes. Moreover, the caregiver intervention literature shows that most of these interventions show some level of effectiveness, yet it appears that there is no single method that is clinically significant and effective across caregivers. However, in general, a combined form of individual educational and psychotherapeutic interventions tends to result in better outcomes than do group interventions. Moreover, psychoeducational and psychotherapeutic interventions are found to be effective in reducing

levels of depressive symptomatology as well as anxiety and anger. In addition to psychoeducational and psychotherapeutic interventions, behavior management training, stress management, and use of certain medications are known to be effective in increasing sense of well-being and quality of life of caregivers as well as in delaying nursing home placement of care recipients with Alzheimer's disease.

The majority of caregivers are likely to benefit from gaining knowledge about the diseases, caregiving tasks, and available resources. Some caregivers can obtain additional help by receiving training in coping and problem-solving skills as well as in behavioral management of care recipients. In addition, some studies have found that providing a mixed form of interventions that are given to both caregivers (e.g., psychoeducational intervention) and care recipients (e.g., medication, therapeutic activities) can also be effective. Furthermore, a fairly new approach to caregiving interventions involves using technologies such as enhanced telemedicine, computers, and the Internet. Table II provides an overview of some of the types of interventions described previously, including definitions and examples.

TABLE II  
Definitions and Types of Caregiving Intervention

Type	Definition	Example
Psychoeducational intervention	This aims at improving knowledge of caregivers about themselves, care recipients, and the environment. This type of intervention focuses on teaching caregivers to develop skills to deal and cope with stressors.	Problem solving, coping skills training
Psychotherapeutic intervention	This involves individual counseling to caregivers by trained professionals. It is mostly designed to teach coping skills and problem-solving techniques with the cognitive behavioral approach.	Cognitive behavioral therapy
Supportive intervention	The main goal is to build a support system or network among caregivers and to create an environment for them to discuss and share problems and feelings related to caregiving.	Support group
Service-based intervention	Goals include facilitating caregivers' use of formal services, improving competence of care recipients, and delaying institutionalization of care recipients. Services provided include respite care for caregivers and drug and activity therapy for care recipients.	Respite care, adult day care, medication therapy, and memory clinic (for care recipients with dementia)

Source. Adapted from Schulz, R., O'Brien, A., Czaja, S., Ory, M., Norris, R., Martire, L. M., Belle, S. H., Burgio, L., Gitlin, L., Coon, D., Burns, R., Gallagher-Thompson, D., & Stevens, A. (2002). Dementia caregiver intervention research: In search of clinical significance. *The Gerontologist*, 42, 589-602.

## 8. CONCLUDING COMMENTS

Family caregiving is a topic that demands greater attention from researchers, clinicians, and those concerned with public policy. The aging of the baby boom generation may stimulate greater attention to the possibility of changes in public policy that provide more support for caregiving families. For example, in other countries, family caregivers may receive pay in recognition of their efforts, adjustments to their retirement benefits, vouchers that can be used to purchase community services, and other forms of support that recognize the value of family caregivers in preventing expensive and unnecessary institutional care.

### See Also the Following Articles

Coping ■ Stress

### Further Reading

- Aneshensel, C. S., Pearlin, L. I., Mullan, J. T., Zarit, S. H., & Whitlatch, C. J. (1995). *Profiles in caregiving: The unexpected career*. San Diego: Academic Press.
- Aranda, M. P., & Knight, B. G. (1997). The influence of ethnicity and culture on the caregiver stress and coping process: A sociocultural review and analysis. *The Gerontologist*, 37, 342–354.
- Coon, D., Gallagher-Thompson, D., & Thompson, L. W. (Eds.). (2002). *Innovative interventions to reduce dementia caregivers' distress: A clinical guide*. New York: Springer.
- Haley, W. E., Allen, R., Reynolds, S., Chen, H., Burton, A., & Gallagher-Thompson, D. (2002). Family issues in end-of-life decision making and end-of-life care. *American Behavioral Scientist*, 46, 284–297.
- Haley, W. E., Roth, D. L., Coleton, M. I., Ford, G. R., West, C. A. C., Collins, R. P., & Isobe, T. L. (1996). Appraisal, coping, and social support as mediators of well-being in Black and White family caregivers of patients with Alzheimer's disease. *Journal of Consulting and Clinical Psychology*, 64, 121–129.
- Kiecolt-Glaser, J. K., Marucha, P. T., Malarkey, W. B., Mercado, A. M., & Glaser, R. (1995). Slowing of wound healing by psychological stress. *The Lancet*, 346, 1194–1196.
- Mace, N. L., Rabins, P. V., & McHugh, P. R. (1999). *The 36-hour day: A family guide to caring for persons with Alzheimer's disease, related dementing illnesses, and memory loss in later life* (3rd ed.). Baltimore, MD: Johns Hopkins University Press.
- Mittelman, M. S., Ferris, S. H., Shulman, E., Steinberg, G., & Levin, B. (1996). A family intervention to delay nursing home placement of patients with Alzheimer's disease: A randomized controlled trial. *Journal of the American Medical Association*, 276, 1725–1731.
- Pearlin, L. I., Mullan, J. T., Semple, S. J., & Skaff, M. M. (1990). Caregiving and the stress process: An overview of concepts and their measures. *The Gerontologist*, 30, 583–594.
- Roth, D. L., Haley, W. E., Owen, J. E., Clay, O. J., & Goode, K. T. (2001). Latent growth models of the longitudinal effects of dementia caregiving: A comparison of African-American and White family caregivers. *Psychology and Aging*, 16, 427–436.
- Schulz, R., & Beach, S. R. (1999). Caregiving as a risk factor for mortality: The Caregiver Health Effects Study. *Journal of the American Medical Association*, 282, 2215–2219.
- Schulz, R., O'Brien, A., Czaja, S., Ory, M., Norris, R., Martire, L. M., Belle, S. H., Burgio, L., Gitlin, L., Coon, D., Burns, R., Gallagher-Thompson, D., & Stevens, A. (2002). Dementia caregiver intervention research: In search of clinical significance. *The Gerontologist*, 42, 589–602.
- Sorensen, S., Pinguart, M., & Duberstein, P. (2002). How effective are interventions with caregivers? An updated meta-analysis. *The Gerontologist*, 42, 356–372.
- Stephens, M. A. P., & Franks, M. M. (1995). Spillover between daughters' roles as caregiver and wife: Interference or enhancement? *Journal of Gerontology: Psychological Sciences B*, 50, P9–P17.
- Teri, L., Logsdon, R. G., Uomoto, J., & McCurry, S. M. (1997). Behavioral treatment of depression in dementia patients: A controlled clinical trial. *Journal of Gerontology: Psychological Sciences B*, 52, P159–P166.
- Weitzner, M., Haley, W. E., & Chen, H. (2000). The family caregiver of the older cancer patient. *Hematology/Oncology Clinics of North America*, 14, 269–281.
- Zarit, S. H., Reever, K. E., & Bach-Peterson, J. (1980). Relatives of the impaired elderly: Correlates of feelings of burden. *The Gerontologist*, 20, 649–655.
- Zarit, S. H., Stephens, M. A. P., Townsend, A., & Greene, R. (1998). Stress reduction for family caregivers: Effects of adult day care use. *Journal of Gerontology: Social Sciences B*, 53, S267–S277.



# Emotion

Belinda Campos, Dacher Keltner, and Molly Parker Tapias

University of California, Berkeley, California, USA

1. Introduction
  2. Defining Emotion
  3. Measuring Emotion
  4. Social Functions of Emotion
  5. Emotions and Health
  6. Emotions and Work
  7. Emotions in School
  8. Conclusion
- Further Reading

## GLOSSARY

**anger** An emotion elicited by the appraisal that a demeaning offense or injustice has been committed against one's self or one's own; anger communicates the intent to redress the perceived wrong.

**anxiety** An emotion characterized by threat and vigilant cognition that can impair performance when conditions are perceived as providing a challenge that exceeds capacity; low levels of anxiety can enhance performance, but high levels of anxiety can impair ability.

**emotional expression** Displays of experienced emotion in the face, body, voice, gesture, and gaze; expressions of emotion convey important information to others about the sender's current state, likely behavior, and enduring dispositions and help to shape social interactions in important ways.

**happiness** The broadest and most recognizable positive emotion; the display of happiness is characterized by the Duchenne smile, which involves movements of muscles around the mouth and eyes.

**interpersonal** Emotion processes that occur at the level of ongoing social interaction where two or more people are interacting; in this more social context, emotions coordinate social behavior and smooth the progress of social relations.

**intrapersonal** Emotion processes that occur at the level of the individual, including the cognitions, physiology, perceptions, and behaviors associated with individual response to environmental opportunities or challenges.

**social function** The particular problem that an emotion was designed to solve and how it does so by structuring coordinated social interactions; recent theory argues that emotions solve problems related to reproduction and the raising of offspring, the problems of group governance and cooperation, and the negotiation of social hierarchies.

Emotions are distinct, biologically based psychological entities characterized by short-lived coherent patterns of perception, experience, physiology, and communicative action. At both the intrapersonal and interpersonal levels, emotions shape experience and outcomes in important domains of life, including health, work, and school.

## 1. INTRODUCTION

What role do emotions play in human social life? Western philosophy has traditionally regarded emotions with suspicion, as a disruptive force. The Stoics advocated distancing oneself from an emotional life,

whereas more recently Hobbes and Freud believed that following natural inclinations would lead to uncontrolled emotion and societal havoc. Still others, such as Aristotle, took the perspective that at least certain emotions (e.g., anger, jealousy) should be controlled or moderated.

Mounting evidence and theory within psychology has led to the opposite view of emotion, that is, that emotions are well-organized coherent systems that enable humans to respond quickly and effectively to environmental and social problems and opportunities. This article outlines this latter perspective, termed a social functional approach, and delineates how social functional approaches lead to an improved understanding of how emotions influence health, work, and school.

## 2. DEFINING EMOTION

Emotion is easily recognized and difficult to define. Currently, emotion researchers define an emotion as a short-lived, biologically based pattern of perception, experience, physiology, and communication that occurs in response to specific physical and social challenges and opportunities. This definition helps to differentiate emotions from related phenomena. Whereas emotions are elicited by fairly flexible interpretations of stimuli and have specific intentional objects, moods have less specific causes and endure for longer periods of time. Sentiments are consistent emotional responses to objects that can endure for lifetimes and be passed from one generation to the next (e.g., intergroup hostility). Emotion-related traits, such as hostility and shyness, reflect consistent ways of responding emotionally to broad classes of stimuli. Sensations (e.g., pain, an itch) are not explicitly related to the individual's strivings.

Most emotion researchers now agree that emotions serve adaptive functions but differ in the questions for which they seek answers. Psychologists who study the structure of emotion systems consider the origins of specific emotions, the behavior associated with emotion experience, and the personal and social consequences of behavior prompted by emotion. Psychologists who study questions of communication and cross-cultural universality focus on facial displays of emotion, taking care to document the distinct muscle movements that characterize some emotions. Their counterparts who are more interested in cultural variation use ethnographic or lexical methods that emphasize unique aspects of emotional experience

such as emotion words with no ready translations outside of the society of origin. Researchers who study appraisals of events that lead to emotional experience use narrative methods to study the patterns of stimulus interpretation that differentiate emotions. These various lines of inquiry have led to remarkable discoveries during the past 20 years or so regarding the role of emotion in development, psychopathology, memory, and brain function.

## 3. MEASURING EMOTION

Emotions have many facets and can be measured in many ways. Researchers interested in the expression of emotion can capture facial expression with the Facial Action Coding System (FACS) developed by Ekman and Friesen or can rely on ethologically significant gaze, posture, vocal, and/or gestural activities. These methods are fairly nonobtrusive and allow researchers to study emotion outside of the laboratory. Researchers interested in emotion-related appraisal or the understanding of emotion have turned to narrative methods and interview techniques and elaborate coding schemes to discern the meaning processes that give rise to and differentiate emotions from one another. Categorizing emotion words into prototypes of an experience is a method used to understand how individuals conceptualize everyday emotion experience. Other researchers have recorded activity in the autonomic nervous system to ascertain when emotions occur and how they differ from one another, for example, by assessing heart rate, respiration pattern, blood flow to the periphery, and skin conductance. Finally, a great deal of research continues to rely on self-reports of emotion experience through the administration of either formal Likert-type scales, such as the PANAS (which measures the disposition to feel positive or negative emotion), or through more direct inquiry of emotion experience.

## 4. SOCIAL FUNCTIONS OF EMOTION

Darwin and James laid the foundation for the study of emotion in significant and enduring ways. In *The Expression of the Emotions in Man and Animals*, Darwin contended that emotions must be serviceable. That is, emotions should be functional adaptations that prepare

the organism to act and communicate a readiness to engage in a particular behavioral action. This preparation should be expressed through nonverbal displays that should be universal across human cultures. Darwin also argued that particular expressions of emotion should be directly connected to the behavior they prepare the individual to enact. Disgust, for example, is a serviceable habit that prepares an individual to expel harmful contents before they are ingested and whose display includes protrusion of the tongue. This thinking laid the foundation for the study of facial expression and functional accounts of emotion that are so widespread today.

In his 1884 essay titled “What Is an Emotion?,” James extended some of Darwin’s claims in proposing that each emotion should be defined by a specific physiological response. This radical notion fell out of favor but is now supported by studies indicating that amusement, anger, disgust, embarrassment, fear, love, surprise, and sadness all are associated with distinct physiological responses in the autonomic nervous system. More controversially, James also claimed that emotional experience followed rather than preceded emotion-relevant behavior. For example, crying would precede feelings of sadness. During the mid-20th century, this counterintuitive idea generated a great deal of interest among psychologists who prioritized cognitive and information explanations over motivational and emotional explanations of human behavior.

What distinguishes contemporary social functional approaches to emotion? First, a social functions perspective builds on Darwinian theorizing about emotion to consider how the challenges and opportunities of the human evolutionary environment led to the development of specific emotions. Importantly, this perspective is complemented by a consideration of how human cultural practices elaborate on human biology to influence emotion experience and meaning. Second, a social functions approach considers the effects of specific emotion processes at the intrapersonal level of individual experience and at the interpersonal level of ongoing social interaction among individuals. Thus, there is consensus that the experience of emotion guides information processing in functional ways, triggering specific memories and patterns of perception and categorization. The autonomic nervous system activity of emotion is thought to support specific kinds of action (e.g., fight or flight). And the communication of emotion conveys specific feelings, intentions, and dispositions, evoking specific and adaptive responses in observers. Finally, a

social functions approach is concerned with the consequences of emotion, emphasizing how emotions facilitate the successful navigation of the social environment.

In contrast to previous approaches that viewed emotions as disruptive forces, a social functional perspective regards emotions as evolved answers to the problems and opportunities presented by human social life across our evolutionary past. In this view, emotions are specific efficient responses tailored to address the problems of physical survival and social living. Emphasis is placed on the utility of emotion in the contexts in which it evolved, whether it remains useful in current environments or has become inappropriate or maladaptive. Framing emotion in this way generates a series of important and interrelated questions about emotion:

1. Why do humans have emotions?
2. How do emotions influence survival and adjustment?
3. What are the interrelated parts of the emotion system?
4. What are the beneficial consequences of emotion?

A social functions approach naturally directs attention to the role of emotion in navigating the complications of human social life. Diverse theorists converge on three classes of problems and opportunities that emotions were designed to solve. Several emotions, such as disgust and fear, help humans to meet the problems of survival, including avoiding predators and toxins. Still other emotions, including filial love, desire, compassion, and rage, help humans to meet the problems of reproduction that center on finding mates, maintaining monogamous bonds, and raising offspring to the age of viability. Finally, still other social emotions, such as gratitude, anger, embarrassment, and awe, help humans to meet the problems and opportunities related to group governance. More specifically, these emotions are thought to promote and maintain cooperative alliances and help humans to fold into social hierarchies and mete out justice and punishment. The article draws on this perspective to outline an understanding of seven emotions that will prove to be particularly relevant to health, work, and school settings.

#### 4.1. Anger

Anger is historically regarded as one of the most disruptive emotions. Functional accounts of anger have shed light on the beneficial aspects of this misunderstood emotion. Anger has been shown to be elicited by the perception that an injustice has

been intentionally committed against the self by another. This attribution of blame for an unfair outcome motivates an individual to (a) seek redressive action and (b) signal the intention to aggress if redressive action is not taken. To this end, anger is received as a threat by the target. According to emotion theorist Robert Frank, anger has two beneficial consequences. First, anger prepares an individual to aggress in response to serious transgressions, if need be, by directing blood to the face and arms. More important, signaling that one is genuinely likely to aggress is threatening and reduces the likelihood that a given individual will be cheated or taken advantage of by another. Signaling honest anger may be all the self-defense that is necessary to motivate a target to take redressive action or flee. Indeed, incidents of anger leading to actual aggression or violence are infrequent relative to the emotional experience.

#### 4.2. Anxiety

Unlike anger, anxiety is not always defined as an emotion. It has not been documented to have a distinct nonverbal display or to be accurately decoded. Instead, anxiety has been related to fear, either as a subset of the broader fear category or as a more generalized mood-like state. The experience of anxiety appears to be a concomitant of threat-related cognitions when an individual perceives a challenge in the environment that may exceed resources. Small amounts of anxiety have been shown to enhance performance on a challenging task, but profound levels of anxiety have been shown to impair the ability to perform.

#### 4.3. Embarrassment

The emotion of embarrassment is a clear example of individual experience diverging from larger social consequence. Embarrassment is marked by a brief display consisting of a downward gaze, a brief non-Duchenne smile, and the movement of the hand to the face. It is felt in response to unintended inappropriate behavior such as social gaffes, pratfalls, and uncontrolled bodily functions. Calling someone by the wrong name and tripping in public are typical triggers of embarrassment. Although embarrassment is an unpleasant individual experience, the audience present for an embarrassing incident is frequently amused or sympathetic toward the target. For this reason, embarrassment has been posited to serve an appeasement function. By feeling and displaying embarrassment, a person signals

to others that he or she realizes the inappropriateness of the behavior and is asking to be absolved of blame. At the level of the social interaction, the display of embarrassment restores social balance. Studies have shown that individuals who display embarrassment are better liked than their nonembarrassed peers and show greater concern for social norms. Adolescent boys who display embarrassment, for example, are less likely to engage in delinquent behavior than are their less embarrassment-prone peers.

#### 4.4. Fear

The emotion of fear prepares an organism to flee from threats that pose an immediate danger or are perceived to be likely to overwhelm the system. Displayed through raised eyebrows, opened eyes, and an open mouth, the biological basis of fear has been well documented. In contrast to anger, which sends blood to the arms, fear sends blood to the legs, facilitating a flight response. In studies of fear response, young monkeys exposed to either a snake or a flower for the first time expressed a fear of snakes immediately, suggesting that they are prepared to feel a fear of snakes. Interpersonally, humans have been shown to be sensitive to the fear cues of others, valuing the warning that fear displays can provide. When a facial expression of fear is presented among a large group of smiling faces, the fear face is identified more quickly than is a happy face among a large group of fearful faces. The work of LeDoux has connected fear to the amygdala, an almond-sized structure of the brain shown to process fear-related information. Amygdala damage is associated with deficits in perceiving and attending to obvious threats.

#### 4.5. Happiness

The emotion of happiness comprises the most general and recognized level of positive emotion. Happiness is universally recognized and signaled by the Duchenne smile, a combination that lifts the muscles of the mouth and crinkles the skin around the eyes. Unlike other types of smiling, Duchenne smiling reflects an individual's internal positive emotional state. The disposition to experience happiness has been associated with higher levels of life satisfaction and a cognitive style marked by optimism and diminished concern with social comparison. There is also some evidence that displays of happiness promote positive social relationships, perhaps by evoking positive social responses in others. A recent study showed that

women whose college yearbook pictures showed Duchenne smiling married earlier, had more positive relationships with partners and children, and reported higher levels of life satisfaction up to 30 years later.

#### 4.6. Love

Love is traditionally regarded as an enduring attitude of affection toward another or a style of relating to romantic partners, and recent research on love has found evidence of emotion-like properties. Love is signaled through a series of affiliation cues that generate movement toward the love object and has been theorized to promote commitment between love partners. Consistent with this theorizing, the self-reported experience and display of love correlate with constructive conflict resolution and perceived trust during moments of relationship threat. Feelings of love are also associated with a greater willingness to engage in prorelationship behavior, even at a cost to self-interest.

#### 4.7. Sadness

Felt in response to a perceived uncontrollable loss, sadness is marked by eyebrows pulled in toward the nose and a downturned mouth. Initially, sadness is marked by crying and other intense displays of grief, but prolonged sadness shuts down the body. This shutting down of the body is theorized to conserve the body's energy until a time when a person is better able to cope with the loss. Sadness is associated with cognitions of having little control over personal outcomes. For example, individuals who attribute positive life events to external causes and regard negative life events as having been brought on by internal characteristics of the self are prone to episodes of sadness.

The preceding seven emotion states do not cover the range of emotions currently studied by emotion researchers, nor do they encompass the states whose status as emotions generates the most agreement. Instead, these seven states comprise emotions whose relevance to the applied settings of health, work, and school has been most clearly documented.

### 5. EMOTIONS AND HEALTH

The literature on emotions and health is vast but can be summed up schematically with two fairly robust claims. First, positive emotions promote positive health and a

longer and more satisfying life. Second, negative emotions impair health and place individuals at risk for poor health outcomes ranging from depression to heart disease. These empirical summaries are based on studies of specific emotional states (e.g., fear) but more typically emerge out of studies of emotion-based dispositions such as anxiousness, hostility, and happiness.

The emphasis on the beneficial aspects of positive emotions notwithstanding, positive emotion routes to better health remain understudied relative to negative emotions. Rather, the disposition to positive emotion and the cognitive concomitants of positive emotional experience have been related to better health. For example, dispositional optimism, which is related to happiness, is associated with longer life in the face of terminal illness. The positive illusions of individuals who overestimate the likelihood of surviving cancer or a stroke appear to act as protective resources, prolonging life and helping individuals to find positive meaning despite illness.

The relationship of positive illusions to health has been the subject of lively debate. In one study, optimism was measured in men recovering from coronary artery bypass surgery. A man's dispositional optimism was related to the release of liver enzymes that promoted recovery, and these men were quicker to recover the ability to sit, walk, and return to work than were their less optimistic counterparts. Other studies show that dispositional optimists are more likely to engage in health-protective behaviors such as taking vitamins, eating a well-balanced diet, and exercising. Still, researchers have contended that happiness and positive illusions should negatively affect health, perhaps by promoting denial or underestimation of the likelihood that illness will happen. In support of this view, unrealistically high levels of optimism have sometimes been related to a diminished likelihood of engaging in health-protective behaviors. Nonetheless, positive illusions remain one avenue through which positive emotional states have been shown to lead to improved health outcomes.

Turning to the negative emotional dispositions, it has become increasingly clear that anger, sadness, and (more recently) shame can have deleterious effects on health. Although previous research connecting health outcomes to the disposition to experience anger or sadness has oversimplified the relationship, the connection exists. Prolonged episodes of negative affective states, rather than a brief episode of negative emotion, are chiefly responsible for the relationship between negative emotion and health. Small doses of

negative emotion, such as anxiety and guilt, are associated with engaging in preventive health behaviors and being willing to seek help. However, more enduring chronic affective states, such as depressed mood and chronic anxiety, have been linked to impaired immune functioning, coronary heart disease, memory decrements, and other major health conditions.

Similarly, the propensity to experience frequent and intense episodes of anger is commonly related to poor health outcomes. The association of "Type A" behavior with heart disease is well known in scientific and popular circles. More recent research has shown that one component of the Type A profile, the propensity to experience hostility and anger, is the principal mediator of this association. Studies of Type A individuals who are less prone to hostility show that the goal-driven persistence that marks Type A may actually promote compliance of rehabilitative regimens, whereas the hostile component of Type A predicts future coronary episodes.

Prolonged episodes of sadness are also predictive of heart disease, but the emotional experience of sadness and its related cognitions have been most directly linked to depression. An attributional style that blames the self for negative life events and perceives little ability to control or create positive outcomes has been associated with the development of depression. More directly, the tendency to ruminate on negative events by persistently directing thoughts toward negative events and possible negative outcomes leads to a greater likelihood of suffering depression. In one study of coping in the aftermath of a strong earthquake, individuals who reported ruminating about their quake experience during the weeks afterward also reported more depressive symptoms and were more likely to seek health services. This coping style is more characteristic of women, and indeed, women make up a disproportionate number of depression sufferers.

Depression is associated with both increased negative affect (e.g., sadness, anxiety, anger) and decreased positive affect. Administration of selective serotonin reuptake inhibitors (SSRIs) tends to decrease experience of negative affect and increase experience of positive affect in clinically depressed populations. These drugs produce parallel effects in normal populations. Nondepressed individuals administered SSRIs for 4 weeks were rated as less hostile by clinicians, self-reported less negative affect, and showed more affiliative behavior in a dyadic social interaction than did individuals administered a placebo.

The experience of shame and self-blame may uniquely relate to negative immunological changes. Relative to a control condition, individuals induced to feel shame by writing about a past traumatic experience for which they blamed themselves showed increases in proinflammatory cytokines, negative immunological response, and the stress hormone cortisol. Self-reports of shame, but not of guilt or other negative affect, specifically correlated with these physiological changes. A study of HIV-positive gay men found that shame-related cognitions, such as poor self-worth and excessive concern with social rejection, are associated with lower CD4 count (a negative immunological marker) and higher AIDS mortality.

The literature connecting the disposition to feel negative emotion to poor health is extensive, but emerging advances in this area suggest that parts of this story have yet to unfold. Guided by social functional analyses emphasizing the beneficial aspects of emotion, researchers have begun to examine how negative emotion experience might affect health in positive ways. One proposed path through which negative emotions may positively affect health is by promoting "toughness" or resistance to the harmful effects of stress. For example, mice exposed to repeated stressors at an early age showed faster patterns of both physiological arousal and recovery, suggesting an increase in the effectiveness of the stress response over time.

In a related vein, the disclosure and free expression of negative emotions have also been shown to promote health. According to Pennebaker, one source of ill health is a repressive coping style that is characterized by an effortful inhibition of emotional thoughts, feelings, and behaviors. In support of this view, repressive coping has been associated with the release of cortisol. Focusing primarily on trauma sufferers who have not shared their experiences with others, Pennebaker found that sharing traumatic experiences by either talking or writing in journals leads to reduced physiological arousal and fewer visits for health care during the ensuing months. Similarly, other studies have shown that confiding or disclosing traumas, even secondhand traumas, is associated with better immune system functioning.

Consistent with a social functional account of emotion, negative emotion is not harmful per se and can be an adaptive response to a situation. Grief, for example, is not associated with the negative immunological changes that characterize depression. Grief may be an adaptive response to loss, whereas depression is maladaptive. Studies of HIV-positive individuals who have

lost partners or close friends to AIDS find that depressed mood is associated with negative immunological changes and more rapid HIV progression, whereas high levels of grief without depressed mood are not.

## 6. EMOTIONS AND WORK

A great deal of modern-day social interaction takes place in work environments. Consequently, a consideration of emotion in the workplace is particularly benefited by an interpersonal perspective. In the workplace, individuals must initiate relationships, strive for distinction and status, seek cooperation, seal commitments, and optimally, be personally satisfied with their work lives. These endeavors are facilitated by the communicative aspects of emotion, where people observe and respond to each other's emotions in ways that often help to resolve or change the relational problems of everyday life.

How do emotions influence work satisfaction and productivity? As with health, positive emotion is also associated with beneficial outcomes. The expression of positive emotion has been shown to positively relate to both work satisfaction and achievement. In one study of 272 employees, the expression of positive emotion was associated with more favorable supervisor evaluations, higher pay, and more support from supervisors and coworkers 18 months later. A related issue, and perhaps a conceptually dominant one, concerns the contextual factors that support the experience and display of positive emotion. Research suggests that much of the variance associated with job satisfaction is attributable to a perceived sense of efficacy in carrying out a task. Perceived efficacy is related to lower levels of chronic stress and higher levels of subjective well-being. Individuals who are given tasks that challenge, but do not exceed, their ability may achieve a state of "flow" in their work, where interest is high and the tasks are engaging. In contrast, perceiving the challenges of the work environment to be overwhelming can lead to unhappiness and high levels of job stress.

Serious negative occupational outcomes are also closely tied to interpersonal emotion. Job burnout is a response to interpersonal stressors on the job, where an overload of contact with people results in changes in attitudes and behaviors toward them. Field research has established that burnout involves three dimensions, with each dimension highlighting different emotions. Anger and anxiety are expected to be high in the dimension defined as emotional exhaustion, that

is, feelings of being emotionally overextended and drained by one's contact with other people. Reduced experience of empathy and compassion is implicated in the depersonalization dimension, defined as an unfeeling and callous response toward people on the job. Decreased experience of pride appears to be central to the third dimension, where feeling low competence and low successful achievement in one's work leads to a reduced sense of personal accomplishment.

Another area of study has examined required emotional displays such as expressing positive emotion toward customers by workers in the retail industry or negative emotion toward debtors by workers at collection agencies. Termed "emotional labor," the use of positive emotional expression has been shown to elicit liking and a desire on the part of the customer to reciprocate in kind. On the other end of the spectrum, debt collectors are pressed to convey irritation to debtors and adjust their expressed emotions in response to debtor demeanor. This emotionally difficult work can clash with the collectors' genuine feelings toward debtors. This use of emotion may be task-effective, but emotional labor has been associated with decreases in felt positive emotion for expressing workers, and some organizations openly encourage employees to develop cognitive reappraisal skills to better cope with the emotional demands of the job.

How do emotions help individuals to navigate the social interactions of the workplace? How do they help them to initiate relationships, manage hierarchy, and procure the cooperation of coworkers? Anger and gratitude play important complementary roles in promoting cooperative alliances and creating stable long-term relationships. The initiation of cooperative relationships is assisted by liking and gratitude, which motivate the initiation of bonds and reward cooperation, respectively. Expressed anger can punish actions that threaten to undermine cooperation such as inappropriately taking credit and overusing resources. Threats that induce fear rather than anger have been found to be successful in changing the behavior of others. However, threats risk inducing anger rather than fear, and eliciting a strong angry response might pose a worthwhile risk only when interest and positive emotional interaction have proven to be ineffective.

The formation and maintenance of cooperative alliances take place within an existing hierarchical structure, and successful management of hierarchy is important to workplace success. Research suggests that the emotions of high-status individuals are of

primary relevance for two reasons. First, the outcomes of subordinates are strongly tied to individuals at higher ends of the hierarchy in most work environments. Second, the emotional experience of high-power individuals has been shown to differ from, and exert an effect on, the emotional state of subordinates. High-power individuals are more disinhibited in their emotional expression, expressing both negative and positive emotions and being less attune to how their emotional expressions are perceived by others. Status is conferred and reinforced by displays of anger rather than sadness, in part because the display of anger is associated with greater personal competence. Over time, the emotions of lower status people have been shown to converge to be more similar to the emotional experience of high-status people. This convergence is proposed to help coordinate the thoughts and behaviors of relationship partners, increase mutual understanding, and foster social cohesion. In the workplace, this emotional cohesion is likely to foster stable, mutually satisfying hierarchies.

In a workplace, decisions must be made and courses of action must be agreed on by relevant parties. How do emotions influence workplace decision making? One means is through emotion-related appraisals. Happy people view the world optimistically, perceiving situations to pose low risks and expecting positive outcomes. Fearful people perceive the world as threatening and are vigilant about risks and negative outcomes. These emotion-specific appraisal tendencies produce emotion-congruent effects on perception, reasoning, and decision making. People who feel happy readily notice and remember positive information, including other people's happy expressions. Bolstered by the accessibility and salience of positive information, happy people tend to view risks optimistically and make more risk-seeking choices. They label people using more positive categorizations (e.g., "good worker") and are more reliant on mental short-cuts in their decision making, and this can lead to less thorough information processing. In contrast, sad people tend to notice and remember negative and sad information, memories, and facial expressions. They tend to process information thoroughly when making judgments, presumably because doing so might help to explain or resolve their underlying negative mood. Appraising the world as threatening, fearful people assess risks pessimistically and make more risk-averse choices than do happy people. However, not all negative emotions lead to more careful decision making or judicious risk assessment. Anger is associated with

certainty and feeling control over a situation, and angry people behave more like happy people, assessing risks optimistically and making risk-seeking choices. Perceiving the world as under their own and others' control, angry people are more likely to blame individuals than to blame situational factors for negative events.

## 7. EMOTIONS IN SCHOOL

A large literature has addressed the factors that undermine academic achievement and social adjustment in school. The subset of the literature that relates to emotion has chiefly examined how anxiety leads individuals to underperform in, avoid, and mistrust particular academic domains or school in general. At a more interpersonal level, research on peer relations suggests important roles for anger, fear, sadness, and embarrassment in social adjustment.

Anxiety is considered to be the major cause of student underperformance on examinations. Students who consistently experience severe levels of anxiety are proposed to have a trait tendency known as test anxiety that involves increased anxiety and stress when specifically taking tests. Test anxiety has been associated with reduced motivation, suppressed immune functioning, and impaired test performance. Several mediators have been proposed, drawing on appraisal tendencies associated with anxiety and fear. Feeling afraid heightens a person's vigilance for threats in the environment. Studies find that test anxiety can impair performance by increasing a student's susceptibility to threatening external distractors. Other studies find that internally generated anxious thoughts disrupt performance by reducing the available cognitive resources in taking the test.

Anxiety is also considered to be a major cause of stereotype threat, a specific type of academic underperformance. Stereotype threat occurs when a negative stereotype about a group of which a person is a member becomes personally relevant, usually in the context of an experience the person is having. Stereotype threat is the resulting sense that the person can be judged or treated in terms of the stereotype or that the person might do something that would inadvertently confirm the stereotype. At each of level of academic skill, students sharing certain social identities get lower scores than do other students. Female students are outperformed by male students in math at each level of skill. Relative to White students, African American, Native American, and Hispanic American students

underperform on tests measuring overall academic ability. The anxiety-related processes believed to underlie stereotype threat have been proposed to be a primary cause of these groups' impaired performance. Self-reported anxiety has sometimes been found to mediate these groups' impaired performance, although in some cases only for those individuals who are sensitive to the group stereotype. One study found that African Americans did not report more anxiety than did Whites during an intellectual ability test; however, African Americans' blood pressure increased significantly from baseline relative to that of Whites.

How can the anxiety of stereotype threat be mitigated? Over a decade of research has found that the performance of students vulnerable to stereotype threat is improved when stereotype threat is removed from the test-taking situation, for example, when students are told that a particular test produces no group differences in performance. Other means of reducing stereotype threat include creating an academic setting that communicates identity safety, for example, when a teacher tells minority students that he or she will hold them to high academic standards and believes the students can meet them. The importance of combating stereotype threat is evident in the link with negative outcomes in academic domains that are negatively stereotyped for an individual's group identity. Students experiencing stereotype threat tend to avoid the domain where the stereotype applies, self-handicap in that domain, and/or situationally disengage their self-esteem and ability appraisals from performance on tests in that domain. The chronic experience of stereotype threat risks instilling devaluation of success in that domain, a disengagement of identity from that domain, and a distancing from careers relevant to that domain.

Other emotions play critical roles in another dimension of school, that is, psychological adjustment and the negotiation of peer relationships. Anger is an important component of conduct disorder and externalizing problems in children. Children's self-reported tendency to experience and express anger correlates with their "acting out" behaviors in normal and clinically diagnosed populations. Children who engage in antisocial behavior are rated by parents, teachers, and peers as being more hostile and angry. Anger-prone children are found to expect hostility from their peers and respond to ambiguous situations with hostile behavior, which elicits rejection from other children. In contrast, children experiencing internalizing disorders, such as depression and anxiety, tend to feel fear and sadness. For example, measures of depression in fifth

graders are associated with self-reports of sadness as well as anger in the case of boys.

The social-moral emotions, those that encourage moral behavior, socialization, and adaptive assimilation into society (e.g., embarrassment, shame, guilt), appear to motivate children to avoid social transgressions. Less guilt, for example, is positively associated with externalizing problems such as antisocial behavior and is negatively associated with moral development. One study of the psychological adjustment of adolescent boys tested for the roles of anger, fear, and embarrassment by coding facial expressions during a structured social interaction (an intelligence test). Boys categorized by teachers as having externalizing problems displayed increased expressions of anger, whereas boys categorized as having internalizing problems displayed increased fear. Interestingly, both groups of boys displayed less embarrassment during the task than did nondisordered boys, indicating that embarrassment signals normal social-moral development.

## 8. CONCLUSION

The three domains of health, work, and school are important contexts where life outcomes are shaped. Emotional experience and disposition shape individual experience and social interaction in ways that influence these life outcomes.

### See Also the Following Articles

Agreeableness ■ Emotion and Culture ■ Intelligence, Emotional ■ Interpersonal Attraction ■ Interpersonal Perception

### Further Reading

- Darwin, C. (1998). *The expression of the emotions in man and animals* (P. Ekman, Ed.). Oxford, UK: Oxford University Press. (Originally published 1872)
- Ekman, P., & Davidson, R. J. (Eds.). (1994). *The nature of emotion: Fundamental questions*. New York: Oxford University Press.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.
- Greenberg, M. A., Wortman, C. B., & Stone, A. A. (1996). Emotional expression and physical health: Revisiting traumatic memories or fostering self-regulation? *Journal of Personality and Social Psychology*, 71, 588–602.
- James, W. (1884). What is an emotion? *Mind*, 9, 188–205.

- Keltner, D., & Haidt, J. (2001). Social functions of emotions. In T. J. Mayne, & G. A. Bonanno (Eds.), *Emotions: Current issues and future directions* (pp. 192–213). New York: Guilford.
- Mayne, T. J. (1999). Negative affect and health: The importance of being earnest. *Cognition & Emotion*, *13*, 601–635.
- Salovey, P., Rothman, A. J., & Rodin, J. (1998). Health behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 633–683). New York: McGraw-Hill.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.



# Emotional and Behavioral Problems, Students with

Thomas J. Huberty

Indiana University, Bloomington, Indiana, USA

1. Introduction
2. Conceptualizations of Child and Adolescent Psychopathology
3. Mental Disorders in Children and Adolescents
4. Interventions  
Further Reading

## GLOSSARY

*comorbidity* The co-occurrence of two or more mental disorders in the same person.

*externalizing behaviors* Behaviors that are undercontrolled (e.g., hyperactive, oppositional, delinquent, aggressive, disruptive).

*internalizing behaviors* Behaviors that are overcontrolled (e.g., withdrawal, anxiety, somatic complaints, obsessions).

*symptoms* Behaviors that are inappropriate or excessive and are of clinical significance.

*syndrome* A cluster of symptoms that occur together in a consistent and recognizable pattern.

This article presents characteristics of students with emotional and behavioral problems within the conceptualization of child and adolescent psychopathology. Implications for children in clinical and school settings are included.

## 1. INTRODUCTION

Students who have significant emotional and behavioral problems necessitating professional intervention often are described as being “emotionally disturbed,” “behaviorally disordered,” “emotionally and behaviorally disordered,” or some similar term. Although these labels are in common use, they are broad and ambiguous and offer little help in understanding the nature, extent, prevalence, and implications of these problems in children and youth. This article describes current conceptualizations of children’s emotional and behavioral problems and implications for schooling and effective interventions.

## 2. CONCEPTUALIZATIONS OF CHILD AND ADOLESCENT PSYCHOPATHOLOGY

The development of research and knowledge in child and adolescent psychopathology has burgeoned during the past two decades or so. The current state of the research suggests that children’s emotional and behavioral problems can be conceptualized as internalizing or externalizing. Internalizing problems are characterized by anxiety, depressed mood, somatic complaints, withdrawal, and similar behaviors. These behaviors

often are referred to as “overcontrolled” because the children show inhibited behaviors and spend much energy in controlling behaviors. These problems also are similar to what are commonly termed “emotional problems.” Externalizing problems refer to behaviors such as impulsiveness, aggression, inattention, hyperactivity, and distractibility. These behaviors are termed “undercontrolled” because the children lack the ability to control the behaviors. They are often referred to as “behavior problems.” Internalizing and externalizing behaviors can be conceptualized as being on a continuum where the range of behaviors shown may be from very withdrawn (internalizing) to highly active and perhaps even aggressive (externalizing). However, this range does not represent a true continuum because children may show many behaviors simultaneously, although some may predominate. For example, 25 to 40% of children with conduct disorder (CD) also are depressed, although the conduct problems may be observed more readily. Research has shown that the correlation between internalizing and externalizing patterns ranges from approximately .25 to .40. Moreover, some behaviors are associated with both internalizing and externalizing problems, for example, social skills deficits and academic difficulties.

When children demonstrate internalizing or externalizing behaviors of such seriousness that personal, social, and/or academic functioning is affected, they may be referred to mental health professionals. Children with externalizing behaviors are more likely to be referred for mental health care because they are more disruptive at home and in the classroom. Children with hyperactivity, aggression, and other externalizing behaviors account for approximately 50% of all referrals to mental health professionals. Inattention without hyperactive behaviors is seen most readily in the classroom. Children with internalizing problems are less likely to be referred because they generally are not disruptive and the problems are less likely to be detected easily by parents or teachers. However, exceptions to this trend do occur, for example, severe phobia or depression that leads to suicidal ideation or behavior.

Another important conceptualization of behavior is trait and state, which complements the internalizing–externalizing perspective. “Trait” refers to behaviors that are characteristic of children and are seen across a variety of settings and circumstances. “State” refers to behaviors that are a manifestation of the impact of the environment. The state–trait concept is important for helping to understand why children’s behaviors vary across settings. In schools, for example, it is common for children to exhibit

off-task behaviors in one classroom but be more attentive in another classroom. Thus, the difference in behaviors might not be due to a trait; rather, it may reflect factors such as teacher management skills (state). Therefore, interventions may focus on the teacher and the setting rather than on the children. Children who have a high degree of a trait, however, may be more likely to show the behaviors in specific situations. Spielberger has conducted extensive research on anxiety and anger as having state and trait dimensions. Anxiety and anger tend to be constant traits, placing children at an increased likelihood of becoming anxious or angry in specific situations (state). Interventions often are directed toward reducing the magnitude of the trait as well as helping children to cope with anxiety and anger-producing situations.

## 2.1. The Taxonomy of Mental Disorders

A taxonomy is an orderly classification of objects, people, or other phenomena according to their presumed natural relations with each other. A taxonomy classifies according to mutually exclusive criteria; that is, the criteria can apply to only one specific member of a class, making the class unique. The concept is most often used in the physical sciences, such as botany and zoology, but it is also applied to human behavior. The taxonomy of mental disorders most commonly used in the United States is the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). There are more than 300 possible diagnoses of mental disorders in the DSM-IV, and determining which one is the most accurate description of the behaviors presented to the clinician can be challenging. The primary reason for the difficulty in formulating clear diagnoses is the problem of comorbidity; that is, many behaviors shown by children are seen in more than one disorder. When more than one disorder exists within the same individual (i.e., not mutually exclusive), they are considered to be comorbid disorders. Comorbidity is the rule rather than the exception in children and youth, occurring in more than half of all cases referred to mental health professionals.

## 2.2. Behaviors, Symptoms, Syndromes, and Disorders

Children’s emotional and behavioral problems typically are first noticed by parents or teachers, who may refer these children to a physician, a psychologist, or other mental health professional. Symptoms are behaviors of

clinical significance that interfere with functioning. When behaviors occur together to form distinct patterns, they are termed “syndromes.” When syndromes occur with such intensity, frequency, or duration that personal, social, or academic performance is affected negatively, the children may be given a diagnosis of a mental disorder by a psychologist or physician.

As an example, consider the case of “Eric,” a 6-year-old boy in the first grade. As a preschooler, Eric exhibited some behaviors of concern such as high activity level, impulsiveness, and noncompliance with parents’ requests in and out of the home. However, the parents were able to manage these behaviors successfully, and they did not view them as problematic. In school, Eric was required to participate in groups, take turns, and follow directions. Instead of these prosocial behaviors, a pattern of noncompliance, difficulties in participating and cooperating, blurting out answers, talking without permission, and interfering with other children developed. These behaviors were considered symptoms of a problem. When the symptoms occurred together in an identifiable pattern, they comprised a syndrome that was of clinical and educational significance because it interfered with the performance of Eric and others. He was evaluated by a psychologist, who considered the behaviors/syndrome in light of current diagnostic criteria and assigned a diagnosis of attention deficit/hyperactivity disorder (ADHD)–combined type. In this example, individual behaviors became symptoms that occurred together to become a pattern (syndrome). Because the syndrome was determined to be at a clinically significant level of impairment, a formal diagnosis of a mental disorder of ADHD was given.

### 3. MENTAL DISORDERS IN CHILDREN AND ADOLESCENTS

The majority of mental disorders listed in the DSM-IV can be seen in children and adults. The concepts of internalizing and externalizing patterns are applicable because most of the disorders can be described in those terms. There are some disorders, however, that are considered to be applicable specifically to children and adolescents, for example, separation anxiety disorder (SAD) and CD.

#### 3.1. Internalizing Disorders

The most common internalizing disorders in children and adolescents are the same types that are seen in

adults, although there may be some differences in symptoms. Internalizing disorders include conditions such as major depressive disorder, dysthymia, and bipolar disorder. Anxiety disorders occur in approximately 5 to 10% of children, and the prevalence of depression is approximately 3 to 5%. Table I summarizes characteristics of anxiety problems.

Recent research indicates that bipolar disorder may be more common in children and adolescents than was previously believed. Bipolar disorder was formerly known as manic–depressive disorder and shows a range of mood symptoms from high-energy mania to severe incapacitating depression. However, it is less common than anxiety or depression. Another internalizing anxiety disorder is posttraumatic stress disorder (PTSD), which was initially identified in soldiers who experienced extreme combat exposure, which was followed by “flashbacks” of the events, difficulty in sleeping, and generalized anxiety. In children, PTSD symptoms can be seen as a result of experiences such as abuse, divorce, and other traumatic events, although flashbacks are less common in children. The only anxiety disorder unique to childhood is SAD, which is an unwillingness to separate from parents at a level inappropriate for the age or developmental level of the

TABLE I  
Characteristics of Anxiety in Children

<i>Cognitive</i>	<i>Behavioral</i>	<i>Physiological</i>
Difficulties with concentration	Shyness Withdrawal	Trembling or shaking
Memory problems	Frequent asking of questions	Increased heart rate
Overreaction and “catastrophizing” relatively minor events	Frequent need for reassurance Need for sameness	Excessive perspiration Shortness of breath
Worry	Avoidant	Dizziness
Irritability	Rapid speech	Chest pain or discomfort
Perfectionism	Excessive talking	Flushing of the skin
Thinking rigidity	Restlessness, “fidgety”	Nausea, vomiting, diarrhea
Hypervigilant	Habit behaviors such as hair pulling and twirling	Muscle tension
Fear of losing control	Impulsiveness	Sleep problems
Fear of failure		
Difficulties with problem solving and academic performance		

children. It is common for children to show some anxiety about leaving their parents, even during the early days of school. When the behaviors exceed what is expected developmentally and the children cannot separate in a manner typical of most children, SAD may be an appropriate diagnosis. The presence of SAD may indicate parent-child or family issues; for example, a parent may have difficulty in allowing a child to have increased independence and may encourage and reinforce the child for refusing to separate. A similar pattern may exist in cases where a child exhibits "school refusal" (sometimes referred to as "school phobia"). If there is no rational basis for a child to not want to go to school (e.g., being the victim of bullying), it may suggest parent-child or family problems.

Depression occurs in less than 1% of preschool children and increases during the school years from approximately 0.3 to 2.5% in prepubertal children. By adolescence, however, the incidence of depression increases to approximately 4.3%, with lifetime prevalence rates of major depressive disorder in adolescents ranging from approximately 8.3 to 18.5%. Clearly, depression is a major concern with children and youth and is associated with suicidal ideation and attempts. Depression is a leading cause of death in adolescents after accidents. Some of the major characteristics of depression are presented in Table II.

### 3.2. Social Correlates of Internalizing Patterns

Children who have internalizing patterns, such as anxiety and depression, may tend to withdraw from social interactions as a way in which to cope with chronic feelings of distress. Withdrawal, social ineptness, and avoidance of interactions due to low self-esteem may be evident. These children may be perceived as "lazy" and unmotivated, despite the fact that they likely want to have social relations but are uncomfortable in initiating interactions due to perceived incompetence and lack of social skills. There is some evidence that anxious and withdrawn children are at higher risk for being the victims of bullies at school and may demonstrate reactive behaviors such as school refusal, sleeping problems, inconsistent performance, lack of assertiveness, poor peer relationships, low self-esteem, and low self-efficacy.

### 3.3. Academic Correlates of Internalizing Behaviors

Children who show anxiety and depression tend to have academic problems that can lead to more distress, causing a cycle of problems that persist over time. These children tend to avoid academically challenging tasks, withdraw from classroom activities, have poor

TABLE II  
Characteristics of Depression in Children

<i>Cognitive</i>	<i>Behavioral</i>	<i>Physiological</i>
Arbitrary inferences	Withdrawal	Hyposomnia or hypersomnia
Selective abstraction	Lack of assertiveness	Increased or lowered appetite
Overgeneralization	Impaired social relations	Weight loss or weight gain as a function of change in eating habits
Catastrophizing	Lack of energy	Fatigue
"All-or-none" thinking	Lowered performance at school	Cries easily
Magnification and minimization	Loss of interest in hobbies, activities, and friends	
Personalization of too many or inappropriate events	Low self-esteem	
Absolutist thinking	Seclusiveness	
Dichotomized thinking, leading to inefficiency in generating alternatives and problem solving	Low participation	
Attributional styles: internal, stable, and global or external, unstable, and specific	Irritability	
	Reduced concern about personal appearance	
	Negative statements about self and others	

achievement, lack persistence on difficult tasks, and show signs of “learned helplessness.” They also tend to worry about appearing academically inferior to others, worry about incompetence, do not use effective learning strategies, and blame themselves for academic failure.

### 3.4. Externalizing Disorders

The most common externalizing disorders in children and adolescents are ADHD, CD, and oppositional defiant disorder (ODD). Collectively, these disorders often are referred to as disruptive behavior disorders (DBDs) because they have the common characteristic of showing disruption in home, school, and other settings.

ADHD is the most common DBD and accounts for a large percentage of referrals to mental health professionals. It is shown in approximately 3 to 5% of children and is three to six times more common in boys than in girls. It is a developmental disorder that appears to have some neurological substrates present at birth. With adequate assessment procedures, approximately 50% of cases can be identified by 4 years of age, although many children, especially those who tend to be inattentive, might not be identified until they enter school. There are three subtypes: predominantly inattentive type, predominantly hyperactive-impulsive type, and combined type. The inattentive type tends to be more common in girls than in boys and is more likely to be noticed at home and in situations where high degrees of attention are needed such as in studying, reading, and following directions. The hyperactive-impulsive type is characterized by displaying impulsive behaviors, engaging in off-task behaviors at school compared with peers, acting “fidgety,” making noises, bothering others, and engaging in similar types of behavior. A list of DSM-IV criteria for ADHD subtypes is presented in Table III.

ODD is characterized by a pattern of oppositional behaviors such as noncompliance with requests or directives of parents, teachers, and/or other adults. Most of these children are resistant and defiant to authority and may engage in disruptive behavior but do not demonstrate significant antisocial behaviors. However, boys and girls with ODD are at high risk for developing more serious problems, primarily CD. CD is a well-established pattern of oppositional and defiant behaviors and is accompanied by antisocial behaviors such as stealing, fighting, truancy, and bullying. There are two forms of the disorder: (a) early-onset CD, which occurs before 7 years of age, and (b) late-onset CD, which is first seen during the preteen or early teen years. As might be expected, children with early-onset CD tend to have

TABLE III  
Symptoms of ADHD

<i>Inattention</i>	<i>Hyperactivity</i>	<i>Impulsivity</i>
Fails to attend to details	“Fidgety” Leaves seat without permission	Blurts out answers
Difficulty in sustaining attention	Runs about or climbs on objects inappropriately	Difficulty in taking turns
Does not listen to directions	Difficulty in playing quietly	Interrupts others
Avoids tasks requiring sustained effort	On the “go,” seems driven like a motor	
Often loses things	Talks incessantly	
Distractible		
Forgets things and events		

more serious problems, are more likely to get into trouble with adults who attempt to exercise authority over them, or come into conflict with legal authorities. If the behaviors in early-onset CD are not corrected by late childhood or early adolescence, the prognosis for positive outcomes is guarded at best. Many children with early-onset CD are at high risk for developing antisocial personality disorder as adults, many of whom come into repeated conflict with society, including a high divorce rate, legal difficulties, alcohol or drug abuse, and employment problems. Children with early-onset CD are at a higher risk for school failure and for dropping out of school before graduation. Adolescents with late-onset CD are less likely to engage in significant antisocial behaviors. Instead, they are more prone to behavior problems such as poor academic performance, sexual risk taking, smoking, and noncompliance. Many children who are given a diagnosis of CD have met criteria for ODD at some time in their lives but might not have been given that diagnosis. Although most children with CD have met criteria for ODD, the reverse is less common because many children who present with ODD remain oppositional and do not develop the more significant antisocial behaviors associated with CD. ODD appears to set the pattern for CD by evoking coercive, harsh, and inconsistent parenting behaviors that lead to repeated conflict and management difficulties.

ADHD shows comorbidity with ODD and CD, but the relations are varied. Boys with ADHD are at higher risk for developing CD than are boys without ADHD; this

appears to be the result of high comorbidity of ADHD and ODD. In boys, ODD appears to be a precursor to CD, but ADHD is not a precursor to CD. Much less is known about these associations with girls; the associations may or may not be similar to those with boys. ADHD begins during the preschool to early school years and, if accompanied by early-onset CD, leads to more significant behavioral problems through the elementary school years.

### **3.5. Social Correlates of Disruptive Behavior Disorders**

Children with DBDs are at higher risk for developing poor relationships with typical peers and adults, including parents and teachers. Their behaviors tend to be annoying and unacceptable to others, leading to punitive reactions by adults and social rejection and ostracism by peers. If the pattern continues, these children are likely to be seen as “outsiders” to their peer groups and will seek out other children with similar patterns, leading to reinforcement of the behaviors and even greater social alienation.

### **3.6. Academic Correlates of Disruptive Behavior Disorders**

Children with externalizing behavior problems or disorders are more likely to do poorly in school, have academic skill deficits, exhibit frequent off-task behaviors, have poor peer relations in the classroom, and be disruptive to the educational process. They also may show signs of frustration, inferiority, anger, and aggression. Angry and aggressive children appear to have a greater number of fears, concerns, feelings of being victimized by others’ actions and provocations, and beliefs that others have negative attitudes toward them. These feelings distract angry and aggressive children from focusing on academic tasks and acquiring necessary learning skills. Much as with the tendency toward social alienation, children with these behaviors may cause teachers and others to be less willing to help them, thereby creating more avenues for academic failure. Continued failure is likely to occur, creating a cycle of frustration and alienation. These children may, however, show “bravado” by exhibiting confidence in their academic skills while, at the same time, harboring resentment and anger at teachers whom they believe are not interested in helping them and at their peers who reject them. They may project the reasons for failure onto others and show boredom and lack of interest in school-related tasks.

## **4. INTERVENTIONS**

### **4.1. Programmatic Interventions**

Children who have emotional and behavioral problems might need help at school to assist them in achieving as much as possible. In the 2001 U.S. surgeon general’s report, it was reported that 70% of all mental health services provided to children are delivered in the school setting. In general, there are two avenues for providing interventions in the school setting. The first avenue is through general education such as might be provided by school psychologists, counselors, and social workers. These interventions are based on clinicians’ evaluations of children’s needs and may take the form of individual or group counseling, consultation, or parent involvement.

The other avenue for intervention is determining whether children are eligible for special education services. Public schools in all states are required to have special education programs and services for children whose behaviors interfere with academic performance. These programs and their associated regulations must be consistent with the federal Individuals with Disabilities Education Act (IDEA), which provides assurance that all children who have disabilities are properly identified, assessed, and given appropriate instruction and support. For children with emotional and behavioral problems, IDEA uses the term “emotional disturbance” (ED), which is defined as follows:

- (i) The term means a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree that adversely affects a child’s educational performance:
  - (a) An inability to learn that cannot be explained by intellectual, sensory, or health factors
  - (b) An inability to build or maintain satisfactory interpersonal relationships with peers and teachers
  - (c) Inappropriate types of behavior or feelings under normal circumstances
  - (d) A general pervasive mood of unhappiness or depression
  - (e) A tendency to develop physical symptoms or fears associated with personal or school problems
- (ii) The term includes schizophrenia. The term does not apply to children who are socially maladjusted unless it is determined that they have an emotional disturbance.

Although not all states are required to adopt this specific definition, many use the same or similar

terminology (e.g., “emotional disability,” “behavior disorder”). The IDEA definition is problematic, however, because many of the terms are ambiguous and are not operationalized (e.g., “inability to learn,” “satisfactory,” “inappropriate,” “normal circumstances”). The school psychologist is the professional most responsible for assessment of a child suspected of having ED, although the final determination of eligibility for special education services is done by a case conference committee that includes the parents, the school psychologist, a counselor, teachers, specialists, and others.

The child must have a “condition” that contributes to impairment in educational performance, which most often includes both academic and social development. It is important to note that having a formal DSM-IV diagnosis does not ensure that a child will receive special education services. It is not uncommon for parents to bring a diagnosis from a physician or psychologist to school, asking for special education services, only to be told that the child is not exhibiting problems that are interfering with educational performance. In those cases, the child might not be found eligible for services, despite having a formal psychiatric diagnosis. Conversely, a child could be found eligible for services for students with emotional disturbance and not have a formal psychiatric diagnosis.

A particularly enigmatic provision of the IDEA definition is the “social maladjustment” exclusion clause, which indicates that such children might not be found eligible for services. There is no definition of this term, and this can lead to difficulties for school-based professionals. The term, as originally considered, was to apply to children who exhibited delinquent behaviors that were controllable. Many professionals tend to equate “social maladjustment” with externalizing problems and equate “emotional disturbance” with internalizing conditions. This distinction would suggest that children with oppositional and conduct problems and ADHD would not be eligible for services unless there were accompanying internalizing problems that negatively affected performance. Conversely, because social maladjustment is not defined, one could argue that children with either internalizing or externalizing problems are socially maladjusted. Therefore, operationalizing the social maladjustment criterion for purposes of establishing eligibility is problematic. However, children with externalizing behaviors may qualify for special education services under IDEA or for general education accommodations under Section 504 of the Rehabilitation Act if needed to benefit from their education. They are not automatically excluded from receiving special education services.

If a student is found eligible for special education services for emotional or behavioral problems, an individualized education plan (IEP) is written in collaboration with the parents, school professionals, and the student if appropriate. In the IEP, goals and objectives are written to address the problems that are affecting educational performance. Services may include counseling, behavioral interventions, consultation with teachers, working simultaneously with the parents, and/or other appropriate interventions. The goal of special education services is not to provide comprehensive mental health services but rather to assist the student in performing at an appropriate level and in gaining educational benefit. The content of the IEP specifies what those services will be and how the outcomes will be measured. Counseling services may be provided if they are made a component of the IEP, or they may be included as an additional service under general education.

## 4.2. Individual Interventions

Individual interventions for students’ emotional and behavioral problems are numerous and cannot be discussed extensively here. In general, research shows that cognitive-behavioral, behavioral, and multisystemic interventions are among the most effective. A brief summary of the major types of interventions and their goals are presented in the following subsections.

### 4.2.1. Internalizing Patterns

The majority of internalizing problems are seen in anxiety and depression, which have several common characteristics. Cognitive-behavioral therapy techniques focus on the faulty cognitions and attributions and include components such as attribution retraining, self-talk strategies, self-control methods, self-monitoring, and self-reinforcement. For older students with anxiety, progressive muscle relaxation coupled with systematic desensitization techniques have been shown to be effective in reducing the physiological symptoms of anxiety and in lessening phobic behaviors. Children with depression also tend to be socially withdrawn and to derive relatively little enjoyment from their environment. In these cases, activity increase strategies and reinforcement for increased social interactions may be useful adjuncts to cognitively oriented therapies.

### 4.2.2. Externalizing Patterns

Children with externalizing patterns, particularly those with aggression, hyperactivity, and impulsiveness, are

best addressed with behavioral intervention strategies that emphasize providing consequences for behavior, including rewards and negative consequences. Behavioral contracts, reward systems, token economy methods, reinforcement, and punishments often are effective methods for increasing prosocial behavior while decreasing undesirable behaviors. Aggression replacement training is a method that includes both cognitive and behavioral techniques to address aggression problems. Including parents and teachers in behavior management training, behavioral contracts, and reinforcement programs can be effective in helping students to change behavior over settings. If negative consequences are used, they should be used sparingly and coupled with positive reinforcement. Although physical punishment (e.g., spanking) is used by some parents, research indicates that it has only temporary effectiveness, at best, in stopping a behavior and does little toward teaching prosocial behavior. Other strategies include positive behavioral supports, cognitive-behavioral strategies, and multisystemic therapy.

Children and youth with either internalizing or externalizing problems often have difficulty with social interactions, self-control, and self-regulation of their emotions and behaviors. Therefore, teaching social skills, self-control strategies, and anger management techniques using cognitive-behavioral and behavioral techniques often is necessary to help children function more effectively with peers and adults in the school setting.

#### 4.2.3. Medications

Prescribing medications to address emotional and behavioral problems in children and youth is controversial. The majority of psychotropic medications used with children were developed for adults and then adapted to children with dosage adjustments. In general, the research on short- and long-term effects of medications with children is sparse. The exception to this pattern is stimulant medications used to treat ADHD and associated symptoms. Historically, stimulants have been used to treat ADHD and been shown to be effective in approximately 70% of cases with few short- or long-term side effects. Some of the newer medications to treat ADHD are not stimulants and show promise of being effective as well as of reducing side effects such as sleeping problems. In the author's experience, parents tend to be reluctant to agree to use medication with their children, whereas school personnel are more likely to advocate its use, particularly for behaviors such as inattention, off-task, and hyperactivity. It is clear from research and clinical practice that medications, when

used appropriately, can be effective in helping to address emotional and behavioral problems. Medication is most often prescribed to treat ADHD in children. The research literature demonstrates that medication combined with behavioral, cognitive, and self-management interventions is more effective than medication alone for the treatment of ADHD. Many children with ADHD have concomitant social and academic problems that are not affected directly by medication, thereby requiring behavioral interventions. Medication should be neither advocated nor completely ruled out for all cases; rather, it should be considered with regard to the needs of the individual child. School personnel are in a particularly good position to observe and monitor the effects of medication and can be valuable collaborators with the prescribing physician.

### See Also the Following Articles

Attention Deficit/Hyperactivity Disorders (ADHD) ■ Attention Deficit Disorders: School-Based Interventions ■ Behavioral Assessment in Schools ■ Behavioral Observation in Schools ■ Diagnostic and Statistical Manual of Mental Disorders ■ Emotion ■ Learning Disabilities ■ Mental Measurement and Culture ■ School Violence Prevention ■ Traits

### Further Reading

- Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2002a). Is American student behavior getting worse? Teacher ratings over an 18-year period. *School Psychology Review, 31*, 428–442.
- Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2002b). Ten-year comparisons of problems and competencies for national samples of youth: Self, parent, and teacher reports. *Journal of Emotional and Behavioral Disorders, 10*, 194–203.
- Achenbach, T. M., & Edelbrock, C. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85*, 1275–1301.
- Breen, M. J., & Fiedler, C. R. (2003). *Behavioral approach to assessment of youth with emotional/behavioral disorders* (2nd ed.). Austin, TX: Pro-Ed Publishing.
- Cicchetti, D. (1998). The development of depression in children and adolescents. *American Psychologist, 53*, 221–241.
- March, J. (Ed.). (1995). *Anxiety disorders in children and adolescents*. New York: Guilford.
- Sameroff, A. J., Lewis, M., & Miller, S. M. (2002). *Handbook of developmental psychopathology* (2nd ed.). New York: Kluwer Academic/Plenum.
- Stoff, D., Breiling, J., & Maser, J. (Eds.). (1997). *Handbook of antisocial behavior*. New York: John Wiley.
- Wenar, C., & Kerig, P. (1999). *Developmental psychopathology* (4th ed.). New York: McGraw-Hill.



# Emotion and Culture

**Batja Mesquita**

Wake Forest University, Winston-Salem, North Carolina, USA

**Ashleigh Haire**

University of North Carolina, Greensboro, North Carolina, USA

1. Are There Universal Emotions?
  2. What Components of Emotions Are Universal?
  3. Are Emotions Culture Specific?
  4. Conclusion
- Further Reading

## GLOSSARY

**action readiness** An inclination to achieve a particular change in one's relationship with the social environment; action readiness is an essential part of emotional experience.

**antecedent event** Any event that evokes an emotional response from an individual.

**appraisal** An individual's evaluation of the meaning an event has for him or her; an appraisal of the situation is an essential part of emotional experience.

**cultural model** A culture's construction of reality that organizes relationships and illustrates what a person is and should be; cultural models refer to understandings as well as the ways in which life is arranged.

**dimensions of emotions** Specific parts that make up the overall emotional experience of an individual such as the way in which the person evaluates an event and the way in which he or she is inclined to respond to it.

**facial expressions** Patterns of facial movements that are associated with emotion.

**physiological response patterns** Specific patterns of physiological events, such as muscle activity and heart rate,

that are associated with the experience of a specific emotion.

**prevention focus** A psychological state in which an individual's primary goal is the pursuit of positive outcomes and rewards.

**promotion focus** A psychological state in which an individual's primary goal is the avoidance of negative outcomes and consequences.

**vocal expressions** Nonverbal aspects of vocal utterances that are associated with emotion.

Do people in different cultures experience similar emotions, or are the emotions of people in other cultures different from one's own? This is the primary question addressed in the research on culture and emotions. This question has important practical implications for a variety of issues. For instance, can one accurately recognize the emotional tone of parent-child relations in other cultures? Do Western and non-Western business partners understand each other's emotional expressions correctly when they are at the negotiating table? Is it possible to counsel people from other cultures? Can one interpret the aggressive threats or the populist emotional addresses of the leaders of antagonistic countries in other parts of the world? The question of the cross-cultural diversity of emotions is important in light of any social relations across cultures. To understand the answers to these questions, one must first understand what research shows about cross-cultural similarities and differences in emotional experience.

## 1. ARE THERE UNIVERSAL EMOTIONS?

Many people believe that, regardless of one's culture, there are certain things that we all experience as humans. Emotion is one such example. Quite a few studies in the domain of culture and emotion have aimed to make precisely this point—that some emotions are “basic” to all humans. Researchers who take this approach believe that, regardless of the culture in which one lives, certain emotions are experienced by everyone wherever they may live. In addition, these scientists suggest that each of these basic emotions has unique features. For example, many people believe that anger is a basic emotion and that the events that cause it, and the ways in which an individual responds to it, are fairly similar for everyone. An assumption underlying this viewpoint is that each of these basic emotions is characterized by a unique experience of the emotion itself and that this unique experience is not directly measurable by scientists. However, characteristic features of each of these emotions are measurable, and when they are observed, they identify the emotion being experienced by the individual. Researchers who seek to prove the existence of basic emotions have attempted to establish the presence or absence of these emotion characteristics (and hence the emotions themselves) in various cultures. Features of emotions include unique facial and vocal expressions, behaviors, and physiological arousal patterns. Thus, for these scientists, the question is a straightforward one: Does a particular emotion occur across cultures or not? The search to prove the existence of universal (basic) emotions has yielded some evidence for cross-cultural similarity in expressive behaviors and physiological responses.

### 1.1. Facial and Vocal Expression Research

Most research in support of universal emotions has focused on facial expressions. These studies have made use of a simple but elegant method. People from various cultures are typically asked to look at a picture of a face showing a given expression and to associate it with one emotion word from a list. The list consists of words for basic emotions such as happiness and anger. Across all cultures, people recognize a small number of facial expressions as representative of several emotions. This is true even for people and cultures with minimal exposure to the West.

Recognition studies on vocal expression (modeled after the facial recognition studies) also have found evidence for universality in the recognition of certain types of emotion expression. Individuals from various cultures were asked to identify the vocal expressions of emotions in either a nonsensical phrase or a phrase with meaning. People are clearly able to identify characteristic vocal expressions for a number of emotions (e.g., anger, happiness) even if they do not have information about the content of the utterance.

Cross-cultural studies on facial and vocal expressions show that humans have several universal modes of communicating their emotions. Therefore, certain elements of the emotion repertoire appear to be part of the human makeup. This evidence shows that emotions cannot be completely culturally defined.

However, the solid evidence for universality in facial expressions does not rule out the possibility of cultural differences or prove the existence of basic emotions. For instance, even though people can recognize facial and vocal expressions of other cultures, they are better at recognizing the expressions from their own culture. Thus, people judging faces and voices have an in-group advantage. Clearly, culture plays an important role in an individual's ability to recognize the expression of emotions.

The established recognition of facial and vocal expressions is not necessarily due to a direct correspondence between these expressions and distinct emotions. The results of facial recognition studies definitely suggest that people are able to know something about a person's psychological state from facial expressions (just as they can from anything else that a person does), but there is some evidence indicating that people would not spontaneously conclude that a particular basic emotion is occurring and that people draw this conclusion only when confronted with a scale that forces them to choose among various emotions.

### 1.2. Physiological Response Patterns Research

A few studies have also measured the physiological aspects of the autonomic nervous system associated with so-called basic emotions. In these studies, some individuals from various cultures were asked to relive an emotion from the past, and some were asked to move specific muscles in the face in motions that are assumed to mimic basic emotions. These studies showed some

evidence for similarity in physiological response patterns produced for several emotions across cultures.

However, there is reason to use caution in interpreting these data as conclusive evidence for basic emotions. First, in Western studies of emotions, it has not been resolved that emotions have characteristic physiological states. Scientists have pointed out, for example, that the physiological experience of fear from a loud sudden noise differs considerably from the physiological experience of fear of failing an examination. In addition, researchers have shown that there are no differences in the heart rates of people reliving fear, anger, and sadness expressions. Therefore, if a few studies show cross-cultural similarity in physiology of certain emotions, this does not prove that the same physiological responses always reflect the same emotions.

A second reason to use caution in interpreting these results is that participants from various cultures do not always experience those physiological patterns as emotions. In one study, people were asked to move their faces into the facial expressions typical of various emotions. The physiological responses of Minankabau men in Indonesia corresponded with those of American men who made the same facial expressions. However, the Minankabau reported experiencing no emotions, whereas the Americans reported experiencing emotions. Therefore, even if the pattern of physiology is similar across cultures, the emotional experience might not be.

In sum, scientists who believe that there are basic emotions have searched for unique patterns of features that are characteristic of particular emotions across all cultures. The research on facial and vocal expressions and on physiological response patterns shows that there are universal features of emotions. However, one cannot conclude that emotions themselves are universal because (a) some other features are culture specific and (b) the feeling associated with certain emotion features might be different across cultures.

## 2. WHAT COMPONENTS OF EMOTIONS ARE UNIVERSAL?

During the past two decades or so, some scientists searching for universalities have shifted away from the perspective that there are universal emotions. Instead, these researchers take the position that the building blocks, or “features,” of emotions can be the same cross-culturally and that research should be done to establish the similarity of the features of emotions across cultures.

Two classes of features of the emotional experience are appraisal responses and action readiness responses. Appraisal is an individual’s assessment of the meaning an event has for him or her. For example, a student who fails an exam will probably feel unpleasant (e.g., sad, angry) and be aware that the event has negative consequences. Action readiness is an individual’s impulse to behave in some way toward objects in his or her environment in response to the emotion being experienced. The student might feel like “disappearing” or condemning the professor, even if these reactions are not feasible.

For scientists who investigate the features of emotions, then, people of all cultures are likely to evaluate emotional events in terms of how they rate on certain features (e.g., pleasantness, expectedness, controllability, agency). Where an event falls on such dimensions is also thought to correspond universally to similar types of experience. For example, the appraisal that a situation is positively relevant to oneself is an important feature of a positive emotional experience to everyone regardless of culture; likewise, a situation judged to be negatively relevant to oneself universally implies the experience of a negative emotion. Another example is the aspect of agency (responsibility). An individual who attributes responsibility for an event to another person might feel anger instead of the sadness he or she might have experienced if no such attribution of responsibility were made. Therefore, the same appraisals are thought to lead to the same emotions cross-culturally. Similarly, action readiness aspects contribute to emotional experience. The tendency to aggress toward another person is a central feature of the anger experience, and the tendency to quit doing anything at all is central to sadness.

Although emotion component theorists assume that the building blocks of emotions are largely similar across cultures, they do not assume that people appraise or desire to react to situations in the same manner across cultures. In this way, emotional experience may differ across cultures despite some universality in the building blocks. The extent to which emotions are actually different across cultures is a question that must be resolved by future research.

### 2.1. Universal Dimensions of Appraisal and Action Readiness

There is some evidence that the same dimensions of appraisal and action readiness help to distinguish among emotions across cultures and that they do so at similar levels. The ability to distinguish among emotions

is, to a large extent (~40%), due to appraisal dimensions. Among the dimensions that appear to be universally discriminative among emotions are pleasantness, expectedness, novelty, and agency (responsibility and control). Similar levels of the ability to discriminate among emotions are accounted for by action readiness dimensions. Current evidence suggests a core of action readiness dimensions that include moving away from, moving toward, and moving against the objects of emotions. So, there is some evidence supporting the idea of universal appraisal and action readiness dimensions that make up and help to distinguish various types of emotional experience across the world.

It should be noted that differentiation among emotions on the basis of appraisal and action readiness is only a partial explanation of what distinguishes emotions from one another, and this is true across cultures. There remains the possibility that some important dimensions have not been included, and there may be other important components of emotional experience (in addition to appraisal and action readiness).

## 2.2. Appraisal and Action Readiness Dimensions and Emotions

To what extent are certain emotions characterized by similar patterns of appraisals? For example, are experiences of emotions in the category of anger across cultures associated with the same appraisals? Studies have suggested that a number of emotions (e.g., joy, anger, sadness) share patterns of appraisals that are quite similar across cultures. For instance, joyful situations have been cross-culturally characterized as expected, very pleasant, requiring no action, and enhancing self-esteem. However, in these studies, cultural differences in appraisal patterns, in addition to these core similarities, have been established. Thus, happiness can be appraised as good for social relationships in some cultures (e.g., United States) and as bad for social relationships in others (e.g., Japan). Current research shows that the cultural differences in appraisals are smaller than the similarities. However, there are some reasons why this conclusion might not reflect the true picture. First, most studies have used relatively homogenous samples such as college students from various countries. The differences among students who are exposed to education with Western influences are likely to be smaller than differences among populations in cultures that are less influenced by Western culture. Second, because appraisal dimensions have been suggested by Western theorists, there

may be a Western bias in the proposed dimensions. There is a possibility that non-Western cultures have important appraisal dimensions that have been missed so far. For example, several studies that included East Asian cultures have suggested the importance of the dimension of interpersonal connection, a dimension not typically included in current appraisal research.

Studies addressing the cross-cultural similarity in action readiness have been relatively scarce and less systematic. There is some evidence that a small set of emotions are universally characterized by core action readiness modes. Culture plays an important role, however, in lending specific contextual features that help to clarify differences in action readiness modes. For example, across many cultures, shame has been associated with the urge to “disappear.” In some other cultures, however, shame is associated with aggressive tendencies when it is caused by the behavior of someone very close to the affected individual.

In sum, research has shown that certain emotions have cross-culturally overlapping similarities in their core features. This finding supports some degree of the universality of emotions. At the very least, it suggests that similar patterns of appraisal or action readiness exist across cultures, and so there are similarities in the features of emotions. A limitation to this approach is its nearly exclusive reliance on retrospective self-report studies. So, one cannot be entirely sure to what extent the results reflect the ways in which people in various cultures think about their emotions or the ways in which emotions actually differ. On the other hand, it may be argued that emotional experience does not happen independently from the way in which people think about it and that self-reports provide a window on this experience that cannot be obtained in any other way. An additional limitation to this approach is that researchers of emotion features have been more interested in predicting similarities than in predicting differences. The differences that exist are equally informative about the nature of emotions and have been addressed by another group of researchers.

## 3. ARE EMOTIONS CULTURE SPECIFIC?

During the past two decades or so, some scientists have approached the issue of culture and emotion quite oppositely. Instead of looking for similarities in emotions and their components, these scientists have focused on the

cultural differences in emotions. These culture researchers believe that emotional experiences can be truly understood only by taking into consideration the culture in which they occur. Thus, some researchers try to understand how individuals' emotional responses relate to the values and practices of the culture.

The search for differences in emotions among various cultures begins with an insight into the ways in which cultures organize relationships and understand what a person is and should be. These social arrangements and understandings are referred to as cultural models. For example, achievement is one prevalent middle-class American cultural model. The belief that one should stand out from the crowd, the value placed on positive rewards, and the emphasis on personal influence and responsibility all are integral parts of this cultural model. Culture researchers have established that the particular models that are most important differ according to the culture in which one lives. For instance, this American cultural model can be contrasted with an important middle-class Japanese cultural model that stresses an individual's harmonious fit with his or her social environment. Culture scientists regard these cultural models as key ingredients in the overall emotional experience.

Although these scientists rely on cultural models as a context for understanding and predicting emotional responses, they do not assume that all people within a culture are alike. Because individuals in a culture will regard a given model in different ways, their individual emotional experiences will differ as well. For example, not all Americans value personal achievement equally, and they certainly differ in how they define achievement personally or in how they go about accomplishing success. What is critical, however, is that the overall culture still powerfully endorses the dominant cultural beliefs and social practices for how things are done in that culture, for example, when we are encouraged to "just do it" by a sports equipment company. The cultural models, therefore, set the reality boundaries within which emotions occur and are expressed. According to culture researchers, the models that are important to a culture will influence several aspects of emotions.

### 3.1. Antecedent Events

Cultural models are important in determining the occurrence of the events that cause an emotion to occur. How an event transpires is not part of the emotion itself, but it is an important determinant of which emotion will occur and what the responses of the individual will be. In other

words, certain situations may be especially likely to lead to certain emotions. If these situations occur more often in one culture than in another, the accompanying emotion is also more likely to occur in the former culture than in the latter. Likewise, if a situation is relatively unlikely to occur in a particular culture, the emotions that typically go with that situation are probably less likely to occur as well. How likely an event is to occur depends, at least in part, on how well it agrees with the overall values of the culture. Cultures tend to create more opportunities for events that are consistent with the prevailing models, whereas they tend to suppress opportunities for events that are at odds with these models. For instance, consistent with the cultural models, Americans create and encourage many contexts in which people are made to feel good about themselves because they stand out (e.g., praise, compliments, awards, trophies). The creation of many of those events has been shown to result in high occurrences of happy feelings. Similarly, in cultures that value a harmonious fit with other people (e.g., Utku Inuit, Chewong of aboriginal Malaysia, Tahitians, Japanese), acts of upsetting and frustrating those around a person are largely absent. The absence of frustrations is reflected in the relatively low levels of anger reported in these cultures.

People also selectively pay attention to certain aspects of situations in ways that are consistent with the cultural models. People in cultures that stress autonomy and uniqueness (e.g., United States) focus on good events and good qualities in the self so as to feel unique and independent. Therefore, individuals in this kind of culture are predisposed to strive for positive rewards. This focus on approaching positive outcomes is known as a promotion focus. In cultures where harmonious interaction with others is emphasized (e.g., Japan), individuals strive to avoid the occurrence of negative events by examining what could be improved about oneself. This focus on the prevention of bad outcomes (e.g., not living up to the standards) is referred to as a prevention focus. These cultural differences in predispositions for attending to certain events have implications for the emotions experienced as well. Consistent with their prevention focus, Chinese individuals report relaxation or relief when goals have been met and report anxiety when goals have not been reached. Americans' promotion focus leads them to report happiness when goals have been met and to report feelings of sadness when goals have not been reached. Thus, cultural models form the emotional lives of individuals within the culture by promoting events that are consistent with the models

and by focusing people's attention on particular aspects of situations that occur.

### 3.2. Appraisal

Cultural models also influence the ways in which people appraise or evaluate events. Certain appraisals are more likely to occur in certain cultures than in others. The prevalence of a particular appraisal will also differ across cultures, depending on the various cultural models. For instance, the appraisal of agency (responsibility) is markedly different across cultures. A key aspect of North American cultural models is success through independent personal accomplishment. Claiming responsibility and a personal sense of control are at the center of what it is to be a good person in this culture. In keeping with this model, North Americans typically attribute personal "responsibility" (either to oneself or to others) in many situations (especially positive ones), and they strive for a state of responsibility and control.

In contrast, East Asian cultural models emphasize the role of fate, how events are determined by many factors other than individuals themselves, and how individuals are interwoven with others in their social environment. Individual control and responsibility are not very relevant to the East Asian way of thinking. Accordingly, East Asians do not view control (or lack thereof) as an important aspect of emotional situations.

Thus, agency appraisals are markedly different across cultures. These evaluations of personal control or responsibility have important roles in the emotional experience. One implication of feeling responsibility, or feeling that someone else has responsibility, is that a person can influence or control other people in specific ways that fate or the world in general cannot. Claiming responsibility for an event results in active ways of emotionally reacting to, and coping with, the event. This differs considerably from the case of an event where one believes that the event could not have been humanly controlled.

Cultural models can also account for differences in the most frequent patterns of appraisal. For example, research has shown that there are differences in the events that are associated with feelings of pleasantness. In cultures that stress individual autonomy, pleasant experiences are those that are associated with evaluations of high self-esteem and personal control, whereas in cultures that stress interpersonal engagement and harmony, pleasantness is associated with events that make one feel connected with other people. Thus, pleasantness can be a

different experience in different cultures, depending on what each culture associates with pleasantness.

### 3.3. Relationship Goals and Action Readiness

Cultural models also outline desired goals for relationships with other people, thereby affecting an individual's impulses in a situation. In a series of studies in which people from various cultures reported what they felt like doing in a variety of emotion situations, Americans reported that they desired to (a) affirm or salvage positive self-esteem, (b) actively change the situations to make them consistent with their personal goals, and (c) prioritize their own personal concerns over those of their relationships. For example, in response to negative interactions, Americans reported the urge to break off their relationships and protect self-interests at the expense of their relationships. In contrast, Japanese reported action readiness modes that served to (a) maintain or improve their relationships and (b) adjust to other people's wishes and expectations. For instance, in negative interactions with others, Japanese wanted to reestablish closeness with the other persons and to keep their emotions to themselves.

Cultural differences in emotions are not random. Rather, they make sense when one considers the role of cultural models. The influence of cultural models goes beyond providing the content to fill in the blank of an otherwise standard emotion process in that cultural models affect the entire process of emotional experience. The context in which an emotion occurs, the ways in which the situation is perceived and evaluated, and the coping that is generated by the emotions experienced all are guided by cultural models. Thus, the cultural approach focuses on cultural differences instead of aiming to establish universalities. It is important to study both similarities and differences for a complete body of knowledge about emotions. However, the claim that emotions are primarily universal has been challenged by cultural researchers. During recent years, the methods adopted by cultural studies have established many cultural differences in emotion.

## 4. CONCLUSION

So, can one accurately interpret the emotions of people from other cultures? Few researchers of emotion are likely to answer with an unconditional "yes." However,

scientists' answers differ according to the ways in which they approach the question. Researchers who seek basic emotions believe that the most important emotions are similar across cultures. Researchers who have analyzed the separate building blocks of emotions in search of similarities contend that emotions in other cultures can be understood if one knows how certain situations are appraised and reacted to in other cultures. Finally, culture researchers believe that one must know and understand the major cultural concerns, meanings, and relational practices to have a sense of the emotional experiences in other cultures.

### See Also the Following Articles

Emotion ■ Intelligence, Emotional ■ Motives and Goals

### Further Reading

- Kitayama, S., & Markus, H. R. (Eds.). (1994). *Emotion and culture: Empirical studies of mutual influence*. Washington, DC: American Psychological Association.
- Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system: Activity in the Minangkabu of West Sumatra. *Journal of Personality and Social Psychology*, *62*, 972–988.
- Mesquita, B. (2003). Emotions as dynamic cultural phenomena. In R. Davidson, H. Goldsmith, & P. Rozin (Eds.), *Handbook of affective sciences* (pp. 871–890). New York: Oxford University Press.
- Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *112*, 179–204.
- Shweder, R. A., & Haidt, J. (2000). The cultural psychology of the emotions: Ancient and new. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 397–414). New York: Guilford.



# Emotion in Sport: An Individualized Approach

Yuri L. Hanin

KIHU – Research Institute for Olympic Sports, Jyväskylä, Finland

1. Introduction
  2. Individualized Approach to Performance Emotions
  3. Multidimensional Description of Performance Emotions
  4. Prediction of Emotion–Performance Relationships
  5. Explaining Emotion–Performance Relationships
  6. Individualized Assessments and Monitoring of Emotions
  7. Practical Implications
  8. Conclusion
- Further Reading

## GLOSSARY

**experience** The totality of characteristics, both past and present, that make up the particular quality of a person's performance; three aspects of experience are situational experience (states), relatively stable patterns of experience, and meta-experience (knowledge, attitudes, and preferences for [or rejections of] a person's experiences).

**individual zones of optimal functioning** An action-oriented framework for analysis of idiosyncratic emotional experiences related to athletic performance.

**in–out of the zone principle** The assumption that specific intensity of emotion has a beneficial (or detrimental) effect on individual performance; a high probability of individually successful performance is related to being in the optimal zone of intensity.

**psychobiosocial state** A situational, multimodal, and dynamic manifestation of the total human functioning.

**resources** Psychobiosocial (cognitive, affective, motivational, bodily, behavioral, operational, and communicative) assets determining an athlete's ability to perform up to his or her potential.

Understanding emotion–performance relationships is of crucial importance for applied sport psychologists. This article reviews an individual- and action-oriented framework, termed the Individual Zones of Optimal Functioning (IZOF) model, that describes, predicts, and explains individually optimal and dysfunctional emotional experiences related to athletic performances. To describe emotion–performance relationships, the IZOF model proposes a multidimensional conceptualization of emotions as a component of the psychobiosocial state. Finally, the model suggests several empirically substantiated guidelines for person- and task-relevant assessments focused on emotion self-regulation and optimization of performance.

## 1. INTRODUCTION

One of the key questions in sport psychology is how and why pleasant and unpleasant emotional states affect

sport performance. Indeed, are unpleasant emotions always bad for performance? For instance, can anger sometimes be useful in competition? On the other hand, are pleasant emotions always good for performance? For instance, can self-confidence lead to poor performance through complacency and underestimation of task demands? Moreover, do athletes perform up to their potential when they are stress free or when they are deliberately using competitive stress to their advantage? Is precompetition anxiety bad for all athletes? Finally, is it possible to predict an athlete's performance based on content and intensity of emotional experiences? Researchers, applied sport psychologists, athletes, and coaches often need straightforward answers to these practical questions. Most of the existing models in sport psychology focus mainly on interindividual differences between successful and unsuccessful athletes. On the one hand, there are exceptional athletes who reach the top and stay at the top with remarkable consistency. On the other hand, it is important to realize that even the winners have bad days and often exhibit inconsistent performance. In other words, performers across the full range of skill and experience are quite familiar with occasional performance slumps and "choking." Thus, instead of comparing high achievers with underachievers, an alternative approach would be to focus more on performance experiences of each individual athlete. Such an approach has direct practical implications and can be applied to a particular athlete or a team. This article describes an individual-oriented, research-based framework, termed the Individual Zones of Optimal Functioning (IZOF) model, to show how the understanding of emotion–performance relationships at the individual level can be helpful in optimization of athletic performance.

## **2. INDIVIDUALIZED APPROACH TO PERFORMANCE EMOTIONS**

Successful performance enhancement interventions in sport should (a) focus on unique challenges and concerns facing coaches and athletes, (b) emphasize practical application of research-based principles, and (c) realize that athletes are healthy achievement-oriented people who are "forced" to cope with performance-induced stress. The IZOF model aims to match these requirements with a special emphasis on individual athletes. The following subsections briefly describe the basic assumptions of the model.

### **2.1. The Meaning of the IZOF Acronym**

The IZOF model, developed in high-achievement sport settings during the past 30 years or so by Hanin, is a sports-specific, action-oriented framework. It focuses on analysis of the structure, dynamics, and function of situational emotional experiences (e.g., nervous, worried, calm, confident) related to individually successful and less than successful performances. The model also examines relatively stable patterns of these experiences and meta-experiences (i.e., beliefs about expected effects of feeling nervous, worried, or calm on performance). The IZOF acronym reflects selected features of the model.

#### **2.1.1. Individual**

The model is an idiographic approach that focuses on idiosyncratic emotional experiences of individual athletes (or teams). Therefore, the main emphasis is on the within-individual dynamics of subjective emotional experiences and meta-experiences related to performance.

#### **2.1.2. Zones**

The zone principle implies a specific relationship between the perceived intensity of emotional state and the quality of performance. Individually optimal intensity (high, moderate, or low) reflects an athlete's past performance history and available resources. Being in the optimal zone implies a high probability of individually successful performance. Thus, the zones serve as empirically established individualized criteria for prediction of performance, and the "in–out of the zone" concept is used to evaluate a match (or mismatch) between the currently experienced level of emotion intensity and previously established optimal and dysfunctional zones.

#### **2.1.3. Optimal**

Optimal emotions are defined as most relevant and appropriate for a particular athlete performing a specific task. The optimal performance state results in a total task involvement and the best recruitment and use of available resources. This usually results in a high-quality performance process and achievement of individually successful performance outcomes (although not necessarily "peak or best ever"). The optimal performance state tends to include both pleasant and unpleasant emotions reflecting idiosyncratic strategies that athletes use in recruitment and use of their resources. Therefore,

the functionally optimal state (e.g., situational anger, high anxiety) is different from the “ideal” or preferred state (e.g., “flow”), which usually has only positive characteristics. Finally, the concept of optimality is multidimensional and is related not only to emotion intensity and content but also to time and context dimensions.

#### **2.1.4. Functioning**

The term “functioning” emphasizes the fact that a detailed description of performance emotions is just a tool to understand their optimal and dysfunctional effects on task involvement and the quality of performance process. Optimal function is manifested in an athlete’s efficient recruitment (effort) and use (skill) of available resources. This usually results in energizing and organizing effects on performance. In contrast, emotion dysfunction usually reflects a failure to recruit resources and their inefficient use, resulting in deenergizing and disorganizing effects of emotion on athletic performance.

### **2.2. Basic Assumptions of the IZOF Model**

The IZOF model holds that emotion is a component of adaptive or maladaptive response in person–environment interactions; emotion is an integral part of a person’s psychobiosocial state, performance process, and total functioning; emotion is triggered by appraisals of a person’s relationships with the environment; and emotion is a reflection of critical moments of imbalance in person–environment interactions. The model describes the nature of the emotion–performance relationships as follows. First, emotions are triggered by a person’s appraisal of the probability of achieving expected individually relevant goals. Second, because sport activity is repetitive, situational emotional experiences gradually develop into relatively stable emotion patterns. Third, emotion patterns are specific to the individual, task, and setting, and this specificity is manifested in the form, content, intensity, time, and context dimensions of the psychobiosocial state. Fourth, emotion–performance relationships are reciprocal. Emotions affect performance; however, changes in the ongoing performance process and intermediate (and final) outcomes produce a shift in a personal meaning of the situation, resulting in a shift in emotion content and intensity. A shift in the direction of impact (causality) is usually determined by ongoing appraisals of the performance process. Fifth, prediction of performance should be based on

assessment of interaction effects of a cluster of really experienced emotions (optimal and dysfunctional) rather than on separate effects of selected “pure” emotions. Sixth, because athletes often reflect on their experiences in successful and less than successful performances, they gradually develop meta-experiences. For instance, an athlete could notice that high anxiety sometimes helps him or her to focus better or that it does not affect performance negatively. This knowledge leads to a positive attitude toward anxiety, which is then interpreted as an indicator of this athlete’s readiness for competition. Emotion self-regulation and individual-oriented interventions in the IZOF model are based on the concept of optimal personal meaning and adequate meta-experiences triggering specific emotions.

## **3. MULTIDIMENSIONAL DESCRIPTION OF PERFORMANCE EMOTIONS**

The IZOF model defines the performance-related psychobiosocial state as a situational, multimodal, and dynamic manifestation of the total human functioning. A multi-level and system description of performance-related states includes at least five interrelated dimensions: form, content, intensity, time, and context. Three of these dimensions (form, content, and intensity) describe the structure of subjective experiences and meta-experiences, and two of these dimensions (time and context) characterize the dynamics of performers’ subjective experiences. When viewed collectively, these five basic dimensions provide a framework for a relatively complete description of the phenomenon and can guide data collection, analysis, interpretation, and generation of new ideas.

### **3.1. Form Dimension**

An athlete’s psychobiosocial state manifests itself in seven basic forms (modalities): cognitive, affective, motivational, bodily, motor–behavioral, operational, and communicative. All seven modalities are interrelated and provide a relatively complete description of a performance state, including experiences and their displays (expression or suppression). Thus, emotion is construed as an important component of the performance-related psychobiosocial state. Current individual-oriented research focuses on affective (emotional), motivational, and bodily components of the performance state and their interactive effects.

### 3.2. Content Dimension

The content dimension is a qualitative characteristic of the performance state. The IZOF model as an alternative individual-oriented approach categorizes emotion content within the framework of two closely related but independent factors: a hedonic tone (pleasure–displeasure) and functionality (optimal–dysfunctional impact of emotion on athletic performance). Both factors reflect qualitatively different experiences related to individually successful and poor performances. The four broad emotion categories derived from these distinctions are pleasant and functionally optimal emotions (P+), unpleasant and functionally optimal emotions (N+), pleasant and dysfunctional emotions (P–), and unpleasant and dysfunctional emotions (N–). These four categories provide an initial robust and sufficiently broad structure to generate a wide range of idiosyncratic, individually relevant, and task-specific emotions really experienced by athletes prior to, during, and after successful and less than successful performances. Finally, the content of these athlete-generated emotions can be recategorized to see how idiosyncratic descriptors match or mismatch the existing “basic” or discrete emotion frameworks proposed in mainstream psychology.

### 3.3. Intensity Dimension

The intensity dimension as a quantitative attribute of subjective experiences is expressed in either objective or subjective metrics on a selected component or modality of the performance state. Different levels of intensity can produce different functional effect related to energizing (deenergizing) and organizing (disorganizing) aspects of the performance process. The in–out of the zone notion describes a selected range of emotion intensity producing optimal, neutral, or dysfunctional effects on individual performance. For instance, a high probability of successful performance is expected if an athlete’s actual state is in (or close to) the earlier established optimal zone of intensity. In contrast, if an athlete’s actual state is out of the optimal zone, the athlete is likely to perform below his or her potential.

### 3.4. Context Dimension

The context dimension is an environmental characteristic reflecting the impact of situational, interpersonal, intragroup, and organizational factors on emotion intensity and content in sport setting. Situational impact is observed in emotional reactions triggered in practices

and competitions by athletes’ anticipated or real contacts and interactions with significant others (e.g., partners, coaches, teammates). This dimension also includes culturally coded and culturally determined beliefs of participants about the expected impact of specific emotions on their performance and about the rules of emotion display in a particular subculture.

### 3.5. Time Dimension

The time dimension includes topological (e.g., phases, cycles, sequencing, periodicity, timing) and metric (e.g., duration, frequency) characteristics reflecting the dynamics of performance-related experiences. The short-term dynamics in sporting activity is observed in emotional states prior to preparation for an action, during task execution, and after performance in a single competition (practice). The long-term temporal dynamics is related to emotion–performance relationships during a competitive season(s), a 4-year Olympic cycle, or an athlete’s sports career.

## 4. PREDICTION OF EMOTION–PERFORMANCE RELATIONSHIPS

There are certain myths in sport psychology about emotion–performance relationships based on unsubstantiated beliefs that have been challenged recently. Some examples include the notions that there exists a universally optimal intensity of emotions facilitating performance of all athletes (e.g., moderate anxiety), that all negative emotions are always dysfunctional for athletic performance, and that all positive emotions are always optimal for athletic performance. In contrast, IZOF-based individual-oriented and empirically supported predictions of emotion–performance relationships indicate that these notions are not the case.

### 4.1. Interindividual Variability

As predicted, research revealed a high degree of interindividual variability in the intensity and content of idiosyncratic optimal and dysfunctional emotions accompanying individually successful and poor performances. Thus, different athletes can perform up to their potential experiencing emotions of different content and intensity. Optimal emotional states reflect (a) individual differences in available resources, (b) athletes’ ability to recruit and use these resources

efficiently, and (c) athletes' individual coping strategies to compensate for either a lack or an insufficiency of resources (low situational readiness for competition).

#### 4.2. In–Out of the Zone Principle

The prediction of individually successful, average, or poor performance based on the in–out of the zone principle has received some empirical support in several studies. Thus, a player's emotional state is first described in terms of individually relevant emotion content and intensity. Then, current emotion intensity is contrasted with the previously established individually optimal and dysfunctional intensity zones. A high probability of successful performance is expected when emotion intensity is within the optimal zones and outside dysfunctional ranges. Therefore, a large discrepancy (in intensity and emotion content) between the actual state and the individually established optimal zones indicates a high probability of less than successful performance.

#### 4.3. Interactive Effects

Research indicates that most emotional experiences involve clusters of “mixed” rather than “pure” emotions. For instance, an international-level tennis player, prior to his best ever match, felt high intensity of pleasant optimal emotions (P+). He felt highly determined, confident, excited, dynamic, and comfortable. At the same time, he felt moderately aggressive, alarmed, and somewhat uncertain (N+). Moreover, his unpleasant dysfunctional emotions (N–) (e.g., nervous, afraid, worried, intense) were of low intensity. This pattern was similar to that during his best match; however, he felt alert and quick but not excited too much without premature satisfaction. In contrast, in the worst match of his career, this player felt highly nervous and worried (N–), and these experiences were more intense during the match. Interestingly, at the same time, his optimal pleasant emotions prior to and during the match were of moderate and low intensity, respectively. If applied sport psychologists measured only anxiety, for instance, they would miss the entire picture of emotional clusters. Therefore, the interactive effects of emotions enhancing and impairing sporting activity are of crucial importance for an accurate prediction of emotion–performance relationships. In this case, a high probability of individually successful performance is expected when combined maximum-enhancing and minimum-impairing effects are observed. On the other hand, a high probability of individually average and impaired performance is

expected when a combination of high-enhancing and high-impairing effects or of low-enhancing and low-inhibitory effects is observed. Finally, a high probability of poor performance is expected when low-enhancing and high-inhibitory effects are observed.

#### 4.4. Dynamics and Bidirectionality

Emotion–performance relationships are dynamic and bidirectional. In other words, precompetition emotions affect performance, and ongoing performance affects the dynamics of mid- and postcompetition emotions. Thus, to predict the entire process of emotion–performance relationships, it is important to establish patterns of emotion impact on performance and of performance impact on emotions.

All of these predictions were tested empirically across various samples and sports in different countries. Initially, the IZOF model focused on precompetition anxiety as a discrete stress-related emotion syndrome with a “fixed” emotion content. The main emphasis in the IZOF–anxiety research was on identifying individually salient intensity of state anxiety. Specifically, it was revealed that each athlete had an individually optimal intensity level (which could be high, moderate, or low) and zones of anxiety enhancing the athlete's performance. The analysis of an athlete's past performance history identified these optimal level and intensity zones, which served as tentative individualized criteria in the prediction of current and future performances. Thus, the probability of successful performance was high when current precompetition anxiety was near or within the previously established individually optimal intensity zones. When precompetition anxiety fell outside the zones (i.e., higher or lower), individual performance usually deteriorated. Recently, similar results were obtained in the studies of optimal and dysfunctional effects of situational anger on athletic performance.

Research indicated that an exclusive focus on precompetition anxiety or anger was limited because these syndromes were only a part of athletes' emotional experiences related to performance. Therefore, the IZOF model was extended to the study of pleasant and unpleasant emotions enhancing and impairing individual performance, and more consideration was given to the saliency of other dimensions of emotion (content, time, and context). A recent extension of this line of research tested the validity of the preceding predictions in motivational, bodily, and performance correlates of the psychobiosocial state.

## **5. EXPLAINING EMOTION– PERFORMANCE RELATIONSHIPS**

Most of the existing approaches in sport psychology focused mainly on antecedents and consequences of selected emotion syndromes (e.g., anxiety, anger, depression). In contrast, the main emphasis in the IZOF model was first on an accurate and detailed description of various forms of idiosyncratic emotional experiences in athletes. To explain the antecedents (emergence and dynamics) and consequences (impact on performance) of performance-related emotional experiences, the model uses several constructs that are briefly reviewed in the following subsections.

### **5.1. Appraisals and the Emotion Process**

Person–environment interactions are important in the IZOF model; therefore, a psychosocial perspective is central in the functional interpretation of the dynamics of emotion–performance relationships. First, emotion is conceptualized as an unfolding process reflecting person–environment interactions. Ongoing appraisals of these interactions result in a change in personal meaning of the situation, and this exerts the influence on emotional experiences related to performance. Second, changes in personal meaning reflecting the dynamics of the performance process trigger functional shifts in emotion content and intensity. Specifically, anecdotal evidence and practical experience indicate that functionally optimal pleasant and unpleasant emotions (P+N+) prior to and during activity are usually anticipatory and are triggered by the appraisals of challenge and threat. In contrast, situationally dysfunctional pleasant and unpleasant emotions (P–N–) during activity are usually triggered by perception of “achieved” outcomes (appraisals of gain and loss) before the task is completed.

### **5.2. Emotion Impact on Performance**

Two constructs with their opposites related to energizing and organizing aspects of emotion account for the impact of emotions on the athletic performance process. These include the concepts of energy mobilization (and demobilization) and energy use (and misuse). Therefore, optimal and dysfunctional emotion function in the IZOF model is conceptualized within the framework of two closely related but independent factors: (a) energy mobilization (e.g., optimal effort,

intensity) and (b) energy use (e.g., efficiency, optimal information processing). The former is related to the situational resources available to an individual performer, whereas the latter characterizes the efficiency of using these resources. Based on these two factors, four relatively independent global emotion functions are derived: (a) energizing or energy-mobilizing function, (b) energy demobilizing function, (c) energy use or regulation function, and (d) energy misuse or deregulation function. These four functions provide a framework for interpretation of separate and interactive effects of positive and negative emotions on individual performance. Based on the nature of these interactions, the total impact of emotions on athletic performance can be optimal (regarding effort and skill), paraoptimal (with only effort or skill, but not both, being optimal), or dysfunctional (regarding effort and skill).

From the functional effect perspective, the constructs of energy mobilization and use (and their opposites) seem to be useful in explaining why in some athletes optimal emotions are predominantly pleasant, whereas in other athletes they are unpleasant. For instance, low-anxious athletes are typically “smart users” of available energy and are less distracted by task-irrelevant and energy-wasting concerns. In contrast, high-anxious athletes typically generate more energy, especially in stressful or emergency situations, because they are often less efficient in its use due to a narrow attention focus and an overload in the information-processing function. Thus, unpleasant emotions, such as anxiety, are functionally useful for these athletes in that they help to generate additional energy to compensate for the apparent limitation in information processing or the use of energy. Because the quality of athletic performance is usually related to both the amount of available energy and its efficient use, different athletes can be successful by using different resources. In other words, the same level of performance may be achieved either through the increase of total effort or through skillful (smart) use of available resources (efficiency).

## **6. INDIVIDUALIZED ASSESSMENTS AND MONITORING OF EMOTIONS**

An accurate individualized assessment and monitoring of performance-related emotional experiences provides a basis for predictions of emotion impact and self-regulation. In the following subsections, selected data collection techniques that have been used to assess

individually relevant idiosyncratic emotion content in sport are briefly described: interviews, emotion profiling, self-generated metaphors, and narratives.

### 6.1. Interviews

The first and most popular data collection technique for eliciting idiosyncratic emotion content is structured and semistructured in-depth interviews with open-ended and probing questions activating an athlete’s recall of thoughts and feelings prior to and during best and worst performances. It is usually recommended in pilot and exploratory studies to generate idiosyncratic labels and then to aggregate most selected items into a stimulus list or standardized emotion scale. Identifying idiosyncratic content and personally meaningful labels to describe athletes’ subjective experiences is a clear advantage of these interviews over questionnaires with researcher-generated items and “fixed” emotion content. Experienced and verbally skillful athletes can provide detailed and meaningful accounts of their experiences (how they felt) and meta-experiences (how they interpreted and coped with these feelings) prior to and/or during the competition. Athletes who are less aware of their performance-related experiences might require some assistance initially in structuring and focusing their recall.

### 6.2. Individualized Emotion Profiling

A person’s perspective is central in the reconstruction of athletes’ emotional experiences and meta-experiences related to successful and less than successful performances. The individualized emotion profiling aims to analyze athletes’ past performance history and significant emotional experiences and meta-experiences. Specifically, athletes generate individually relevant emotion words that best describe their optimal (helpful) and dysfunctional (harmful) pleasant and unpleasant emotional experiences accompanying their performances in best and worst competitions. A stimulus list, compiled from existing emotion scales or based on previously generated items, can assist athletes in generating emotion descriptors. The IZOF-based emotion list consists of 46 pleasant emotions (arranged in 14 synonymous rows) and 50 unpleasant emotions (arranged in 14 synonymous rows). Table I shows a group-oriented functional categorization of pleasant and unpleasant emotions based on several studies of athletes representing various sports. However, research also indicates that in each individual case, optimal and dysfunctional emotions can be different from those in the aggregated group data. Therefore, individualized emotion profiling is strongly recommended.

TABLE I  
Functionally Categorized Emotion Stimulus List for Individualized Profiling

<i>Pleasant emotions</i>	<i>Unpleasant emotions</i>
P+ <ul style="list-style-type: none"> <li>• Active, dynamic, energetic, vigorous</li> <li>• Quick, rapid, fast, alert</li> <li>• Confident, certain, sure</li> <li>• Determined, set, settled, resolute</li> <li>• Brave, bold, daring, dashing</li> <li>• Inspired, motivated, stimulated</li> <li>• Excited, thrilled</li> </ul>	N+ <ul style="list-style-type: none"> <li>• Angry, aggressive, furious, violent</li> <li>• Intense, fierce</li> <li>• Tense, strained, tight, rigid</li> <li>• Annoyed, irritated, distressed</li> <li>• Anxious, apprehensive, worried</li> <li>• Concerned, alarmed, disturbed, dissatisfied</li> <li>• Nervous, jittery, uneasy, restless</li> </ul>
P- <ul style="list-style-type: none"> <li>• Delighted, overjoyed, exhilarated</li> <li>• Relaxed, comfortable, easygoing</li> <li>• Calm, peaceful, unhurried, quiet</li> <li>• Cheerful, merry, happy</li> <li>• Glad, pleased, satisfied, contented</li> <li>• Lighthearted, carefree</li> <li>• Nice, pleasant, agreeable</li> </ul>	N- <ul style="list-style-type: none"> <li>• Discouraged, dispirited, depressed</li> <li>• Afraid, fearful, scared, panicky</li> <li>• Doubtful, uncertain, indecisive, irresolute</li> <li>• Helpless, unsafe, insecure</li> <li>• Inactive, sluggish, lazy</li> <li>• Sorry, unhappy, regretful, sad, cheerless</li> <li>• Tired, weary, exhausted, worn out</li> </ul>
Your own emotion: .....	Your own emotion: .....

Individualized emotion profiling involves several steps. First, to identify optimal emotion patterns, athletes select four or five positive items and then four or five negative items that best describe their emotions related to their individually successful performances in the past. Then, athletes identify their dysfunctional emotions related to individually poor performances by selecting four or five positive items and four or five negative items. Although athletes use the emotion stimulus list to generate individually relevant descriptors, they can also add emotion words of their own choice. The focus of recall is either on a specific situation (precompetition) or on repeated experiences across several similar situations. Individually optimal and dysfunctional intensities for each emotion item are established using the Borg's CR-10 scale (ranging from "nothing at all" to "maximal possible").

In sum, individually oriented and task-specific emotion profiling identifies (a) individually relevant emotion content, (b) individually optimal and dysfunctional emotion intensity, (c) individually relevant context, and (d) interaction effects of optimal and dysfunctional emotions. In a comprehensive IZOF-based assessment program, first, the individual's past performance history is examined within the context of the current situation. Then, individually relevant emotion content is generated. Next, optimal (and dysfunctional) intensities for each emotion item are identified. Finally, the emotion profiles (with intensity zones for each individual item) are validated and refined in a series of repeated self-ratings in specific performance situations (practice or competitions). For monitoring emotional states in a field setting, the total emotion profile can be condensed to a few most important markers (e.g., three items for each of the four emotion categories). The major emphasis in individualized emotion profiling is on analysis and reconstruction of athletes' performance-related idiosyncratic experiences and meta-experiences by using their own vocabularies.

Research shows that athlete-generated emotion descriptors usually have strong connotations with none-motion concomitants of the psychobiosocial state such as cognitive (e.g., alert, focused), motivational (e.g., willing, determined, motivated), and bodily (e.g., relaxed, tense, jittery) experiences. Therefore, there is a clear need to extend the focus of individualized profiling from emotions to other modalities of an athlete's state.

Moreover, a mere description of emotional experience (e.g., the same high level of precompetition anxiety in an athlete on two occasions) might not explain why the athlete was very successful in the first competition and less than successful in the second competition. Here, additional qualitative data (e.g.,

motivational domains or causes of these experiences and resulting self-talk) could be instrumental. For instance, a highly anxious athlete in a successful race could feel "psyched up," "perceive the race as important," and "focus on competing with tough opponents." This athlete's self-talk would be clearly self-empowering (e.g., "I can run at this level," "I'll show them all," "It is my day today"). In contrast, prior to a less than successful race, this athlete's thoughts included "doubts in her current shape," "insufficient recovery," "health problems," and/or "poor team climate," and the resulting self-talk was clearly self-defeating in that there was too much focus on opponents' strengths (e.g., "they are going to beat me," "I've got no chance in this race," "today is not my day," "the race is meaningless").

### **6.3. Self-Generated Metaphors**

Although the individualized emotion profiling may be successful in generating idiosyncratic descriptors of emotions, there is still a need for a more holistic description of the psychological meaning of the performance situation and an athlete's role in it. To address this concern, Hanin and colleagues proposed a metaphor self-generation method to examine the feasibility of symbolic representation of performance-related experiences using idiosyncratic athlete-generated metaphors. The approach is based on the symbolizing ability of people and on the notion that people give structure, meaning, and continuity to their activities by symbolizing their experiences.

A metaphor is always related to a certain image, which is holistic and meaningful for a person. Therefore, words, objects, mental images, stories, and the like are the units of metaphor. A symbolic nature of a metaphor provides a picture of one's understanding of something, with an emphasis on most personally relevant and important aspects and shades of meaning. Well-known metaphors describing performance-related states in sport include being "in the zone," "in the groove," "in the cocoon," and "in the flow state" as well as having "choked." Research suggests that athlete-generated idiosyncratic metaphors may be useful in the practice of sport psychology for holistic and personally meaningful description and better understanding of performance-related situational experiences and relatively stable patterns of experiences.

The metaphor self-generation method is an instrument identifying metaphors and accompanying interpretive descriptors of feeling states prior to, during, and after best and worst competitions. Thus, the

focus is on assessment and symbolic representation of idiosyncratic content (quality) of emotional experiences in a high-achievement setting. Several studies using a metaphor generation method in various performance situations revealed that athletes in various sports, and at various age and skill levels, are able to generate highly idiosyncratic images symbolically representing their high or low action readiness in successful and poor performance situations.

Metaphors reflecting available resources (e.g., “a boat with an engine,” “a mountain river”) and ability to recruit resources (e.g., “a horse in a light cart,” “a man who can move a mountain”) have been used to describe athletes’ states in best ever competitions. A lack of resources (e.g., “a sinking boat,” “an empty bottle”) and inability to recruit resources (e.g., “a man with weights on his legs,” “a bird unable to fly”) are characteristic of metaphors that athletes have used to describe their states in worst ever competitions. In addition, metaphors describing best ever competitions have reflected highly efficient use of resources (e.g., “Batman in flight,” “a fisherman catching a golden fish”), whereas metaphors describing worst ever competitions have reflected inefficient or poor use of resources (e.g., “a bear after hibernation,” “a soldier in a burning tank”).

An example of using metaphors to monitor feeling states of a Finnish Olympic-level track and field male athlete in a series of international competitions during the season illustrates a practical utility of the method. A chain of metaphor images generated by this athlete for various contests included “a colt on a spring pasture” (very strong but too excited to control his own performance), “a tractor” (started moving in the right direction), “a Volvo” (good routines, self-confident, and reliable performance), “a Mercedes” (ready and knew what to do, enjoying successful performance), and “a Ferrari” (feeling great, powerful, and in control). These metaphors helped to identify and anchor the athlete’s optimal level of confidence, the right focus, and self-control skills prior to and during competition.

Taken together, research findings indicate that idiosyncratic content and personally relevant meaning of metaphors are either self-empowering (usually in best competitions) or self-defeating (usually in worst competitions) images. Furthermore, the quality of metaphors is culturally coded and determined by the specific contexts in which athletes perceive and experience performance-related episodes. Context-induced beliefs about potential effects of pleasant and unpleasant emotions can have a strong impact on perception of consequences of emotional experiences for athletic performance. Thus,

athletes can perceive competitive stress either as an undesirable factor to be avoided or as a part of their job with which they must cope. Consequently, if athletes believe that high precompetition anxiety is detrimental to their performance, the quality of such anxiety will be different from that of anxiety perceived as an indicator of high readiness for the competition. Therefore, before attempting any intervention program, sport psychologists must identify these context-determined beliefs based on past idiosyncratic experiences.

Research findings also indicate that metaphors differentiate various contexts (e.g., best ever and worst ever competitions) and capture the dynamics of personal meaning across pre-, mid-, and postcompetition situations. For instance, in best ever competitions, metaphors symbolizing strength, power, speed, and skillfulness were generated 5 to 10 times more often than in worst ever competitions. In contrast, in worst ever competitions, metaphors symbolizing weakness, slowness, and powerlessness were generated 5 times more often than in best ever competitions. Similar and even stronger relationships were observed regarding the efficient recruitment and use of available resources.

Idiosyncratic metaphors symbolically representing the content of feeling states experienced throughout competitions were different (low content overlap) across pre-, mid-, and postcompetition situations. These data provide support for the notion that a performance task usually involves three interrelated, but functionally and qualitatively different, stages: preparation for, execution of, and evaluation of performance. The findings also suggest that the functional and personal meaning of these three situations, as represented symbolically by idiosyncratic metaphors, is different for each athlete. A change in meaning triggers a change in emotional states. In other words, emotion dynamics is observed not only in the intensity of emotions but also in the content (quality) of emotions. Therefore, in research and applications, it is important to identify a specific constellation (a set or cluster) of emotion content that is optimal or dysfunctional for an athlete’s performance. Moreover, metaphors reflecting even slight changes in personal meaning of each performance situation could be useful markers for applied sport psychologists in monitoring the dynamics of performance states and in evaluating the effectiveness of individualized interventions.

#### 6.4. Narratives

A narrative is an account of several interconnected occurrences. It is a scheme by means of which people

give meaning to their experiences. It also provides a framework for understanding the past events of one's life and for planning future actions. It is the primary scheme by means of which human existence is rendered meaningful. As Bruner argued in 1991, narratives are self-stories conveying intention and feeling related to a sequence of events; therefore, they provide a holistic picture of unfolding processes. Narratives (as self-stories told by athletes and coaches) can describe concrete performance situations and can identify automatic thoughts and emotional responses. Thus, they help to recreate a high-level performer's experiences, meta-experiences, and available (or unavailable) coping strategies.

The following transcript illustrates how a 16-year-old ice hockey goaltender perceives and describes his problem during important games:

This is perhaps a silly question, but can someone help me? I am very good technically and physically with the qualities needed for a goaltender. I always train much and well. For instance, I trained on the average 12 times per week during the summer season. However, I have one very big problem. In the game, I start feeling very nervous and fear that I will play badly. Then usually easy things become very difficult. . . . This is especially often in the games where it is necessary to win. Sometimes the tension goes away if the play starts well. I know that I am really good. My coach thinks that my flexibility and other qualities are top level in Finland. . . . In training, I can tend off really well, but in really tough situations, I cannot stand it. . . . Now it is again the time when I have to succeed, especially in training games. Based on these results, they will select a future national junior team.

This narrative describes briefly the goaltender's feelings and performance in practices and tough games, accompanied by his general self-ratings and expectations. However, the main advantage of narratives is that they provide a means for holistic and dynamic reconstruction of the unfolding process of athletic performance in a real-life setting. Narratives focus on the functions of emotion feelings and displays in the episodes of everyday life (or a specific activity). Thus, of special value is the fact that storytelling includes not only emotion words as descriptors of real idiosyncratic experiences but also an athlete's interpretations of these experiences and culturally determined beliefs within a specific context (i.e., meta-experiences). Although the main emphasis so far was on emotion experiences, it is important to realize that emotion displays (i.e., expression and suppression of feelings)

also have specific features and stereotyped views in the context of competitive and high-achievement sport. In this setting, athletes also learn to read the body signals of their opponents to identify the opponents' feeling states and action readiness.

## 7. PRACTICAL IMPLICATIONS

Applied sport psychologists assisting athletes in assessments and self-regulation of emotions usually face three issues: (a) identification of emotional states related to individually successful and poor performances, (b) prediction of emotion–performance relationships, and (c) selection of person- and task-relevant techniques of self-regulation. Compelling empirical evidence in support of the IZOF model provides several tentative guidelines on how to deal with these three issues.

First, to identify individually optimal and dysfunctional emotional experiences, establishing an individually relevant cluster (constellation) of emotions and their intensities prior to, during, and after successful and less than successful (poor, average, or customary) performances is recommended. These qualitatively and quantitatively extreme situational experiences serve as individualized criteria in the evaluation of current and anticipated emotional states. In addition, it is important to identify athletes' specific beliefs and attitudes about their emotion impact on performance. Are they aware of such effects? How do they usually cope with stress- and complacency-producing situations? Are these situational emotional experiences random or relatively stable patterns that athletes try to reproduce in important competitions? The main purpose of such individualized assessments is to enhance athletes' awareness and acceptance of these experiences.

Second, prediction of emotion–performance relationships is based on the in–out of the zone notion using categorical or continuous approaches. A categorical approach predicts performance based on the comparison between previously established individual zones and actual scores of intensity. A continuous approach is based on perceived “intensity impact” contingencies along the entire working intensity range of each emotion. Here, the emphasis is on an estimation of partial and total effects rather than on only a selected optimal range of each emotion. In both cases, a decision about emotion regulation (increase or decrease) is based on the magnitude of deviations either from optimal and dysfunctional zones or from a total effect in the selected modality. Furthermore, the intervention should aim not

TABLE II  
Conventional and IZOF-Based Features of Emotion Assessment and Self-Regulation

<i>Basic principle</i>	<i>Conventional feature</i>	<i>IZOF-based feature</i>
Multimodality	One to three components	Seven basic forms
Multirange/-content	Fixed/Same for all	Idiosyncratic
Multizone/-intensity	Moderate or nonspecified	IZOF/Idiosyncratic intensity
Multistage	Mainly precompetition	Pre-, mid-, or postcompetition
Multidirection	Decreases or nonspecified	Individualized up/down
Multitask	Nonspecified	Specified and identified
Multifunction	Nonspecified	Effort or skill focused
Multimethod	One or two researcher generated	"Own," tailored or combined

only at helping athletes to enter or reenter their optimal zones but also at staying away from the dysfunctional zones. Finally, predictions should also consider the total functional impact of emotion on performance that is usually manifested in an increase or a decrease in effort (energy) and in efficiency or inefficiency in the use of available resources.

Third, there are numerous techniques of emotion self-regulation in the practice of applied sport psychology. Effective emotion self-regulation should be based on already described individualized assessments and predictions of emotion–performance relationships. Moreover, a selected method or intervention strategy (technique) should be optimal for both the athlete and the situation. In other words, the method should match previously established individual patterns of coping with emotion-triggering situations. In addition, the effective intervention program usually includes several appropriate methods of self-regulation.

Finally, there are several directions for effective emotion regulation. Most of them focus directly on emotional response by using various mental skills. However, there are other options that include a change of the current situation or its perception (personal meaning) by an athlete or a special organization of athletic activity for an athlete or a team (role expectations and game tactics). Table II shows IZOF-based and conventional (group-oriented) features reflecting several practical guidelines that are important to consider in assessments and regulation of emotional states.

## 8. CONCLUSION

The main purpose of this article was to provide an overview of the IZOF model as an alternative

individual-oriented idiographic approach to studying performance-related emotions. The IZOF model suggests some tentative answers to the questions asked at the beginning of the article.

First, unpleasant emotions as strong situational experiences are not always detrimental to athletic performance. Thus, anxiety, anger, and/or tension could be helpful in generating more energy and effort, and sometimes they could even compensate for a situational lack of needed resources. These emotions, if well channeled on task process, can postpone fatigue, sustain alertness, and keep the right focus. Coaches and athletes, as well as sport psychologists, begin to realize that coping with stress and using stress to enhance performance are possible.

Second, pleasant emotions are not always beneficial for performance, especially in sports requiring sustained focus, effort, and persistency for a relatively long time. A special problem in a high-achievement setting often presents a demotivational impact of excessive complacency following repeated success. Moreover, high self-confidence can sometimes lead to complacency or the underestimation of a "weak" opponent, resulting in insufficient alertness, lack of focus, and/or carelessness. These, in turn, can have a debilitating effect on performance, often leading to unexpected and season-ending injuries. In such cases, idiosyncratic experience is the best indicator of how athletes can perform up to their potential, either stress free or using competitive stress to their advantage.

The IZOF model provides a conceptual framework and methodological tools to describe, predict, assess, and explain situational emotional experiences related to performance, relatively stable patterns of these experiences, and meta-experiences. It is a clearly idiographic and person-oriented approach, with the main focus on

TABLE III  
Comparison of Group-Oriented and IZOF-Based Models

Feature	Group oriented	IZOF based
Analysis level	Global/Contextual	Situational
Target	Groups, samples	Individual athletes
Emphasis	Between-individuals	Within-individual
Content	Global/Narrow	Task and person relevant
Performance	Unrelated	Related to, induced by
Setting	Nonspecific	Sport specific
Time	Static, stable	Dynamic, changeable
Assessment	Group referenced	Self-referenced
Scales	Standardized	Individualized

within-individual comparisons that provide the basis for athlete-tailored and interindividual generalizations. Although the IZOF model is consistent with several existing orientations in psychology, it differs mainly in its emphasis on intraindividual dynamics in emotion–performance relationships (Table III). From the applied perspective, the major advantage of the individualized approach to studying emotion–performance relationships is its ability to capture and explain the unique findings that are often missed or ignored in group-oriented models. However, both individualized and group-oriented perspectives can be useful in applied sport psychology.

### See Also the Following Articles

Anxiety and Optimal Athletic Performance ■ Arousal in Sport ■ Assessment in Sport Psychology ■ Attention and Concentration Training in Sport ■ Cheating in Sport ■ Competition in Sport ■ Decision Making in Sport ■ Emotion ■ Fair Treatment and Discrimination in Sport ■ Goal Setting and Achievement Motivation in Sport ■ Group Dynamics in Sport ■ Intrinsic and Extrinsic Motivation in Sport ■ Performance Slumps in Sport: Prevention and Coping ■ Psychological Skills

Training in Sport ■ Self-Confidence in Athletes ■ Sport Psychology, Overview ■ Successful Athletic Careers

### Further Reading

- Hanin, Y. L. (Ed.). (2000). *Emotions in sport*. Champaign, IL: Human Kinetics Press.
- Hanin, Y. L. (2002). Individually optimal recovery in sports: An application of the IZOF model. In M. Kellmann (Ed.), *Enhancing recovery: Preventing underperformance in athletes* (pp. 199–217). Champaign, IL: Human Kinetics Press.
- Hanin, Y. L. (2003, February). Performance-related emotional states in sport: A qualitative analysis. *Forum: Qualitative Social Research*, 4(1). Available: [www.qualitative-research.net/fqs-texte/1-03/1-03hanin-e.htm](http://www.qualitative-research.net/fqs-texte/1-03/1-03hanin-e.htm).
- Hanin, Y. L., & Stambulova, N. B. (2002). Metaphoric description of performance states: An application of the IZOF model. *The Sport Psychologist*, 16, 396–415.
- Robazza, C., Pellizzari, M., & Hanin, Y. (in press). Emotion self-regulation and athletic performance: An application of the IZOF model. *Psychology of Sport and Exercise*.
- Ruiz, M. C., & Hanin, Y. L. (in press). Metaphoric description and individualized emotion profiling of performance related states in high-level karate athletes. *Journal of Applied Sport Psychology*, 16.



# Employment Discrimination

Mary L. Tenopyr

Bridgewater, New Jersey, USA

1. Basic Concepts in Employment Discrimination
  2. Early Law Relating to Employment Discrimination
  3. Role of Validity
  4. Discrimination in Education
  5. Validation
  6. Evidence of Validity
  7. Fairness and Bias
  8. Special Selection Situations
  9. Sexual Harassment
  10. Conclusion
- Further Reading

## GLOSSARY

**actuarial prediction** Predicting future behavior only on the basis on actual data; very much like predictions made in the insurance industry.

**Age Discrimination in Employment Act of 1967 (ADEA)** A U.S. federal law that prohibits discrimination against persons over 40 years of age.

**Americans with Disabilities Act of 1990 (ADA)** A broad federal law that prohibits discrimination against persons with disabilities in many areas, including employment.

**bias** A systematic effect; that is, the same degree of bias occurs for every member of the group of interest.

**Civil Rights Act of 1964** A comprehensive equal opportunity federal law that includes prohibition of employment discrimination on the basis of race, color, religion, or national origin.

**concurrent study** A criterion-related study of employment methods that involves obtaining predictor and criterion data at approximately the same time.

**criterion-related validity** The relationship between test results and a criterion that is a measure of an outcome of legitimate interest to an employer.

**differential item functioning (DIF)** The possibility that some items on an employment test function differently for distinct groups; it is difficult to conduct proper studies in this area, but it has become apparent that DIF is not found very often.

**disparate impact theory** The requirement that an employer justify the employment method, even if it seems neutral, if the employment method has a disparate impact on a legally defined group.

**disparate treatment theory** Intentional discrimination not necessarily involving tests or employment methods.

**Equal Employment Opportunity Commission (EEOC)** A U.S. enforcement agency created by the Civil Rights Act of 1964.

**fairness** A social term; there are various conceptions of what is fair given that what appears to be fair to one person might not be seen as fair by another person.

**flagging the score** A practice (considered controversial) of communicating to various members of management whether an employment test was administered under specially accommodating conditions.

**holistic assessment** Essentially a method for predicting future behavior on the basis of judgment given assorted data of several kinds; it is akin to clinical prediction by psychologists.

**internal structure** Any of the relationships among items or scales comprising a test; it can affect reliability or replicability of test results.

**predictive bias** An effect that occurs when a group of interest does significantly better or worse than would be predicted from scores for the whole group of applicants; the groups

of interest in this situation are differentiated on some characteristics not relevant to criterion performance.

**predictive study** A criterion-related study of employment methods that yields an indication of how accurately test results can predict criterion scores that are obtained at a later time.

**race norming** Selecting employees in terms of their standing in their own group and not on their standing in the whole job applicant population.

**test** Any procedure used in making employment decisions; this broad definition is frequently used in professional standards and guidelines, and it reflects the fact that employment discrimination is possible over a wide range of employment methods and not just paper-and-pencil testing.

**validity** The accuracy of inferences drawn from test results.

**validity generalization** The extent to which one can generalize from results of analyses and apply them in a particular employment context; with due care, a professional wishing to use tests can review the published validity generalization evidence and make an informed decision about what selection procedures to use in a given situation.

The selection of employees, whether in terms of hire, promotion, or other personnel actions, is a matter of concern relative to civil rights. Issues regarding the application of psychology in making selection decisions have become a major focus of legal activity. Employment discrimination on the basis of membership in some specific groups is prohibited by law. Legal challenges are frequently directed toward the methods used in employee selection. In particular, certain types of psychological tests and their use have come under close legal scrutiny. Concepts that have been associated with testing, such as validity and reliability, have been incorporated into various government guidelines and regulations that have, in turn, greatly influenced the development of laws relating to civil rights. Testing concepts and their application have been subject to considerable debate, research, and (in some cases) redefinition. Bias in selection is now better understood, and associated concepts have been evaluated thoroughly. It is clear that there are different psychological considerations associated with the selection and treatment of different groups. Despite advances in knowledge relative to employee selection, there are still research needs related to selection of some groups. Nevertheless, the large amount of psychological study that has been devoted to understanding employment

discrimination and its many aspects has added greatly to the database of psychological knowledge.

## 1. BASIC CONCEPTS IN EMPLOYMENT DISCRIMINATION

There are many situations in which there are fewer desirable rewards than there are seekers of those rewards. For example, except in unusual circumstances, there are fewer job openings than there are candidates for the jobs. Thus, there must be a means of selecting from among the job seekers. An analogy to a gatekeeping function may be drawn; in this case, the employment process serves as a gate, and the gatekeeping involves the application of psychology. In particular, the psychology of individual differences and its associated procedures have been widely applied in making employment decisions.

No two adult individuals are exactly alike, and each can be expected to behave in unique ways in many situations. Even identical twins, depending on the circumstances, can show noticeable differences in behavior. Within every definable group in society, there are individual differences, and often there is a wide distribution of these differences in, for example, scores on a test.

In addition, relative to some psychological processes and achievements, there are consistent differences among groups on average. Many of the gatekeeping functions in employment involve procedures that indicate average differences among groups. The term "average" is important because it represents a situation far less dramatic in terms of discrimination than many believe. For example, because of individual differences, there will generally be a number of different scores within each definable group. The distributions of scores for the separate groups will probably overlap. That is, given two groups, some members of the group that has lower scores on average will score higher than the group that has higher scores on average. Obviously, when there are average differences among groups on a test or another selection procedure, there is the potential for discrimination on the basis of group membership. Discrimination based on group has undoubtedly been a feature of every employing society in the world. So long as there are discernable group differences—and there always will be some differences with regard to some aspects of the employment process—there is the possibility of discrimination.

## 2. EARLY LAW RELATING TO EMPLOYMENT DISCRIMINATION

In many ways, the laws regarding employment discrimination started with a relatively narrow focus in the United States. However, during recent years, new legal developments have resulted in the scope of the issues being increased, and new aspects of psychology have come to the forefront. For example, new research has been done relative to the employment of the aging and of persons with disabilities, and the results of that research have been widely applied.

The abolishment of slavery by the 13th Amendment to the U.S. Constitution in 1865 and the Reconstruction Era after the Civil War brought forth the first legal approaches to employment discrimination. Thus, race was the focus of early laws in this area. Unfortunately, equal treatment of Black Americans in most aspects of life really did not receive much impetus until the 20th century. Significant early events of this era were President Franklin Roosevelt's 1941 order to end discrimination in the defense industry, President Harry Truman's 1948 order to desegregate the military services, and President Dwight Eisenhower's 1957 order to send troops to enforce desegregation of a high school in Arkansas. Thus, the stage was set for the emergence of the modern civil rights movement that involved a major march on the U.S. capital in 1963, numerous other marches and demonstrations, and a series of major urban confrontations.

These events greatly influenced congressional action on the Civil Rights Act of 1964, a comprehensive equal opportunity statute that included prohibition of employment discrimination on the basis of race, color, religion, or national origin. Many of the aspects of this broad law and subsequent amendments and court interpretations have influenced the employment practices of all major employers into the 21st century. The 1964 act had profound effects on the direction and focus of applied psychology in the workplace. This came about mainly due to an amendment prompted by a then recent situation in Illinois. A Black job applicant had filed a complaint with the state's fair employment commission because he was denied employment as a result of failing a cognitive ability test. A commissioner ordered the test replaced by a more culturally balanced one. This case received considerable national attention and was the impetus for considerable congressional debate, which ended in the eventual passage of an amendment that, on the surface, appeared to protect employers' use of ability tests in employment.

This amendment stated that it is not unlawful for an employer to give or act on the results of any professional developed ability test, provided that the test is not designed, intended, or used to discriminate. Mainly as a result of this amendment and other developments in the civil rights area, there was an extraordinary growth of interest in differences in group cognitive ability, especially in applied areas. Academic journals were flooded with research articles, and various theories and mathematical concepts of discrimination were debated. Furthermore, new studies related to group differences were extended beyond those relating to cognitive abilities. For example, as women moved into physically demanding jobs, studies of sex differences in various job-related physical aspects multiplied. Because the Civil Rights Act of 1964 specifically designated testing as a matter of concern, various agencies at both the federal and local levels began to promulgate guidelines on employee selection. All of these emphasized testing, and all other employee selection procedures, such as the interview, were defined as tests. The position taken by the U.S. Equal Employment Opportunity Commission (EEOC), an enforcement agency created by the 1964 act, was that a "professionally developed ability test" is one that is "professionally developed and used." This definition, which implied that validity (i.e., the accuracy of inferences drawn from test results) is the primary basis for professional use, appeared in the brief EEOC guidelines issued in 1964. In 1970, the agency issued detailed prescriptions for studying and evaluating validity. Furthermore, professional organizations began to issue their own guidelines on employee selection, and various professional standards on testing were modified.

## 3. ROLE OF VALIDITY

Validity, its definition, methods for conducting validation studies, and methods for assessing possible discrimination have become increasingly complex over the years. Testing standards have been revised not only to respond to technical advancements but also to address new issues associated with testing of groups that became the focus of civil rights acts passed after 1964. Nevertheless, validity remains a unifying theme in many of the laws involving possible discrimination in employment. Many local laws, which are even broader in group coverage than are federal laws, incorporate validity concepts.

The emphasis on validity in the 1970 EEOC guidelines was codified through the landmark 1971 U.S. Supreme Court decision in *Griggs v. Duke Power*

*Company*. In this case, the court ruled that if a test tends to exclude Blacks, the employer has the obligation to demonstrate that the test is manifestly related to the requirements of the job. Thus, business necessity could justify employment procedures. Also, the court addressed the question of relative qualifications and how high employment standards could be set. The novel feature of the *Griggs* decision was the introduction of disparate impact theory. Under this theory, if an employment method, although neutral of its face, has a disparate impact on a legally defined group, the employer must justify the employment method. Intention to discriminate is not a necessary element of a disparate impact case. Intentional discrimination is covered under a separate theory called disparate treatment theory. The emphasis on validity was continued in various government pronouncements at both federal and local levels. The most influential of these was the Uniform Guidelines on Employee Selection Procedures issued by several federal agencies in 1978.

The complicated further development of case law and legislation in the area of employment discrimination resulted in change to the conduct of a case brought under disparate impact theory. The Civil Rights Act of 1991 delineated three steps in a disparate impact case. First, the plaintiff must prove that the use of the selection procedure has a disparate impact on his or her legally defined group. The burden of proof then shifts to the defendant, who must prove that the procedure is job related (i.e., valid). The plaintiff then has the option of showing that a specific equally valid selection procedure not having the same degree of adverse impact could have been used. The 1991 act incorporated case law that eliminated an onerous provision of the 1970 EEOC guidelines requiring that an employer using a selection procedure that had a disparate impact must prove that there was no suitable alternative. In formal logic, this is known as “proving the negative”—which cannot be done.

#### 4. DISCRIMINATION IN EDUCATION

Constitutional rights apply to both the education and employment venues. However, many of the laws relating to discrimination in the field of education are not the same as those governing employment. Nevertheless, many concepts developed in law relevant to education have been applied in employment contexts. For example, “diversity” as a suitable employment goal evolved

from a U.S. Supreme Court decision in 1978, when the issue was admission to medical school. In a subsequent major and narrowly tailored decision on university admissions in 2003, the court allowed race and other factors to be taken into account in university admissions. The reason why the court allowed for this action is to achieve student body diversity. This is to be done on a time-limited basis, and no fixed quotas may be set. In addition, weighting techniques, such as point systems that favor some groups over others, are prohibited. The court supported a holistic system involving a number of methods such as testing, essay writing, and interviewing. This action brings into focus the major psychological question of the validity of clinical prediction as opposed to actuarial prediction. The holistic assessment endorsed by the court is essentially a method for predicting future behavior on the basis of judgment given assorted data of several kinds and is akin to clinical prediction. Actuarial prediction, very much like predictions made in the insurance industry, involves predicting future behavior only on the basis on actual data. For example, the quality of work that would be done by a job applicant, if hired, is predicted from actual test scores.

#### 5. VALIDATION

In an employment situation, the inferences made through the application of employee selection procedures are predictive in nature regardless of the nature and format of the methods used in making the personnel decisions. Throughout this discussion, the term “test” is used to cover any procedure used in making employment decisions. This broad definition of test is frequently used in professional standards and guidelines and reflects the fact that employment discrimination is possible over a wide range of employment methods and not just paper-and-pencil testing.

The term “validity” has had a number of definitions over the years. Today, validity is not seen as a property of the test. Instead, it refers to the accuracy of inferences made on the basis of test scores. For example, if it is inferred that persons with low test scores will not do the job as well as will persons with higher test scores, and this inference is confirmed as a result of proper study, one may say that the inference is accurate and, hence, valid.

Validity is not an all-or-none concept; there are varying degrees of validity. Validation is the process of conducting a study to assess the degree of validity. In this case, the term “validation” is not used in its other sense as an approval process.

The legal use of the term “job related,” as opposed to “valid,” clearly and appropriately narrows the scope of any validation evidence to that related to the workplace activities of the concerned persons. It should be noted that activities other than actual job performance can be the focus of the test. For example, tenure, training grades, accidents, and absences may be legitimate concerns of the employer, who may use tests designed to predict them.

Furthermore, it is important to recognize that validity in psychology, as in many fields, is never perfect. There are always errors in prediction. However, the lack of perfect prediction need not necessarily deter an employer from using a test. How useful a test will be depends very much on the particular situation. Under some conditions, a test with relatively low validity can be quite useful to the employer. Certain factors, such as labor market conditions and job difficulty, must be taken into account. Professional judgment, as well as available data, should inform decisions on test use.

## 6. EVIDENCE OF VALIDITY

There are five generally recognized sources of evidence. Sources for evidence of validity are based on (a) content, (b) response processes, (c) internal structure of the test, (d) relationships of test scores with other variables, and (e) consequences of test use.

### 6.1. Evidence Based on Content

Content validation involves an initial thorough analysis of the work in terms of work behaviors and activities and/or knowledge, skills, abilities, or other personal characteristics related to job demands. The work analysis may be as detailed as need be, depending on the situation. Again, professional judgment is required. A work analysis for a parachute packer job should be rigorously detailed. For other jobs that are less rigid in terms of critical requirements, a broader analysis might be appropriate.

A selection procedure developed on the basis of content need not necessarily cover the whole job. If it is critical that every task involved in a job be done precisely, every task should be covered in minute detail. However, for many jobs, only some of the job duties are of critical interest, and only these parts may be covered by the test. For example, a person doing word processing may also answer the telephone on occasion. The employer may test on word processing

skills and not test on telephone answering skills. Measuring abilities of various sorts presents particular problems. In general, content validation is appropriate only for relatively narrow abilities. If one wants to measure broad abilities such as leadership, content should not be the primary source of evidence.

Developing a test on the basis of content is essentially a sampling process in which a sample of the work or the worker characteristics of interest is taken. Evaluation of the adequacy of the sampling is difficult. Validity is established mainly on scrutiny of the processes used in sampling from the content area of interest and the procedures used in developing the test. Professional judgment is usually the only feasible way in which to assess the adequacy of the evidence. However, some psychologists contend that evidence can be gathered through judgments regarding job tasks or the skills, knowledge, abilities, and other characteristics required for the job and their linkage to the content of the test.

Many of the disputes involving employment discrimination have arisen in situations involving validity evidence based on content. Because of the large judgmental component involved, there is opportunity for both employment discrimination and disputes about the adequacy of the sampling process used and the product developed. Also, many of the tests used in the public sector are based on content, and much of the case law regarding employment discrimination has been generated in the public sector. In particular, selection for public safety jobs that are sought by thousands of applicants, such as police work and fire-fighting, is a source of contention. Most public sector merit systems have legal rules and rigid requirements for employee selection. For example, job applicants are put in rank order in terms of standing on tests, and the final selection must be from among the top three candidates. Furthermore, lists of candidates may be used to fill openings for months or even years, and applicants might have a long wait before a new test is developed and they have the opportunity to be tested again. Private employers generally have more options in the use of selection procedures and, consequently, can make adjustments to the selection process when there are questions about employment discrimination.

There are other issues related to applied psychology in the use of content-based evidence. One of these is whether it is appropriate to rank order candidates on the basis of scores on a content-based test. It appears that most industrial psychologists believe that such ranking is permissible. Another related issue is how to set “cut” (i.e., passing) scores on such tests. There

are still disagreements about how high passing scores should be set. The main issue is whether they should be consistent with the average proficiency of the workforce or whether an employer trying to improve the company's workforce may set higher standards. Another is whether changing test formats, such as moving from a paper-and-pencil test to a more complicated and costly procedure involving pictures or exercises that simulate job duties, will reduce disparate impact on various groups. The Civil Rights Act of 1991 allows cost to be a consideration in determining whether specific alternative selection procedures should be adopted.

A specific source of contention is the application of content-based physical ability tests to both genders. The differences in performance between the genders on many physical ability tests is marked on average, and issues related to ranking, passing scores, and the possibility of alternative selection procedures becomes magnified. For example, a content-based test for firefighters might involve activities such as dragging a firehose up a flight of stairs. There have been questions about whether simpler physical tests, including those involving activities such as push-ups, can be justified on the basis of content. Persons with certain types of physical disabilities have raised similar questions about such tests.

Some employers have representatives of various groups that are likely to be subject to discrimination review the test content to detect possible discriminatory items or those that might be offensive to particular groups. There are questions about the effectiveness of this procedure, and more research concerning the value of this method in preventing discrimination is needed.

### **6.2. Evidence Based on Response Processes**

Where employment is concerned, evidence based on response processes (e.g., how long an examinee requires to respond to a particular type of item), is not so relevant as most of the other types of evidence of validity. Response processes do have some value in determining how a test is functioning and what abilities the applicant is using in taking some types of tests.

These processes have been assessed by various methods such as having an examinee talk about what mental techniques he or she is using while trying to solve an arithmetic problem. Physical measurements based on recording eye movement are often used, as are observations by judges. It has been suggested that

such monitoring of response production may eventually be able to discern differences in the ways in which different groups respond to the same test material. However, relevant literature on research in employment settings is lacking.

### **6.3. Evidence Based on Internal Structure**

The third type of evidence is based on the internal structure of the test. An example of internal structure would be the extent to which test items are in the same format such as solving analogies problems. Internal structure can affect reliability or replicability of test results. Such reliability is important in supporting validity. In general, validity is limited by lack of reliability of test results. Reliability may be estimated by various methods. The choice of exact estimation procedure depends on a number of factors. In some respects, a finding of validity is, to some extent, evidence of reliability.

Relative to discrimination, the possibility that some items function differently for distinct groups is often studied. This effect is called differential item functioning (DIF). It is difficult to conduct proper studies in this area, but it has become apparent that DIF is not found very often. When found, reasons for such differential functioning are not usually easily identified.

### **6.4. Evidence Based on Relationships with Other Variables**

The fourth type of evidence, relationships of test scores with other variables, has been the major source of statistically based definitions of discrimination and of much confirmable information in psychology about group differences. This type of evidence will undoubtedly be a mainstay in studying employment discrimination and related topics for some time to come.

In an employment situation, the relationship of most interest is the relationship between test results and a criterion that is a measure of an outcome of legitimate interest to the employer. This relationship is called criterion-related validity, and the generic predictive model of validity is merely a generalization of the criterion-related model. Despite the fact that criteria can take many forms, supervisors' ratings appear to be the most commonly used criteria. The design of a criterion-related study can involve a number of specific procedures, but these can be subsumed under the general headings of "predictive" and "concurrent" studies.

A predictive study yields an indication of how accurately test results can predict criterion scores that are obtained at a later time. For example, job applicants may be tested and the results might not be used in hiring. Then, at a later date, job performance or other job outcome data that form the criterion may be collected. A concurrent study involves obtaining predictor and criterion data at approximately the same time. Typically, present employees are used in these studies. In the area of employment, it is very difficult to conduct predictive studies and complete them in a reasonable period of time. Consequently, much of the validation relevant to employee selection is based on concurrent designs. For cognitive ability, it appears that validation results from predictive and concurrent studies do not vary much from each other. More studies need to be done relative to other types of tests.

Relative to discrimination, one of the major findings resulting from the application of the criterion-related model is that validities in the cognitive ability area do not differ greatly between Caucasians and African Americans. As the United States becomes more multicultural, there is a need for research with newly emergent racial and cultural groups. Also, it is necessary to determine the extent of group differences relative to tests that reflect characteristics other than cognitive ability.

Differences between gender groups are a matter of concern. However, they are difficult to study because during test development, particularly of the most widely published cognitive ability tests, efforts have been made to reduce some gender differences. Still, there are some well-documented findings relative to possible discrimination. Women are at a disadvantage when selection procedures involve some spatial and mechanical abilities. Also with respect to cognitive ability tests, women generally have a narrower range of scores than do men. Some of the practical effects of this phenomenon are that there are more males than females classified as mentally retarded, and there are more males than females who qualify for prestigious scholarships and other honors that are awarded on the basis of high-level cognitive ability.

### **6.5. Evidence Based on Consequences**

Within a broad social context, concern about the consequences of testing, particularly that some types of testing can result in unequal distribution of rewards to various groups, has led some major researchers to

define validity partially in terms of consequences of test use. In an employment context, consequences may influence decisions regarding the use of particular selection practices. However, when validity is defined as it has been historically, consequences alone do not form acceptable evidence of validity. Socially undesirable consequences in employment should prompt a thorough investigation of possible sources of problems and corrective actions available.

### **6.6. Other Evidence-Related Issues**

With the possibility of pressure to substitute clinical prediction for actuarial prediction in employment operations, there is a real need to conduct further studies on the decision processes involved in clinical judgments as they relate to employment. Questions about validity and effects on groups are a matter of concern. Past findings clearly demonstrate that the actuarial prediction process produces valid results to the same extent as does clinical prediction. In many cases, actuarial prediction is associated with greater validity. Among the major problems that have plagued clinical prediction in the past is the basing of judgments on methods that have not been shown to produce valid results. Despite the fact that use of multiple methods forms the basis for clinical prediction, the possibility that some of these methods do not add to the accuracy of such prediction should be considered. This is particularly important when tests that are not designed for use in employment inform the clinical judgment process used in selecting new hires from among job candidates. Another problem is that judges might not weigh the available information optimally. Too much weight can be attached to the results of methods that yield marginal validity.

An extraordinary amount of evidence has been accumulated regarding the validity that can be expected from the application of various employee selection methods. A mathematical technique known as meta-analysis has afforded the opportunity to analyze the results of accumulated research so that general conclusions relative to various research findings can be drawn. In employment psychology, meta-analysis has been applied to the accumulated validity studies done over a period of many decades. The extent to which one can generalize from results of these analyses and apply them in a particular employment context is known as "validity generalization." With due care, a professional wishing to use tests can review the published validity generalization evidence and make an

informed decision about what selection procedures to use in a given situation. The published validity generalization research results have been based on a substantial amount of data in some areas, including those involved with cognitive ability; however, in some other areas, the evidence is still accumulating. Nevertheless, it is clear that most professionals consider the validity generalization data first before they make decisions about the use of selection procedures.

## **7. FAIRNESS AND BIAS**

Although the term “fairness” was used in early federal and local guidelines on testing, it has become customary in applied psychology to distinguish between “fairness” and “bias.” The former is a social term; what appears to be fair to one person might not be seen as fair by another person. Thus, there are various conceptions of what is fair. Conversely, latter term has a specific meaning in employment psychology. This definition is based on results of research with the criterion-related model and is generally accepted in applied psychology.

First, fairness may be viewed as equitable treatment in the employment process. For example, all applicants should have the same selection procedures applied to them under the same conditions. Second, fairness may be conceptualized as all applicants having the same opportunity to learn material covered by employment tests. With a few exceptions, this notion, conceived in the context of educational achievement testing, does not apply to employee selection. Third, fairness may be viewed in terms of whether the employment process provides equal outcomes for different groups. In other words, the same proportion of each group involved in the process should be selected. In the interest of affirmative action, various techniques have been applied in attempts to achieve equal outcomes. One method is to add points to the scores of members of selected groups. Another method is race norming, that is, selecting employees in terms of their standing in their own group and not on their standing in the whole job applicant population. In this case, a person from a group that has low average scores would be more likely to be selected than would a person with the same test score if the latter person belongs to a group with higher average scores. Other mathematically complicated methods have been used with the objective of reducing group differences in selection. For example, a method known as the sliding band, which seeks to lessen the disparate impact associated with ranking systems in the public

sector, has been developed. Considerable controversy has arisen regarding the application of these and other methods developed in the context of affirmative action. Fourth, fairness may be seen as a lack of “predictive bias,” which is defined technically as occurring when a group of interest does significantly better or worse than would be predicted from scores for the whole group of applicants. The groups of interest in this situation are differentiated on some characteristics not relevant to criterion performance. Examples of such characteristics might be race and gender. Bias is a systematic effect; that is, the same degree of bias occurs for every member of the group of interest. So far as African Americans and Caucasians are concerned, when cognitive ability tests are used in employment, bias is seldom found. This situation might not be seen when other groups are compared. The rapidly changing composition of the labor force, coupled with the need for large samples of research participants, renders the proper conduct of studies of possible bias very difficult except in military organizations and other large employers.

A number of other issues are related to more general questions of fairness and bias. It is well known that many groups are judged in terms of stereotypes. Gender, ethnic, racial, and other stereotypes can affect employment decisions. Considerable research in the area of social psychology has been devoted to stereotypical thinking as it applies to employment. Some controversy still prevails regarding the application of this research.

As a result of women being paid less than men, there are both federal and (some) local laws requiring equal pay for equal work that involves the same duties. To determine whether the work done is indeed equal requires the work analysis methods developed by applied psychologists. Legal issues concerning pay equity for the genders continue. There have been major efforts to extend the laws to require equal pay for jobs of comparable worth. The determination of comparable worth for jobs with different job duties is extremely challenging for industrial psychologists.

## **8. SPECIAL SELECTION SITUATIONS**

### **8.1. Aging**

Employment issues involving possible age discrimination are covered by special laws and regulations both at the federal level and in many local jurisdictions. The Age Discrimination in Employment Act of 1967 (ADEA) prohibits discrimination against persons over

40 years of age. As the population ages and people extend their working lives past what was once considered normal retirement age, issues relevant to employment discrimination against the aging have increased both in scope and in intensity. The use of measures of cognitive ability is a particularly sensitive issue. It is now known that many of the major abilities decline with age more slowly than had once been believed. However, the possible onset of debilitating diseases, such as Alzheimer's disease, must be considered in some cases. Also, there may be interactions among mental functioning, physical condition, and other variables (e.g., gender).

Test results might not have the same meaning for the aging as they do for other groups protected by law. At the ages when individuals over 40 years old are likely to seek employment, major age-related deficits in mental functioning are relatively unlikely. When there are deficits, the more likely case is that the individuals involved have had long histories of obtaining low scores on measures relating to cognitive functioning. Another possibility to be considered is that discrimination against older persons may be associated with other characteristics such as gender and ethnic origins.

It appears that it would be unrealistic to strictly apply all of the law that has been developed under the Civil Rights Act of 1964 and its progeny to the aging population. The application of disparate impact theory is particularly controversial. The issues relative to mental functioning of older persons are more complicated than most of those involving younger people.

Considerable legal attention has been directed toward the various processes involved in terminating the employment of older persons. Any techniques used in making termination decisions, such as performance appraisals, should be treated as tests and subjected to the same professional scrutiny. The economic issues involved in retirement and other forms of termination have been the focus of considerable legal attention. Among the major situations that involve the potential for age discrimination is the provision of health care benefits to retirees. An understanding of the needs for mental health care for the elderly can inform employer decisions regarding health care benefits for both retirees and older persons still employed. Psychologists can provide useful input in this area.

## 8.2. Individuals with Disabilities

The Americans with Disabilities Act of 1990 (ADA) has received considerable attention by applied

psychologists because the statute applies to both mental and physical disabilities. This broad law prohibits discrimination against persons with disabilities in many areas, including employment. This law involves a number of ambiguities and, thus, is subject to continuing legal interpretations. Some of the requirements of the law remain to be clarified. It appears that the courts are progressively limiting the scope of ADA. Furthermore, solid research bases to support actions under this statute are unavailable. To conduct proper research in this area, one needs a substantial number of research participants with the same disability and the same level of symptoms. Such large groups of people are generally unavailable. Basically, ADA recognizes the individual nature of disabilities and the need to treat each person with a disability in a different way. ADA is basically an extension of earlier acts dealing with rehabilitation and employment of veterans.

Under ADA, disability is defined as a physical or mental impairment that substantially limits at least one major life activity. According to EEOC guidance, among the activities covered are learning, working, thinking, concentrating, and interacting with others. Specifically excluded are certain conditions such as current illegal drug use, a number of conditions dealing with sexual behavior and gender identification, certain compulsive disorders such as kleptomania, and impairments that are expected to be temporary. In addition, employers are not expected to hire individuals who may be a danger to themselves or others. Questions about diagnoses clearly involve the application of psychological knowledge about mental impairments. In addition, making predictions about future behavior is necessary.

The Americans with Disabilities Act of 1990 applies only to qualified persons with disabilities. These individuals must be able, with or without reasonable accommodations, to perform the essential functions of the job. Determining the essential functions of the job may involve an adaptation of the work analysis methods used in conventional employee selection situations. Such analysis is not legally required, but many employers conduct it. During prehire questioning, such as might be done on an application blank or during an interview, the employer cannot ask questions that would require the applicant to indicate the existence of a disability. However, the employer can give a psychological examination unless the test is used for medical reasons, in other words, to determine whether the individual is mentally impaired or otherwise mentally disabled. Diagnostic psychological tests and related procedures may be used in a medical

examination, which can be given only after a job offer has been made, with the same questions being asked of all applicants for the same job category.

In testing situations before a job offer is made, the job candidate may ask for an accommodation relative to the medium of testing, test format, testing conditions, or testing time. For example, a person with impaired vision may ask for a test booklet with enlarged print. The most commonly requested accommodation is extra time to take the test. Accommodations are to be done on an individual basis. Nevertheless, psychologists try to be reasonably consistent in devising accommodations for similar types of impairments. In many cases, psychologists collaborate with physicians and other professionals in developing accommodations in testing and other aspects of the employment process.

Most employers ask job candidates for documentation regarding disabilities that are not obvious when either tests or job accommodations are requested. Documentation regarding mental disabilities is usually carefully scrutinized. Many employers have internal requirements for documentation. For example, a family practice physician's diagnosis of a learning disability might not be accepted. Accommodations in working conditions on the job are often developed with input from several types of specialists. Some employers have standing committees to make decisions about job accommodations. Most job accommodations are not particularly expensive. In any case, the employers are not expected to undergo undue hardship. In general, legal interpretations of what constitutes hardship have been quite narrow.

The issue of indicating whether a test was administered under accommodating conditions should be recorded, and whether this information should be communicated to various members of management is quite controversial. This is called "flagging the score." At a minimum, a psychologist usually makes some record of the accommodation and maintains it in a filing system that is separate from other test score files. Various other legal and practical issues are related to the employment of persons with disabilities. In particular, violence in the workplace might be related to the failure of the employer to recognize that a person claiming a disability was a danger to himself or herself and others. Other aspects of workplace security, such as developing a crisis management plan in the event of disaster (natural or otherwise), may involve paying special attention to the needs of persons with disabilities. Of obvious importance are plans for the evacuation of such individuals. Psychologists can provide

input to a number of other aspects of such plans and can advise on training and practice needs.

The role of disabilities with respect to a number of other laws may require the attention of psychologists. Included among such laws are the federal Family and Medical Leave Act and the various workers' compensation and unemployment compensation laws, all of which may have some features that can lead to discrimination against persons with mental disabilities.

### **8.3. Linguistic Minorities**

Discrimination on the basis of national origin may occur due to the language in which a test is written or administered. Linguistic considerations in employment are not straightforward. One of the first distinctions to be made is that between a language and a dialect. Some have contended that tests should reflect dialects shared by sizable groups of people. As the United States becomes more multicultural, attending to all of the various dialects, whether regional or group associated, would be an onerous task for even the largest employers. If job proficiency in the language of the workplace is required, employment processes usually use the language needed to communicate on the job. A distinction should be made between oral communication and written communication. For example, a given individual's lifestyle may involve proficiency in writing but not in speaking. This could occur in cases where English was the language of the schools attended, but some other language was spoken in the home. Others may speak in a mixture of two languages, and others with backgrounds in two languages may have difficulty in using both. Any translations of tests should be done by persons qualified not only in the languages but also in psychometrics. There are a number of procedures necessary to ensure that the original test and the translated version yield equivalent results.

## **9. SEXUAL HARASSMENT**

A major issue in the area of employment discrimination is sexual harassment. This subject has been a major source of litigation since the early 1970s. Sexual harassment can be physical, verbal, or visual. It is not always directed toward women, but in any event it is always unwelcome. Legally, the conduct should be, at a minimum, pervasive or severe enough to create a hostile or abusive work environment and be

such that a reasonable person would find it to be offensive. In legal terms, sexual harassment unreasonably interferes with work performance, is physically threatening or humiliating, and involves more than isolated utterances. Voluntariness is not a defense against a sexual harassment charge; a person might submit to sexual advances due to fear of an unfavorable tangible job action such as termination.

The U.S. Supreme Court defined two types of sexual harassment. The first of these is “quid pro quo,” where sexual involvement is explicitly or implicitly made a condition of employment or where submission or rejection of harassing conduct is used as a basis for employment decisions about the employee. The second type of harassment is “hostile work environment,” where the conduct has the purpose or effect of interfering with the employee’s work performance.

More recently, the courts have formulated standards for assessing employer liability in sexual harassment cases. A tangible employee action constitutes a significant change in employee status, such as hiring, firing, failing to promote, or reassigning with significantly different responsibilities, or a decision causing a significant change in benefits. If the alleged harasser is a supervisor and there is a tangible employee action, the case is considered quid pro quo and the employer is potentially vicariously liable. If there is no such action but the harassment is severe or pervasive, the employer may present an affirmative defense by showing that reasonable care was exercised to prevent sexually harassing behavior and to ensure that such behavior was corrected promptly and that the employee failed to take advantage of preventive or corrective opportunities or to otherwise avoid harm. There are variations of these rules for alleged harassment by coworkers or outsiders such as vendors associated with the employer.

Many employers have established firm policies to prevent and correct such harassment. At a minimum, these policies involve accessible and effective complaint processes, training for all employees, and detailed investigative procedures. The sensitivities involved in carrying out these policies and the possibility of sensational disclosures require the application of psychological knowledge. During the mid-1990s, Congress extended the so-called federal rape shield law to civil proceedings. Thus, a woman’s current and past sexual behavior is admissible in court only if the defendant can show that its evidentiary value outweighs any prejudicial effect on the plaintiff. Thus, sexual harassment cases, which are civil cases, may

involve all of the applications of psychology now applied in rape cases. As would be expected, there are a number of legal gray areas in sexual harassment cases. Perhaps this is one of the reasons why the courts and legislatures have been reluctant to extend current laws beyond relationships between persons of the opposite sex. Even disgusting and gross behavior, as has occurred in same-sex horseplay at work, has not been considered legally as sexual harassment. Civil rights laws are not being considered to be general civility codes.

## 10. CONCLUSION

Although there are other pertinent connections between psychology and employment discrimination, it is clear that psychological knowledge and methods can have a major role in preventing and understanding such discrimination. In addition, knowledge in other psychological areas that are not generally considered to be applied can inform activities involved in possible employment discrimination. Applied psychology has become more highly developed in many respects due to the research and debates spurred by the civil rights movement and the emphasis on employment discrimination. The database of psychological knowledge has been greatly enhanced by these efforts in applied psychology.

### See Also the Following Articles

Prejudice and Discrimination ■ Sexual Harassment

### Further Reading

- Age Discrimination in Employment Act of 1967, as amended, 29 U.S.C. 621.
- American Educational Research Association, American Psychological Association, & National Council in Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- American with Disabilities Act, 42 U.S.C. 12101 (1990).
- Civil Rights Act of 1964, § et seq., as amended, 42 U.S.C. § 2000, et seq.
- Cleveland, J. N., Stockdale, M., & Murphy, K. R. (2000). *Men and women in organizations: Sex and gender issues at work*. Mahwah, NJ: Lawrence Erlbaum.
- Federal Glass Ceiling Commission. (1995). *Good for business: Making full use of fact-finding report of the Federal Glass Ceiling Commission*. Washington, DC: U.S. Department of Labor.

Griggs v. Duke Power Company, 401 U.S.C. 424 (1971).

Lockwood, N. (2003). *The aging workforce: The reality of the impact of older workers and eldercare in the workplace*. Alexandria, VA: Society for Human Resource Management.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green, OH: Author.

Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. 1607 (1978).



# Employment Interviewing

Deborah L. Whetzel

Work Skills First Inc., Glen Allen, Virginia, USA

Michael A. McDaniel

Virginia Commonwealth University, Richmond, Virginia, USA

1. Overview
  2. Psychometric Characteristics of the Interview
  3. Conclusion
- Further Reading

## GLOSSARY

**critterion-related validity** The correlation between a predictor (e.g., test, interview) and the criterion that it is designed to predict (e.g., job performance).

**critical incidents** Structured stories about good and poor job performance, including the situation in which the incident occurred, the behavior describing the performance, and the result of the incident.

**incremental validity** The validity of one predictor over and above that of another predictor.

**interrater reliability** An estimate of the reliability of a measure assessed by the correlation between ratings made by two raters independently.

**meta-analysis** A set of statistical techniques used to summarize research results across studies to draw generalizable conclusions.

**past behavior interview question** A question in which the interviewee is asked to describe his or her performance in a similar, but more specific, situation (e.g., "Describe a time when you had to give bad news to a client. What was the situation? What did you do? What was the result?").

**situational interview question** A question that poses a situation to the interviewee and asks how he or she would likely perform in that situation (e.g., "What would you do if . . .?").

This article is organized into two sections: overview and psychometric characteristics of the interview. The overview section provides a definition of interviews, a description of various levels of structure, and methods for enhancing the structure of the interview. It also provides examples of questions using the job of a civil aviation security inspector. The psychometric section describes research on reliability, validity, correlations with other constructs, subgroup differences, and incremental validity beyond general mental ability.

## 1. OVERVIEW

### 1.1. Definition

An interview involves interactions (face-to-face, telephone, or written) between an interviewer and an interviewee. An interview typically is initiated to achieve one or more objectives (e.g., obtaining data in response to a survey, disciplining employees, providing feedback as part of performance appraisal, career counseling, selecting employees from a pool of applicants) and occurs as part of a sequence of activities (e.g., reviewing resumes, interviewing a subset of candidates, and administering cognitive ability tests for the purpose of selecting applicants; reviewing training curricula and interviewing employees for the purpose of assessing training needs). The interview's objectives and the sequence of activities in which the interview takes place are known as the context of the interview. In this article, interviews are

described in the context of selection, that is, interviews used to select employees from a pool of applicants. In this setting, interviews typically focus on past, present, or future behavior, beliefs, opinions, attitudes, and/or convictions of applicants. Information about an individual's previous experience, education, and vocational aspirations reported during the interview, as well as behavior observed during the interview, is considered the content of the interview. Interviewers are required to interpret information and draw inferences about traits, attributes, attitudes, and skills that the applicants possess. These inferences often take the form of ratings on dimensions of personal attributes or rankings of applicants.

### 1.2. Levels of Structure

There are two broad types of employment interviews: structured and unstructured. Structured interviews have a fixed format or a fixed set of questions, and the questions are based on a job analysis. Conversely, unstructured interviews do not use a standard procedure for the administration of questions or scoring. In fact, questions may vary from applicant to applicant.

In 1992, Huffcutt reviewed the literature on interview structure and described four progressive levels of structure. Level 1 was the typical unstructured interview characterized by no constraints on the questions and a global evaluation of responses. Level 2 imposed limited constraints by specifying the topics to be covered by the questions and some degree of structure on response evaluation. Level 3 required the prespecification of questions, although applicants were not asked precisely the same questions because different interview forms were used or interviewers were allowed to choose among alternative questions and to probe responses to the specified questions. An example of this kind of interview is the behavioral patterned description interview as described by Janz in 1982. Responses were evaluated using some degree of structured response evaluation. Level 4 involved asking applicants precisely the same questions with no deviation or follow-up probes, and responses were scored according to benchmark answers. In 1980, an example was provided by Latham and colleagues, who discussed situational interviews.

Situational and past behavior are two primary types of structured interviews. Situational questions provide a situation to the interviewee and ask how he or she would likely perform in that situation (e.g., "What would you do if . . .?"). The situations resemble those in situational judgment tests (SJTs). Past behavior questions provide a broad situation for the interviewee.

The interviewee is asked to describe his or her performance in a similar, but more specific, situation (e.g., "Describe a time when you had to give bad news to a client. What was the situation? What did you do? What was the result?"). Both of these kinds of questions derive from critical incidents provided by subject matter experts (SMEs). Critical incidents are structured stories about good and poor job performance. The three-part structure of critical incidents involves having the SMEs respond to three questions: "What was the situation/task?," "What did you do?," and "What was the result?"

Interviews can be very appealing to organizations because questions can be created to tap cognitive and noncognitive skills as well as abilities. Interviews can focus on a range of applicant qualities such as communication skills, interpersonal skills, decision making, and substantive knowledge. Examples of situational and past behavior interviews are provided in [Tables I and II](#), respectively.

### 1.3. Methods for Enhancing the Structure of Interviews

In 1997, Campion and colleagues reviewed the use of structure in selection interviews and identified 15 components of structure that may enhance either the content of the interview or the evaluation process in the interview:

1. Base questions on a job analysis.
2. Ask the exact same questions of each candidate.
3. Limit prompting, follow-up questioning, and elaboration on questions.
4. Use better types of questions (e.g., situational, past behavior, background, job knowledge).
5. Use a longer interview or a larger number of questions.
6. Control ancillary information.
7. Do not allow questions from the candidate until after the interview.
8. Rate each answer or use multiple scales.
9. Use detailed anchored rating scales.
10. Take detailed notes.
11. Use multiple interviewers.
12. Use the same interviewer(s) across all candidates.
13. Do not discuss candidates or answers between interviews.
14. Provide extensive interviewing training.
15. Use statistical prediction rather than clinical prediction.

**TABLE I**  
**Situational Interview Example**

**Critical Incident**

During an air carrier assessment, an inspector noticed that the screeners at checkpoints appeared to be unsure about the proper use of metal detector equipment and procedures for screening passengers. The inspector interviewed all of the screeners on-site and discovered that the screeners on duty at the metal detectors had not received required training. The inspector immediately informed all air carriers represented at the checkpoint and requested that carriers implement additional screening procedures. The screeners who had not been trained were removed immediately.

**Situational Question**

You are conducting an air carrier assessment and you notice that several of the screeners at the metal detectors appear to be unsure of how to operate the equipment. What would you do?

**Response Anchors to Situational Question**

<i>Highly ineffective</i>			<i>Average</i>			<i>Highly effective</i>		
<ul style="list-style-type: none"> <li>• Have the inspector on the next shift train the screeners</li> <li>• Do nothing, assuming that the screeners will get better using the machinery with practice (e.g., learning by doing)</li> </ul>	<ul style="list-style-type: none"> <li>• Tell another inspector about the problem</li> <li>• Watch for multiple instances in which screeners appear to be unsure of how to operate the equipment to make sure that you are correct in your assessment</li> </ul>	<ul style="list-style-type: none"> <li>• Inform the air carriers being represented at the checkpoint</li> <li>• Have the untrained screeners removed and/or trained immediately</li> </ul>						
1	2	3	4	5	6	7	8	9

These components of structure are based on research literature and continue to be replicated. The more of these components that are incorporated into the interview, the more structured the interview becomes.

**2. PSYCHOMETRIC CHARACTERISTICS OF THE INTERVIEW**

**2.1. Reliability**

An interview is reliable to the extent that it is free of measurement error. Typically, reliabilities of interviews are assessed by interrater reliability, that is, the correlation of the interview ratings made by two independent raters. There are at least two meta-analytic reviews that investigated the reliability of interviews. In 1995, Conway and colleagues found the average reliability of structured interviews to be .67. In 1994, McDaniel and colleagues found mean reliabilities of .84 for job-related structured interviews (across 167 coefficients) and .68 for job-related unstructured interviews (across 20 coefficients).

**2.2. Validity**

The validity of the interview is typically assessed by correlating an interview score with a measure of job performance. This is one application of criterion-related validity. McDaniel and colleagues broke down interviews by content and found that situational interviews were superior to job-related and psychological interviews with respect to job performance (criterion-related validity coefficients of .50, .39, and .29, respectively). Structured interviews, regardless of content, were more valid (.44) than unstructured interviews (.33) for predicting job performance criteria. Individual interviews (.43) were more valid than board interviews (.32).

In 1995, Pulakos and Schmitt found that experience-based interview questions yield higher validities (.32) than situational questions (-.02). This result was replicated by Huffcutt, Weekley, and colleagues in 2001. Their behavioral description interviews (BDI) (i.e., experience-based interviews) were a reasonably strong predictor of leadership performance in training for Canadian military officers, whereas their total situational interview (SI) scores were not ( $r = .49$  vs  $.22$  after correcting for range restriction). This result was replicated again by Huffcutt and colleagues using a

**TABLE II**  
**Past Behavior Interview Example**

**Critical Incident**

A senior inspector wanted on-the-job training (OJT) for new inspectors in the office on the procedure used by the legal counsel to close a case or hold informal hearings. The senior inspector contacted the legal counsel's office and discussed the matter with several lawyers who handled security cases. The lawyers agreed to provide OJT for the new inspectors. Consequently, new inspectors now routinely attend informal conferences with legal counsel and receive firsthand instruction on how informal hearings are conducted and how to close a case.

**Past Behavior Question**

Tell me about a time when you needed to have people trained as soon as possible and you did not have the expertise to provide the training yourself. What did you do?

**Response Anchors to Past Behavior Question**

<i>Highly ineffective</i>			<i>Average</i>			<i>Highly effective</i>		
<ul style="list-style-type: none"> <li>• Have new inspectors “learn it on their own” on the job</li> <li>• Perform work on own that could be delegated if inspectors were trained</li> <li>• Give more assignments to top performers rather than train new Inspectors</li> </ul>			<ul style="list-style-type: none"> <li>• Read a book and learn the subject area yourself</li> <li>• Identify and suggest training needs for new inspectors in a variety of areas of expertise</li> <li>• Coach new inspectors as they learn new things</li> </ul>			<ul style="list-style-type: none"> <li>• Identify individuals who are knowledgeable in the subject area and ask them to teach the new inspectors</li> <li>• Provide developmental opportunities to new inspectors who wish to learn a new function</li> <li>• Create opportunities for inspectors to try things they have not done before</li> </ul>		
1	2	3	4	5	6	7	8	9

sample of district managers. The BDI total scores correlated with the performance evaluation ( $r = .31$ ), whereas the SI total scores did not ( $r = .02$ ).

**2.3. Correlations with Other Constructs**

In 2001, Huffcutt, Conway, and colleagues developed a taxonomy of seven types of constructs likely to be measured by interviews. Analysis of 338 ratings from 47 actual interview studies indicated that basic personality and applied social skills were the most frequently rated constructs, followed by mental capability and job knowledge and skills. These authors also found that high- and low-structure interviews tend to focus on different constructs. Furthermore, they suggested that the reason why structured interviews tend to have higher validity is because they focus more on constructs that have stronger relationships with job performance. High-structure interviews focus more on applied mental skills, direct job knowledge, applied social skills (e.g., communication, interpersonal skills, leadership) and organizational fit, whereas low-structure interviews focus more on general mental ability, background credentials (e.g., education, training, experience), some aspects of

personality (e.g., agreeableness, emotional stability), and physical attributes. For Huffcutt and colleagues' analysis across all studies, the validity coefficients ranged from .24 to .58. As expected, the mean validities for constructs rated in more structured interviews were higher (.39) overall than the mean validities for constructs rated in less structured interviews (.24).

Two meta-analytic studies documenting the relationship between interview ratings and cognitive ability scores are described later.

**2.4. Subgroup Differences**

Most employment-related assessments show race differences in mean scores. On average, White job applicants obtain higher mean scores than Hispanic job applicants, who in turn have higher mean scores than Black job applicants. These mean score differences are typically large enough to sharply curtail the number of Blacks selected for employment. Mean race differences in employment interview scores are often smaller than other measures because the interview is often the final selection procedure used in a multiple-procedure process. For example, if a written job knowledge test is used, Blacks and Whites will likely show large mean

score differences. However, if only those who score high on the job knowledge test are interviewed, the Blacks and Whites who are interviewed will be more similar in job-related skills than will the pool of all applicants who took the job knowledge test. This effect, known as “range restriction,” will result in an interview showing small or no race differences. Pulakos and Schmitt found very small mean differences between subgroups on interview scores. However, they used job incumbents in both of their samples, suggesting that selection to enter the job may have absorbed most of the adverse impact. Pulakos and Schmitt generally concluded that interviews yielded fewer subgroup differences and widely recommended their use. However, in 2002, Roth and colleagues found that two forms of a behavioral interview were associated with standardized ethnic group differences of .36 and .56 when corrected for range restriction. These differences were three to five times as large as the previously estimated value of .10 obtained by Huffcutt and Roth in 1998. Roth and colleagues attributed these differences largely to the fact that applicants were previously screened on measures with high subgroup differences (e.g., cognitive ability), similar to the Pulakos and Schmitt study. They recommended that when examining the psychometric properties of a predictor, analyses should focus on the applicant level, especially applicants who have not been prescreened on another predictor.

### 2.5. Incremental Validity beyond General Mental Ability

In 2000, Cortina and colleagues found that interview scores contribute to the prediction of job performance over and above cognitive ability and conscientiousness to the extent that they are structured, with scores from highly structured interviews contributing substantially to prediction. Specifically, they found that the incremental validity for Level 1 (lowest level of structure) interview scores ranged from 0.9 to 2.2% of the variance in job performance. Level 2 (moderate level of structure) interview scores explained an additional 1.8 to 6.2% of the variance in job performance. Level 3–4 (highest level of structure) interview scores explained an additional 12.3 to 22.2% of the variance in job performance over and above cognitive ability and conscientiousness. Pulakos and Schmitt also found that interviews had incremental validity beyond mental ability.

In 1996, Huffcutt and colleagues explored the extent to which employment interview evaluations reflect

cognitive ability. A meta-analysis of 49 studies found a corrected mean correlation of .40, suggesting that on average 16% of the variance in interview constructs represents cognitive ability. In fact, their results suggest that interview ratings that correlate higher with cognitive ability tend to be better predictors of job performance.

## 3. CONCLUSION

Employment interviews are frequently used selection tests. Employment interviews vary in their degree of structure. Structured interviews permit a more systematized collection of job applicant information than less structured interviews, resulting in the former providing better predictions of future behavior. The reliability of interviews is typically assessed by correlating the interview ratings of two independent interviewers. Interview structure tends to enhance reliability. Interviews can be used to measure a variety of job-related personal characteristics, including personality, applied social skills, job knowledge, and cognitive ability. Like most selection procedures, employment interviews tend to show mean differences in scores as a function of race. There is some evidence that interviews provide incremental prediction beyond that of general mental ability. In general, it is recommended that employment interviews be structured and based on knowledge of the job.

### See Also the Following Articles

Competence at Work ■ Cooperation at Work ■ Job Analysis, Design, and Evaluation ■ Work Motivation ■ Work Role, Values Sought in the

### Further Reading

- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*, 655–701.
- Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*, 565–579.
- Cortina, J. M., Goldstein, N. B., Payne, S., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325–351.
- Huffcutt, A. I. (1992). *An empirical investigation of the relationship between multidimensional degree of structure and the*

- validity of the employment interview. Unpublished doctoral dissertation, Texas A&M University.
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*, 897–913.
- Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology, 83*, 179–189.
- Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459–473.
- Huffcutt, A. I., Weekley, J. A., Weisner, W. H., DeGroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology, 54*, 619–644.
- Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*, 577–580.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422–427.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599–616.
- Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. *Personnel Psychology, 48*, 289–308.
- Roth, P. L., Van Iddekinge, C. H., Huffcutt, A. I., Eidson, C. E., & Bobko, P. (2002). Corrections for range restriction in structured interview ethnic group differences: The values may be larger than researchers thought. *Journal of Applied Psychology, 87*, 369–376.



# End of Life Issues

Rebecca S. Allen, Laura L. Phillips, and Kristen Payne

University of Alabama, Tuscaloosa, Alabama, USA

1. Introduction
  2. Definitions and Differentiation of Hospice and Palliative Care
  3. Predictors of Mortality
  4. Advance Care Planning within a Family Context
  5. Familial Caregiver Strain and Growth Potential
  6. Therapeutic Assessment
  7. Therapeutic Interventions and Treatments
  8. Conclusion
- Further Reading

## GLOSSARY

**advance care planning (ACP)** The process of preparing for medical status change events at the end of life.

**advance directive** A legal document that explains a competent individual's wish for future medical care (as in a living will) or that designates a surrogate decision maker who can make medical treatment decisions for the individual when he or she is unable to communicate these treatment preferences independently (as in a durable power of attorney for health care, also known as a health care proxy).

**anticipatory grief** A process of grieving for losses encountered during the course of illness and prior to death.

**hospice care** Palliative and supportive services that provide physical, psychological, social, and spiritual care for dying persons and their families; patients must have a prognosis of 6 months or less to live, receive no life-prolonging medical treatments, and have an identified primary caregiver.

**palliative care** Extends the principles of hospice care to a broader population who can benefit from receipt of aggressive comfort care earlier in the disease process than the point at which life expectancy is 6 months or less; provision of palliative care is not exclusive of receipt of life-prolonging medical treatments or independent living situations.

**treatment delivery** An aspect of treatment implementation that pertains to the degree to which project staff adhered to treatment guidelines and documented their treatment package (i.e., development of interventionist training manuals).

**treatment enactment** An aspect of treatment implementation that pertains to the usefulness of the treatment to the individual and/or family as indicated by their spontaneous use of treatment skills in everyday life.

**treatment receipt** An aspect of treatment implementation that pertains to the individual and/or family's understanding of treatment components and acquisition of skills.

Everyone must deal with the deaths of those they love as well as his or her own mortality. Yet one rarely considers the certainty of the end of life until confronted with the immediate needs of a loved one who is dying. This article reviews the differentiation of hospice and palliative care, predictors of mortality, advance care planning within a family context, caregiving stressors as well as positive aspects of caregiving, and assessment and treatment issues surrounding the end of life.

## 1. INTRODUCTION

Issues surrounding the end of life are receiving increased attention due to the “graying of America.” In 2000, the average life expectancy for all individuals at birth in the United States, regardless of gender or race, was 77 years. The National Center for Health Statistics reported that the leading causes of death in 2001 were heart disease, cancer, stroke, and respiratory disease. Alzheimer’s disease was listed as the eighth leading cause of death, although this is likely an underestimate given that Alzheimer’s is rarely listed as the direct cause of death on death certificates. Fully 80% of individuals who die do so in an institutional setting such as a hospital or nursing home. However, individuals report a desire to die at home.

One reason for the increased interest in end of life issues has to do with the U.S. Supreme Court’s 1990 decision in *Cruzan v. Director, Missouri Department of Health*, requiring “clear and convincing evidence” of an individual’s wishes regarding medical treatments that may be used to prolong life (e.g., artificial nutrition and hydration). Many cultural and psychosocial issues influence individual preferences for end of life medical treatments, receipt of hospice or palliative care, nursing home placement, and the experience of caregiving. These include a need for education regarding the benefits and drawbacks of life-prolonging (or life-sustaining) treatments in comparison with receipt of hospice and palliative care, socioeconomic factors, religiousness and spirituality, prior experience with end of life issues, and style of family decision making.

## 2. DEFINITIONS AND DIFFERENTIATION OF HOSPICE AND PALLIATIVE CARE

Hospice care includes palliative and supportive services that provide physical, psychological, social, and spiritual care for dying persons and their families. The National Center for Health Statistics reported that hospice patients in the United States are still predominantly cancer patients (e.g., 58% in 2000). However, guidelines to determine hospice eligibility have been developed for severe chronic illnesses such as heart disease, chronic obstructive pulmonary disease, cerebrovascular disease, and dementia. There are dramatic racial/cultural differences in the receipt of hospice care. Hospice patients are most often elderly White individuals who live with a

primary caregiver in a private or semiprivate residence and are covered by Medicare or private insurance. Only 5% of hospice services are paid for by Medicaid. Although the Medicare hospice benefit provides for care during the final 6 months of life, the average length of service in hospice programs nationally is only 46.9 days (median = 15.6 days).

A primary restriction to the provision of hospice care is that patients must have a life expectancy of 6 months or less. In addition, no life-prolonging treatments (e.g., artificial nutrition and hydration, chemotherapy, ventilation) can be ongoing during the provision of hospice care. This may be a primary barrier to receipt of hospice services by individuals who identify themselves as members of racial/ethnic minority groups. For example, African Americans are more likely to initiate life-prolonging medical treatments than are Whites. Another barrier to receipt of hospice services is the absence of an informal or formal in-home caregiver.

In contrast, palliative care extends the principles of hospice care to a broader population who could benefit from receiving aggressive comfort care earlier in the disease process. Palliative care is not restricted to individuals with a life expectancy of 6 months or less. Furthermore, provision of palliative care is not exclusive of receipt of life-prolonging medical treatments or independent living situations. Therefore, palliative care models are ideal for more medically frail elders who may benefit from aggressive comfort care well before they qualify for current hospice guidelines. Ideally, palliative care would segue into hospice care as the individual’s illness progresses. However, recognizing the signs of disease progression that can signal a need for palliative care can be difficult. Recent research has attempted to clarify some of these signs and symptoms.

## 3. PREDICTORS OF MORTALITY

Events that can signal medical status change include a nonchronic condition that calls for medical follow-up by a physician (e.g., weight loss, difficulty in breathing, heart trouble, chest pain, and trouble in swallowing) and a disease process that calls for a medical decision involving the individual’s family. In 2003, Covinsky and colleagues examined data from the final 2 years of life of frail older adults in the community and found that declines in functional status were evident at least 1 year prior to death. The degree of functional impairment was greater for those with cognitive impairment.

Medically frail individuals are at risk for various types of cognitive impairment, including dementia, confusion associated with opioid medication, delirium, and (in the case of cancer) brain metastases. In 2003, Schonwetter and colleagues found that advanced age and impaired nutritional and functional status were associated with shortened survival among patients admitted to a community hospice with a diagnosis of dementia. Three-fourths of patients meeting local Medicare fiscal intermediary guidelines for hospice eligibility died within the required 6-month window.

In addition, continued weight loss has been associated with an increase in the likelihood of dying. Weight loss is a common symptom in many types of cancer, although fluid buildup is not uncommon. Older age and feeding dependence are also associated with increased risk of mortality. However, the most prevalent symptom of impending death during the final 48 hours appears to be respiratory symptoms.

Perhaps the most common symptom of medical status change, and certainly the most troubling for the patient and the family caregiver, is pain. Individuals with more severe cognitive impairments might not be reliable in, or even capable of, self-reporting pain or discomfort. Individual pain reports might not coincide with proxy pain reports. Pain is common among nursing home residents regardless of proximity to death. In 2002, Happ and colleagues noted that most nursing home residents died without family members present and with little documented evidence of pain or symptom management. According to Engle and colleague's 1998 research, African American nursing home residents may be at particular risk for inadequate pain relief.

The National Hospice and Palliative Care Organization (NHPCO) has recommended continuous assessment of an individual's needs and treatment preferences in the context of family values. This process may be considered continual treatment planning for the end of life, and some have argued that this should occur within a family context.

#### **4. ADVANCE CARE PLANNING WITHIN A FAMILY CONTEXT**

Advance care planning (ACP) is the process of preparing for medical status change events at the end of life. ACP typically involves discussion and assessment of the understanding that the person with life-limiting illness has regarding his or her medical history, current condition,

values, preferences, and resources. ACP tends to emphasize decisions regarding the withholding of life-prolonging medical treatments rather than treatments that are desirable. It is important to note that ACP is not merely the completion of an advance directive, although completion of an advance directive may be a part of ACP.

Older adults tend to believe that others will take care of their future health care. This belief may, in turn, preclude completion of advance directives or interfere with ACP. Individuals who possess advance directives tend to be older, White, more educated, classified as middle class, and in poor health.

ACP is typically conducted in a family context because family members are often involved in end of life care. Family members tend to be present throughout the individual's illness and play an important role in end of life decisions. Sometimes, family members advocate the individual's end of life treatment preferences when he or she cannot voice these wishes independently. However, this substituted judgment is not always an accurate representation of the individual's wishes. This discrepancy may be influenced by the lack of prior discussions of end of life preferences with caregivers or family members; however, a growing body of research suggests that even with such discussions, surrogate decision makers do not always respond with the same treatment preferences as indicated by their loved ones. Many factors may contribute to this discrepancy, including differences in how healthy and unhealthy individuals view illness and quality of life. Although research is still inconclusive as to whether ACP and completion of advance directives aid in maintaining patient control over the dying process, there seems to be some suggestion that at least ACP may ease the minds of patients, caregivers, and family members concerning end of life issues.

#### **5. FAMILIAL CAREGIVER STRAIN AND GROWTH POTENTIAL**

On the topic of caregiving, research to date has focused on caregiving for chronic illnesses, most predominantly Alzheimer's disease and other forms of dementia. However, there are differences in the impact of caregiving across various diseases and racial groups. When reviewing research on caregiving at the end of life, important themes emerge, including the relationship between caregiver and patient; the search for meaning/purpose, spirituality, or religiousness; and the adjustment to bereavement.

In general, caregivers experience more negative health outcomes than do noncaregivers. Specifically, caregivers are at increased risk for developing negative mental and health outcomes such as depression, perceptions of burden, lowered immune system function, and mortality. Female caregivers experience less participation in social and recreational activities, and more movement out of the workforce, than do noncaregivers. Research has suggested that Alzheimer's disease produces similar negative effects on caregivers' well-being and mental health when severity of illness is considered.

Other important aspects to consider when evaluating the long-term effects of caregiving are the quality and type of the relationship between caregivers and the individuals for whom they are providing care. In 1998, Lawrence and colleagues found that better relationship quality was directly related to lower levels of depression and a decreased sense of role captivity for caregivers. Familial relationship has been found to influence the average number of hours per week spent in the caregiving role and the extent of burden experienced. Specifically, spouses tend to spend more time caregiving than do adult children or other family members. In addition, spousal caregivers report more fatigue and more sleep difficulties than do nonspousal caregivers.

Regarding racial differences, African Americans are more likely to be cared for by members of their extended families than are Whites, whereas adult children are a prominent source of care for both African Americans and Whites. African American caregivers also experience higher levels of overall coping mastery than do White caregivers. In contrast, Aranda noted in 2001 that Hispanic/Latino caregivers generally, but not always, report higher levels of depressive symptoms than do other ethnic/racial groups.

Beyond the impact of type of relationship or the length of care provided, the course of the illness also affects the relationship and caregivers' response to caregiving. Although a good deal of research has focused on caregiving for chronic illness, very little has focused on caregiving at the end of life. Most of what is known in this area comes from clinical reports of the dying process or observations of experienced clinicians who work with families through hospice and palliative care programs. These qualitative accounts portray the extremely difficult circumstances experienced at the end of life by caregivers, who are simultaneously caring for a loved one while also dealing with the anticipatory grief related to their loved one's impending death as well as the frustrations of dealing with an often unresponsive health care system.

Notably, few studies address potential positive aspects of palliative caregiving for persons at the end of life. Caregivers may face needs that could be identified as existential or spiritual, including finding purpose and meaning, forgiving and receiving forgiveness, maintaining hope, saying good-bye, and coming to terms with expectations for what may occur after death. In 2003, Haley and colleagues found that hospice caregivers who have high levels of social activity and social support, and who are able to find meaning and benefit in caregiving, reported lower levels of depression and higher levels of life satisfaction. This recent finding suggests that activities in which caregivers find meaning and benefit in caregiving are beneficial.

The impact of religiousness and spirituality on coping has been a particularly neglected area of research. Religiousness and spirituality may reduce the sense of loss of control and helplessness that can accompany end of life caregiving. Religious and spiritual beliefs provide a cognitive framework that reduces stress and increases purpose and meaning in the face of illness. Spiritual and religious activities may reduce the sense of isolation and increase caregivers' sense of forgiveness or control over the illness. However, religiousness and spirituality might not always have a positive influence. For example, if a person who is dying feels remorse over past failings and believes in a punishing God, he or she may experience great anxiety.

When death comes after a sustained period of chronic illness and family caregiving, family members often react with a mixture of relief, grief, and guilt. Relief at the end of the caregiving process and the end of the loved one's suffering is accompanied by a deep sense of loss. Family members experience increased stress following a decision to terminate end of life care, but in cases where there is no verbal or written advance directive, the stress levels tend to be higher and last longer. Family members may experience anticipatory grief during their loved one's dying process. This anticipatory grief process may be beneficial to adaptive coping after bereavement, but this has not been supported consistently by research.

Although most people go through the grieving process effectively and eventually resume their normal functioning, a considerable portion (perhaps 10–25%) experience some form of pathological grief. A number of risk factors for complications with bereavement have been found, including the quality of the relationship between the bereaved individual and the deceased loved one (worse outcome in conflictual relationship), gender (worse outcome in male caregiver), and the

nature of the death (worse outcome in sudden, unexpected, or socially unacceptable death). The impact of caregiving does not end at death, nor is this impact the same for every caregiver. Substantial racial/ethnic differences in subjective reactions to death have been found. Specifically, African American dementia caregivers are more likely than White dementia caregivers to initiate life-prolonging medical treatments, are less likely to have their loved one die in a nursing home, and report less acceptance of their loved one's death and greater perceived loss. Consequently, more research is needed to understand the impact of caregiving, and an individualized approach to working with these families in clinical settings is needed.

## 6. THERAPEUTIC ASSESSMENT

Careful assessment is a precursor to effective treatment, and effective treatment leads to enhanced quality of life for individuals with life-limiting illnesses. The importance of fine-tuned information on patient prognosis becomes more apparent as treatment options increase. Communication of such information to individuals and their families allows time to explore family values and goals for future medical treatment. Such medical treatment can be viewed as life sustaining or life prolonging, depending on the perspective of each individual and family.

Assessment of patients' preferences regarding future medical treatment can be conducted formally through the use of advance directives or informally through discussions with family members. One assessment tool that may be helpful in promoting informal discussions and in clarifying wishes for future medical care is the Life Support Preferences/Predictions Questionnaire (LSPQ). This instrument was designed to identify treatment preferences in specific illness scenarios. Each scenario differs from the others in severity, level of pain, prognosis, and/or impairment. Following presentation of each scenario, respondents are asked to indicate which treatment options they would desire given that situation. Data from such assessments can provide guidelines for appropriate treatment and disease management.

Clinicians also may wish to conduct needs assessments with individuals and family members throughout the course of the individuals' chronic life-limiting illnesses. Continuous and comprehensive assessment of individual and family perception of need can provide a medium for matching specific individual and family characteristics with appropriate therapeutic interventions. Data from

such assessments can provide education and support for individuals living with illnesses at the end of their lives. Such information can also provide education and support for family members regarding their own risk for disease and available preventive measures.

## 7. THERAPEUTIC INTERVENTIONS AND TREATMENTS

Many families and individuals with life-limiting illnesses have difficulty in accepting the inevitable consequences of the diseases and may resist engaging in interventions such as advance care plans, psychotherapy, support groups, and respite services or seeking information about the benefits of palliative care to alleviate suffering. Clinicians must recognize the potential of such resistance and approach individuals and families with the utmost sensitivity and empathy. However, once individuals and families are engaged in discussions of therapeutic interventions at the end of life, the opportunity for meaningful interaction and the potential for individual growth and growth in relationship become more apparent. Some interventions that target various stages of life-limiting illnesses and end of life treatment issues are reviewed briefly in the following subsections.

### 7.1. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments

The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) project, conducted in 1995 by the SUPPORT principal investigators, implemented and evaluated a physician-assisted intervention for dying patients in hospitals. The intervention was aimed at addressing inadequate communication regarding prognoses and patient care preferences. Specifically, the intervention was aimed at (a) providing accurate and timely prognostic information to physicians, (b) documenting the views of patients and families surrounding the patients' prognoses and treatment options, and (c) providing skilled nurses to carry out discussions and communicate relevant information.

To assess whether the intervention had a positive effect, SUPPORT researchers focused on (a) when and if "do not resuscitate" (DNR) orders were written, (b) patient-physician agreement regarding preferences to withhold resuscitation, (c) the number of days patients spent in intensive care units, (d) the frequency and

severity of the pain experienced by patients, and (e) the amount of hospital resources used during patients' hospital stays. The research investigators found no differences on any of the five outcomes between those who received the intervention and those who did not.

The specific reasons why the intervention did not work are unclear, but there are some suggested possibilities. The SUPPORT investigators recognized that one limitation in their ability to detect a positive impact of this treatment might be that individuals, families, and physicians were generally satisfied with the care received at that critical time. According to the researchers, another possible explanation is that the hospitals in which the SUPPORT study was conducted might have provided better care than is typically available to the general public.

Although these are plausible explanations, another possibility is variability in the implementation of the intervention itself in terms of treatment delivery by the investigators, treatment receipt by the participants (i.e., understanding), and treatment enactment (i.e., the degree to which the information provided by the SUPPORT team was used by the families in the study). SUPPORT was the first intervention of its kind to submit to the rigors of scientific testing and represents seminal research in this area. The investigators chose to maximize generalizability by conducting the research in as natural a way as possible in their chosen hospitals. However, in so doing, the experiences of individuals and families in the treatment group varied dramatically.

Delivery of the treatment was allowed to vary based on the SUPPORT nurse's discretion and ranged from informational support to intense emotional support. In addition, the nurse did not always meet with the families or the patients, nor did the nurse discuss the same topics with each participant. This variation in treatment can be beneficial to the individuals involved, but it makes it difficult to determine whether the treatment as a whole was effective or ineffective due to the variety of interventions delivered.

The only measure of receipt of treatment was the physicians' indications of receipt of the reports regarding DNR or the placement of the reports in the medical charts. Although the placement of these reports in the charts could be conceptualized as treatment enactment, there is little information on whether these reports were ever actually used by those in constant contact with the patients or families. Efforts to standardize the treatment provided by the interventionists, measures of treatment receipt for both physicians and patients/families, and a measure addressing the use (enactment) of the

treatment are needed before such interventions can be deemed beneficial or ineffective.

## **7.2. Advance Directives, Values Assessment, and Communication Enhancement**

In 2001, Ditto and colleagues implemented an intervention that consisted of patients completing an advance directive in the presence of family members. In some groups, discussions were also encouraged to facilitate communication between patients and family members regarding medical treatment preferences. One group had advance directives with discussions, one group had advance directives without discussions, one group had discussions without advance directives, and a control group had neither advance directives nor discussions. The results of this study showed that none of the interventions resulted in an increase in patient-family member agreement when family members were asked to predict patient preferences. Although the study revealed no increase in agreement, it did reveal an increase in comfort and understanding as reported by patients and family members regarding end of life decisions. As with the SUPPORT study, data regarding treatment implementation were not reported systematically, although there seemed to be variability in the delivery of treatment to the various groups.

## **7.3. Meaning-Centered Group Interventions**

William Breitbart and colleagues are currently testing an intervention designed to enhance a sense of meaning and purpose and to improve quality of life among advanced cancer patients. The treatment is delivered in a group format (eight sessions) and targets the creation of a supportive environment to facilitate understanding of the possible sources of meaning before and during cancer illness. Basic concepts include the achievement of transcendence through creativity (e.g., work, deeds, causes), experience (e.g., nature, art, relationships), and attitudes toward suffering. Group sessions include information about sources of meaning, a storytelling life project, and hopes for the future. Anecdotal reports from Breitbart and colleagues indicate that the intervention may be effective in decreasing hopelessness and desire for hastened death while increasing meaning and spiritual well-being. No treatment implementation data are yet available.

#### **7.4. Individual Problem-Solving Skills Training**

Another type of intervention used at the end of life is problem-solving skills training. Susan McMillan's "Caregiver Intervention to Improve Hospice Outcomes" is one example of a problem solving-based intervention. This intervention takes place over a period of 9 days, during which time three intervention visits occur. In these visits, caregivers receive the "COPE" intervention involving the following: Creativity (viewing problems from different perspectives to find new strategies), Optimism (having a positive but realistic attitude toward the problem-solving process), Planning (setting goals and thinking out steps to reach those goals), and Expert information (learning what caregivers need to know about the patient's condition, when to seek professional help, and what can be done by caregivers to help the patient).

The three problem areas on which McMillan focused regarding the hospice patient's condition were pain, constipation, and dyspnea. For each of these problems, a specific module was designed for caregivers to learn assessment and problem-solving techniques they could implement to help the patient. In addition, caregivers were asked to keep a patient symptom diary in which they recorded patient symptoms to monitor and share this information with the hospice staff as needed.

Treatment implementation was measured in this study. Delivery of the treatment was monitored by case conferences to reduce the possibility of drift from the protocol or cross-contamination of the treatment conditions. In addition, visits in the intervention group were tape-recorded, and 10% of the tapes were reviewed monthly to ensure adherence to protocol. At monthly case conferences, the possibility of an undesired change in hospice staff behavior was assessed by a question regarding any noticed changes by caregivers with respect to the hospice staff's behavior toward the patient.

Treatment receipt, or caregivers' understanding of the intervention, was verified by the interventionist's inquiry, "What would you do if . . .?" Clarification regarding specific skills was provided to caregivers if needed to facilitate adequate treatment receipt. At the end of the intervention, receipt was assessed by asking caregivers to rate their ability to manage symptoms successfully and comparing their responses prior to the intervention with their responses following the intervention.

Treatment enactment can be examined through the caregivers' diary to determine the extent to which caregivers complied with the daily assessment and

recording of patient symptoms. If caregivers found the information and monitoring system provided by the project to be helpful, they would be expected to use the skills in providing care to the patient.

#### **7.5. Ongoing Studies by the Authors**

Our team has developed two interventions that combine problem-solving skills training with social support and meaning-centered therapeutic activity. One intervention is the care integration team (CIT), which focuses on caregivers of terminally ill patients enrolled in hospice care. The intervention consists of four weekly home visits and has a randomly matched minimal support treatment control group. The intervention for caregivers focuses on use of problem-solving skills in four topic areas: (a) symptom management, (b) communication with hospice staff, (c) caregiver self-care, and (d) anticipatory grief. Treatment implementation measures are included in this ongoing study.

Unlike the CIT intervention, the legacy intervention focuses on both the individual with life-limiting illness and the caregiver. The legacy intervention is an effort to help patients and caregivers to find meaning through intergenerational activities. This is achieved through the joint development of the patient's "personal legacy" for future generations. This legacy can take multiple forms: a family photo album, a family cookbook or scrapbook, audiotaped or videotaped stories of the older adult's life, personal letters to younger family members, and the like. The intervention consists of three weekly 60- to 90-minute sessions in which the caregiver and patient learn about what a legacy is, select a legacy, obtain materials to create the project, and review their experience with members of the research team. A control group receives weekly 5-minute telephone calls and answers questions about spontaneous intergenerational activities to promote family ties. Encouraging families to develop lasting memories with their loved one at any stage of the disease process can enhance meaning in relationships and provide a focus to heal prior conflicts.

These interventions were designed with the three components of treatment implementation in mind. Both the CIT and legacy projects use audiotapes of intervention sessions (10%) in evaluations by an independent professional clinician not involved with the research project. The independent evaluator monitors the tapes and transcripts for interventionists' adherence to research protocol in treatment delivery. Both projects have caregiver and interventionist training manuals that are provided as guides to treatment components.

The CIT intervention measures treatment receipt by knowledge assessments at the first week of training and immediately following the intervention. If caregivers do not achieve an 80% score on these assessments, further training is provided to ensure receipt of the skills. Both the CIT and legacy interventions also use contact logs to document the content and length of all sessions so as to measure receipt of treatment, including who initiates the contact and the purpose of the contact.

The CIT project assesses treatment enactment by interventionists' review of caregivers' personal tracking forms. The interventionists then rate caregivers' compliance with all aspects of treatment on a Likert-type scale evaluation. For treatment enactment, interventionists working on the legacy project note the extent to which a legacy was completed by the family, again using a Likert-type scale evaluation.

## 8. CONCLUSION

This article has reviewed issues associated with personal and familial stressors and potential for growth when facing end of life issues during adulthood. Because of increased numbers of individuals living to very old age today, research and clinical attention are being directed at end of life care. Of necessity, end of life issues such as ACP, choice of medical treatment, assessment of treatment needs, and delivery of therapeutic treatments involve both individuals with life-limiting illnesses and their family members. Knowing when to initiate aggressive curative or aggressive palliative treatments is difficult given variable predictors of mortality across disease categories. This article has attempted to provide a broad overview of current clinical research efforts surrounding end of life issues. Without question, these topics will continue to be a focus for research, service, and policy efforts during the coming decades.

## Acknowledgments

Funding from the National Institute on Aging (K01AG00943) to R. S. Allen supported preparation of this article. Special thanks are extended to the staff of the Center for Mental Health and Aging for assistance in this preparation.

## See Also the Following Articles

Elder Caregiving

## Further Reading

- Allen, R. S., DeLaine, S., Chaplin, W. F., Marson, D. L., Bourgeois, M., Dijkstra, K., & Burgio, L. D. (2003). Advance care planning in nursing homes: Correlates of capacity and possession of advance directives. *The Gerontologist*, *43*, 309–317.
- Allen, R. S., Haley, W. E., Small, B. J., & McMillan, S. C. (2002). Pain reports by older hospice cancer patients and family caregivers: The role of cognitive functioning. *The Gerontologist*, *42*, 507–514.
- Allen, R. S., & Shuster, J. L. (2002). The role of proxies in treatment decisions: Evaluating functional capacity to consent to end-of-life treatments within a family context. *Behavioral Sciences and the Law*, *20*, 235–252.
- Aranda, M. P. (2001). Racial and ethnic factors in dementia care-giving research in the U.S. *Aging and Mental Health*, *5*(Suppl. 1), S116–S123.
- Covinsky, K. E., Eng, C., Lui, L. Y., Sands, L. P., & Yaffe, K. (2003). The last 2 years of life: Functional trajectories of frail older people. *Journal of the American Geriatrics Society*, *51*, 492–498.
- Ditto, P. H., Danks, J. H., Smucker, W. D., Bookwala, J., Coppola, K. M., & Dresser, R. (2001). Advance directives as acts of communication: A randomized controlled trial. *Archives of Internal Medicine*, *161*, 421–430.
- Engle, V. F., Fox-Hill, E., & Graney, M. J. (1998). The experience of living–dying in a nursing home: Self-reports of Black and White older adults. *Journal of the American Geriatrics Society*, *46*, 1091–1096.
- Haley, W. E., LaMonde, L. A., Han, B., Burton, A. M., & Schonwetter, R. S. (2003). Predictors of depression and life satisfaction among spousal caregivers in hospice: Application of a stress process model. *Journal of Palliative Medicine*, *6*(2), 215–224.
- Happ, M. B., Capezuti, E., Strumpf, N. E., Wagner, L., Cunningham, S., Evans, L. K., & Maislin, G. (2002). Advance care planning and end-of-life care for hospitalized nursing home residents. *Journal of the American Geriatrics Society*, *50*, 829–835.
- Lawrence, R. H., Tennstedt, S. L., & Assmann, S. F. (1998). Quality of the caregiver–care recipient relationship: Does it offset the negative consequences of caregiving for family caregivers? *Psychology and Aging*, *13*, 150–158.
- Schonwetter, R. S., Han, B., Small, B. J., Martin, B., Tope, K., & Haley, W. E. (2003). Predictors of six-month survival among patients with dementia: An evaluation of hospice Medicare guidelines. *American Journal of Hospice & Palliative Care*, *20*, 105–113.
- SUPPORT Principal Investigators. (1995). A controlled trial to improve care for seriously ill hospitalized patients: The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). *Journal of the American Medical Association*, *274*, 1591–1598.



# Engineering Psychology

Klaus-Peter Timpe, Hans-Gerhard Giesa, and Katharina Seifert

Technical University Berlin, Berlin, Germany

1. Historical Background
2. The Human–Machine System as the Subject of Engineering Psychology
3. Human Information Processing in Human–Machine Systems
4. Human Reliability
5. Methods
6. Development and Planning of Human–Machine Systems
7. Future Developments  
Further Reading

**task analysis** Any process for analyzing the behavior of a human within a human–machine system that determines the interactions between humans and equipment or machines with the underlying working conditions in mind.

**usability** The degree of ease with which humans can interact with a system or a piece of equipment.

The subject of engineering psychology is the optimization of the exchange of information within the human–machine system with the use of planning, analysis, design, and evaluation.

## GLOSSARY

**cognitive task analysis** Any further development of traditional task analyses that yields information about the knowledge, thought processes, and goal structures that underlie observable task performance.

**compatibility** Consistency in the direction of the movement of relationships and operator expectations.

**human–machine system** A convenient abstraction of the goal-oriented cooperation between persons and technical systems to accomplish a self-determined or assigned task.

**human reliability** The probability of the successful performance of the human activities necessary for a human–machine system.

**interface design** The construction of that technical part of the machine directly interacting with the human operator on the basis of empirically grounded rules and heuristics to achieve efficient human–machine interaction as well as high system performance.

**operator** The human working in a human–machine system.

## 1. HISTORICAL BACKGROUND

Changes in work in all branches of industry, transportation, and services after World War II led to an increasing amount of monitoring, control, supervision, and planning activities for operators. To match the equipment to the needs and cognitive capacities of the human operators, engineers, psychologists, physiologists, and anthropologists have outlined general principles for an optimal layout of technical systems that involve people.

Since the late 1940s, the framework of specialists has focused on that topic developed as a part of human factors engineering called engineering psychology. At the beginning of the 1960s, the subject was examined in Europe, especially in England, France, and Germany. Similarly, many studies were undertaken in the Soviet Union in this new field of technical design under the

label of *ingenemaja psychologika*. In Japan, research in engineering psychology also spread quickly. Similar subjects were worked on in the United States, although socially and culturally different foci were set.

During the initial phase of engineering psychology up to around 1960, systems design research was focused on the developmental design principles of machine displays, controls, and interaction between humans and machines. However, this changed quickly due to the fast development of computing technology. By the end of the 1960s, ever more portions of the mental work of humans were already automated and new questions about the allocation of functions between humans and machines had to be answered. This led away from the “dials and knobs” optimization of the early 1950s to human–machine/computer interaction as an area of engineering psychology. Although the adjustment behavior of humans was initially modeled as an automatic controller, the field of research spread soon to investigations of the operator as a dialogue partner and interactive problem solver. System design issues requirements are currently driven, and will also be driven in the future, largely by trends in technology such as computerization, automation, knowledge, and information processing.

The new field of activity engineering psychology is closely related to other spheres of research such as human factors or human factors engineering (United States), ergonomics (United Kingdom and Germany), and cognitive ergonomics. However, differences in the core psychological matter and methodology of engineering psychology clearly separate it from these subjects in that engineering psychology focuses on information exchange in human–machine systems. The methods and tools used for analysis in the human–machine information exchanges are predominantly derived from cognitive psychology, which developed numerous paradigms to investigate fundamentally human information processes. During recent decades, an increasing trend is to incorporate and transfer methods from other scientific areas to the field of human–machine interaction studies.

## 2. THE HUMAN–MACHINE SYSTEM AS THE SUBJECT OF ENGINEERING PSYCHOLOGY

In general, the numerous definitions of engineering psychology can be summarized as follows: the optimization

of the interchange of information within a human–machine system with the use of planning, analysis, design, and evaluation.

The concept of human–machine systems refers to a convenient abstraction of the goal-oriented cooperation between persons and technical systems to accomplish a self-determined or assigned task. This more abstract definition covers the relevant features of the diverse variations of the goal-oriented information interchange between humans and machines in different situations. The concept of a human–machine system implies that at least one person and at least one machine work together.

The general structure of a human–machine system (Fig. 1) is a control system in which an operator and/or a team make decisions to steer the technical system according to organizational basis, goal, task, and perceived feedback about the environment and process conditions. The behavior of the human–machine system is thereby determined by qualitatively different processes. On the one hand, it is shaped by human data processing based on social, psychological, and biological regularities. On the other hand, it is influenced by technical processes that obey physical laws.

The reception of information concerning process data and environmental properties takes place either directly or indirectly (e.g., a direct visual approach, an indirect instrumental approach). Information can also be entered either with appropriate devices or through language, muscle activities, and the like.

The location of these interferences and/or of the machine can be nearly arbitrary spatially. If this aspect of informational coupling is more intensely differentiated (i.e., if machines and their maintenance or control are handled by persons at different places), the term “distributed human–machine systems” may be used as well. Subtasks, responsibility, decisions, and knowledge are distributed among different human participants in different hierarchical levels so that the classical problem of task allocation arises again.

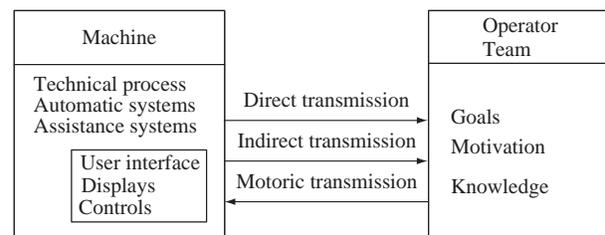


FIGURE 1 Schematic representation of information interchange in human–machine systems.

The substantial components of the technical entity “machine” in human–machine systems are indicating and control units, automated subunits, and computer-aided assistance systems (e.g., assistance or diagnostic systems) as well as the technological process controlled. Information indicators and control components are often integrated, for example, in control rooms, cockpits (for vehicles and airplanes), conning bridges, control desks, and control panels and screens.

(perception), the memory with its substructures (information retrieval, selection, and recognition), and the system for the execution of the (motoric processing) action (Fig. 2). It is assumed that all utterances and actions by the operators are the result of information processing (e.g., comparisons, judging, decisions) in the nervous system. These processes are essentially steered by signals from the environment and by knowledge stored in memory. As shown in Fig. 2, the data are received and decoded by receptors in the eyes, in the ears, and in the haptic system (i.e., the information is perceived). The decoded information reaches the brain, where it is compared with previously stored information. The result of this comparison can be recognition or comprehension, but it can also prompt a process for behavioral decisions that result in sensomotoric activity (e.g., as steering activities). Engineering psychology considers information processing on three different levels: knowledge based, rule based, and skill based.

As a result of research in the field of general psychology, it has been shown that the evaluation of the information processed has a strong influence on

### 3. HUMAN INFORMATION PROCESSING IN HUMAN–MACHINE SYSTEMS

#### 3.1. Circulation of Information in Human–Machine Systems

Three components can be differentiated in information circulation and are represented separately for didactical reasons: the system of the initial processing of stimuli received by the receptors of an appropriate sense

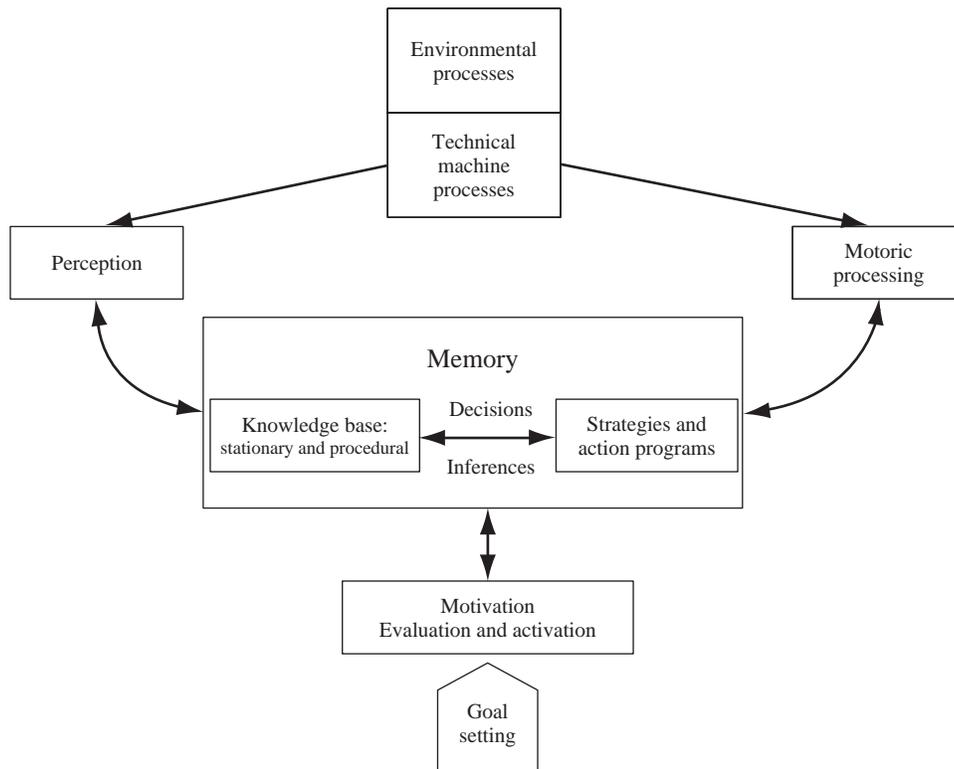


FIGURE 2 Information circulation processes in human–machine interaction (based on a structural model of human cognitive systems).

performance. This is why mental processes cannot be investigated without taking activating processes into account. This means that the interrelationship between information processing and evaluation, as well as activation processes, is an important field of research for engineering psychology (e.g., for measures of workload and situation awareness).

### 3.2. Goal Setting

All actions in a work context need to set a specific goal that is expected to be achieved by the actions. The goal-setting process includes an assessment of a cost-efficiency ratio as well as of a demands/effort ratio. Furthermore, the process is affected by the interpretation, motivation, and values of the operator. This can be described as an anticipation of future action and its result. The goal is to resolve problems in a definite time, to anticipate changes in the state or in the process, and/or to develop a process of execution so as to set means and define results. Goal setting can be processed at different levels of detail, depending perhaps on the amount of experience in the domain of action, on the complexity of the task, and on the degree of freedom implicated in it. Goals can be differentiated as those that are set by or taken over from another authority and those that are self-determined. However, self-determined goals are achieved with more motivation and involvement than are assigned ones. The process of goal setting is completed by means of the resolution to execute a specific action. Goal setting can be supported by the unambiguous design of displays and controls that give the operator useful hints for the current action needed.

### 3.3. Perception and Interface Design

Human perception is distributed over diverse channels of information intake (i.e., the senses) that are integrated into the process of perception. The senses are usually distinguished as the visual sense, auditory sense, somato-visceral sense (including the haptic sense), the olfactory and equilibrium sense, and the sense of taste. The research on perception in experimental psychology during the past 50 years or so has been predominantly directed toward the visual and auditory sensory system. In addition to these two senses, the haptic sense—especially the sense of touch and kinesthetics—has been “rediscovered” for the interface design of advanced human-machine systems (e.g., to synthetically model force feedback in distributed or virtual human-machine systems).

Psychological theories of perception differ in how they assume that information is taken in, describing the process as follows:

- A direct process that rests on the spatio-temporal invariance of the environment (ecological approach of perception)
- An indirect process involving different stages in the active construction of the percept from receptor sensations based on processing rules, the referential system, and the experience of the person involved (constructivist approach)

Current theories of perception often assume that bottom-up processes, as well as top-down processes, are involved in the various levels of perception.

Engineering psychology relies particularly on the findings of Gestalt psychology, the laws of perception, and psychophysics to offer recommendations for the design of icons for visual displays as well as of acoustic feedback. The laws of perception are related to the perceptual grouping of visual and acoustic stimuli depending on their properties. Following are a couple of examples:

- *The law of similarity*: Similar items tend to be grouped together.
- *The law of proximity*: Items with spatio-temporal closeness tend to be experienced as belonging together.

The laws of perceptual organization also explain the figure-ground disjunction and can be applied for adequate display design in human-machine systems. For example, smaller and symmetrical pieces, as well as orthogonally oriented figures, can be more easily perceived as objects.

Psychophysical findings, which are important for the design of human-machine systems (especially for information coding), are the minimal and maximal thresholds of perception for the individual senses, their discrimination thresholds, and their spatio-temporal resolution.

To support the operator in the comprehension of information, this information can be coded into features that are easy to perceive and distinguish (e.g., the color red for “hot”). Most of these kinds of codes are understandable due to knowledge about cultural conventions.

### 3.4. Cognition

In the future, many fields of technical design, diagnostic processes, modern management, and so forth will not remain competitive without the help of technical

aids for knowledge processing, (group) decision making, and other aspects of cognitive tasks. The ultimate aim is to support cognitive tasks as effectively and transparently as possible. Knowledge-based supporting aids and decision support systems have been developed based on findings from cognitive psychology.

### 3.4.1. Knowledge-Based Supporting Aids

A knowledge-based supporting aid is a technical entity for information processing that helps the operator to achieve his or her tasks in a human-machine system by taking over certain subtasks. These knowledge-based supporting aids for technical design have become an important field of engineering psychology in research and practice.

However, this knowledge must first be prepared and presented in such a way that it can be implemented in a supporting system. But this is not a matter of acquiring knowledge for knowledge's sake. Instead, it is a matter of being able to implement the relevant knowledge related to the specific task. A wide range of psychological methods (Table I) are available for the elicitation of the relevant procedural and declarative knowledge that exists, for example, as semantic networks, production rules, or schemata.

Knowledge-based diagnostic systems have been developed and applied in numerous areas, for example, for troubleshooting in medicine, mechanical engineering, or technical design. This field is witnessing a trend where engineers, engineering psychologists, and computer scientists are working together. Although the body of knowledge has been developed particularly by engineering psychologists, it has been implemented by engineers and computer scientists.

### 3.4.2. Decision Support Systems

Findings from the psychology of cognition should be applied to decision support or decision analysis systems. These systems contain axiomatically grounded decision procedures for managers and operators working in complex systems. Often, these working conditions are complicated by time pressure, uncertainty, information overflow, and high importance. Decision support systems in this field are frequently based theoretically on normative and descriptive perspectives on the process supported (e.g., multiattributive value theory [MAVT], multiattributive utility theory [MAUT]).

TABLE I  
Methods in Engineering Psychology

<i>Objective area</i>	<i>Method examples</i>
System validation	Measurement of performance data Physiological measurements Verbal protocols Interviews Questionnaires
Cognitive task analysis	Observations Interviews Verbal reports Analysis of team communication Critical decision method Cognitive walkthrough GOMS
Knowledge elicitation	Association techniques Sorting techniques Structure formation technique Discussion in groups Grid techniques
Human reliability analysis	Analytical methods (e.g., THERP, CREAM, ATHEANA) Time-reliability correlations (e.g., HCR) Structured expert ratings (e.g., SLIM)
Usability testing	Thinking aloud method Constructive interaction Retrospective testing Coaching method Heuristic evaluation
Cognitive modeling	Computer simulations based on cognitive architectures (e.g., SOAR, ACT-R, GOMS)
Workload	Subjective measures (e.g., rating scales, psychometric techniques) Performance measures (e.g., primary task measures, secondary task measures) Physiological measures (e.g., cardiac function, muscle function, eye function, brain function)

In up-to-date systems, a team makes decisions more often. But practical design recommendations are not yet possible for the automated support of team decision making because a generally accepted concept does not yet exist. In the future, team decision support will become an important research and application domain of engineering psychology.

### 3.5. Motoric

Since the 1940s and 1950s, general principles for the choice and design of controls depending on working demands have already been established. Today, this information can be found in handbooks on human factors engineering or ergonomics that pertain, for example, to the optimal control/display ratio, directions of movement relationship, or coding and arrangement of controls, especially in terms of compatibility.

Compatibility is one of the most important factors to be considered in systems design. This is especially significant in stressful situations where learned habits often break down and can cause errors in action. These results from early engineering psychology can be found in all relevant compendia.

With ongoing technological development, and especially with digitalization, new concepts of human-machine interaction have been developed. Today, the research in engineering psychology is paying much attention to setting up design recommendations for speech and gesture interaction in cars and planes as well as in conventional computing. Some of these input devices are recognized as convenient and highly standardized (e.g., mice, touch screens, keyboards). They are manufactured for different practical purposes due to their widespread application in different work domains. The number of specific input devices for defined domains (e.g., endoscopic surgery) is increasing. These devices must fulfill the special demands of specific domains as well as general ergonomic criteria for evaluation. Speech input and automated image processing input (e.g., gesture input) are currently not used universally due to problems in recognition reliability, costs, and unsolved technical constraints (e.g., noise from the environment, changes in the brightness of the surroundings). The cultural aspects of the design of user interfaces, and especially of input devices, are more obvious today than they were previously due to the globalization of industrialization, and they constitute a limiting factor in standardization.

Another well-elaborated research area in engineering psychology is the modeling of the sensomotoric regulation of operators in flight and driving tasks. Formal descriptions of human sensomotoric regulation were sought at the end of the 1940s due to the requirements of military engineering. The goal of this research was to develop an analytical representation of the dynamics of a human-machine system for the choice of optimal system dimensions and to train operators on the model of the new system.

Human sensomotoric control can be described in terms of the mathematical control theory based on a

structural analogy between an automatic control circuit and a human-machine system. The results of this intensive research, well known as tracking experiments, are numerous mathematically descriptive equations for the "control person's" behavior. The crossover model, the quasi-linear model, and the time-optimized model are frequently cited in this context.

Admittedly, such equations are valid only for skill-based level activities. Higher cognitive procedures, particularly the motivational and behavioral strategies of an operator, cannot be described with these equations. Today, research is concerned with developing complex operator models by means of fuzzy logic approaches, artificial neural networks, and so forth.

### 3.6. Attention

Attention is one of the frequently used psychological concepts relating to the main problems of automated human-machine systems design such as the following:

- Holding operators in the loop
- Maintaining the operator's positive situation awareness and availability
- Reducing the consequences of information overflow

The concept of attention refers to the human ability to focus information processing on selected events or contents over a specific amount of time. The selection of information is an important precondition for precise action. Attention lapses in interaction with a technical system may cause dangerous states or situations. But the selection of current important events is more difficult due to potentially important information and time pressure. These circumstances contribute to the mental workload of the operator. Another problem in advanced working domains is to maintain readiness for action at an adequate level of vigilance. This problem is caused by long periods of monitoring an automated system in which nothing that requires action happens or by lengthy tasks (e.g., truck driving for hours during the night). The consequences of decreasing attention or vigilance for the performance of the human-machine system are often the same: Performance falls and strain on the operator increases. Different methods have been developed to quantify actual workload levels.

## 4. HUMAN RELIABILITY

It is estimated that in 60 to 90% of all system failures, human actions are the main cause. In light of

these statistics, the study of human error has emerged as an important discipline of engineering psychology.

Errors in work activities are considered to be a result of inadequate training, poor working conditions, and organizational factors that are also a consequence of learning and adaptation to different situational conditions. This multifaceted interrelation means that no homogenous theory about human errors exists. Taxonomies are required for the measurement of human error. There are two main approaches for classification. The first refers to the manifestation of errors such as observably incorrect actions. These taxonomies are widely used in human reliability analysis where error frequencies must be estimated. The second approach concentrates on variations in underlying psychological mechanisms, that is, the internal causes of human error. The origin of errors resulting from incorrect action is described in various ways in the psychological literature. Both approaches are related to “active errors” that have an immediate and direct impact on the system.

To meet the requirements of the influence of organizational factors, the concept of latent errors has been established. Latent errors originate in decisions taken in the managerial and organizational spheres and may lie dormant for long periods of time.

The concept of human error is closely related to the concept of reliability. The term “reliability” has its origins in the engineering sciences and is based on the theory of probabilities. The frame of reference is the frequency of errors. On the basis of different assumptions concerning error distributions, system structures, and time intervals, the reliability of a system is calculated. The concept of reliability has been transferred from technical systems to human actions. Several attempts have been made to detect human error probabilities just like in technical systems.

Engineering psychology focuses mainly on operators and their cognitive competencies and resources. But it also must consider the organizational and situational aspects of working conditions for to evaluate the reliability of an entire human-machine system.

## 5. METHODS

The most important objective areas and dedicated methods used in engineering psychology are shown in Table I. They are described briefly in this section.

### 5.1. System Validation

System validation examines whether a machine is matched to the needs and cognitive capacities of the human operator. Validation normally takes place in an experimental setting within a controllable environment where disturbing variables can be eliminated or controlled.

During early phases of the development of a new human-machine system, single subsystems are evaluated in the laboratory. The performance of a human-machine system should be investigated under realistic conditions to achieve integrated system validation. The preparation and investigation of early prototypes have become less costly during recent years with the further dissemination and use of simulators. Technical possibilities, such as log file recordings and freezing a situation, make possible techniques of investigation that are not feasible in the real world.

A broad inventory of empirical methods from psychological research can be employed to validate new systems in laboratories or in simulators (e.g., measurement of performance data such as speed and accuracy, physiological measurements, verbal protocols, interviews, questionnaires). Workload and situation awareness are typical concepts used for system validation in engineering psychology.

### 5.2. Cognitive Task Analysis and Knowledge Elicitation

A new area of engineering psychology, cognitive task analysis, concentrates on cognitive aspects of human-machine interaction that are not directly observable. As a further development of traditional task analysis techniques, cognitive task analyses are performed to gain insight into the knowledge representations, thought processes, mental strategies, and goal structures that underlie task performance. The results are used for the design of human-machine interactions (e.g., task allocation, preparation of procedures); for the development of training, testing, and selection tools; for the development of expert systems; and as a basis for computer simulations.

In addition to observations, verbal statements from job experts are the main source of information for data collection. Information about cognitive processes is collected by using interview techniques and verbal reports or with the analysis of team communication. Visualization tools and sorting methods are applied for illustration. Further methods for a cognitive task analysis

are the critical decision method, cognitive walkthrough, and the GOMS architecture.

A special field of cognitive task analysis is knowledge elicitation. A wide range of psychological methods are available for elicitation of procedural and declarative knowledge necessary for the development of knowledge-based supporting aids. In addition to common methods (e.g., interview techniques, questionnaires), other methods employed include association techniques, sorting techniques, structure formation techniques, and grid techniques.

### 5.3. Human Reliability Analysis Techniques

Human reliability analysis is concerned with prediction of the likelihood that human performance within a human-machine system will fail. Numerous methods for human reliability analysis have been developed. Traditional HRA techniques can be divided into analytical methods, time-reliability correlations, and structured expert ratings.

Analytical methods (e.g., THERP) are based on the breakdown of a task into partial tasks or actions. Error probabilities are determined for these tasks, and actions and then integrated into an overall error probability for the analyzed task. Time-reliability correlations (e.g., THERP diagnostic model, HCR model) are used for the assessment of critical events when interpretations, diagnoses, and decisions about necessary actions must be made. They are based on a simplified correlation between available time and error probability. Structured expert ratings are used if no numerical data about the analyzed task exist. Experts can carry out an assessment with the SLIM method, which provides numerical ratings of the strength and importance of a relatively small number of factors that influence performance. Newer human reliability analysis techniques (e.g., CREAM, ATHEANA) focus on errors resulting from cognitive processes.

### 5.4. Usability Testing

In general, usability refers to how well individuals can interact with a product or a piece of equipment. Usability is connected with attributes such as learnability, flexibility, intuitiveness, and likability. Usability tests can be performed as a formative or summative evaluation. Formative usability tests are done to improve a system design as part of the design process.

Summative usability testing aims at assessing the overall quality of a product.

Quantitative and qualitative methods for performance measurement are used for usability testing. Primarily qualitative methods are used for formative usability testing, especially variants of the thinking aloud method (e.g., constructive interaction, retrospective testing, coaching method). Usability testing can be combined with heuristic evaluation to rely on a list of established usability principles when applied by an expert.

### 5.5. Computer Simulations

Models are used to analyze the attributes of a human-machine system on the basis of an abstract description of the structure and the performance of the system. In the engineering field, computer simulations have been increasingly applied in the design and evaluation of new systems. Simulations of the technical subsystem must be connected with human operator models to simulate interactive systems. In engineering psychology, models from cognitive psychology are adapted to include human behavior in the simulation of interactive systems. Cognitive architectures such as SOAR, ACT-R, and GOMS represent theories for simulating and understanding human cognition that are used for the simulation of human behavior. Detailed cognitive task analyses are needed to construct computer simulations that represent human behavior.

## 6. DEVELOPMENT AND PLANNING OF HUMAN-MACHINE SYSTEMS

The goal of engineering psychology is to create demands for human-oriented work design within the framework of modern information technologies. The theoretical foundations for this type of work have already been securely laid and are supported by both experimental and empirical evidence. They have been developed for successful application to systems design:

- Eliminating the negative effects of work (e.g., workload)
- High dependability
- High performance
- High usability
- General well-being and health

The main objective is to create the best possible system and to ensure that numerous components of human

information processing and task performance are included in the job requirements so that well-balanced jobs can be designed.

To achieve this goal, one must move away from the corrective systems design that dominated the initial phases of engineering psychology and move toward a preventive and flexible form of systems design. Therefore, the system's layout should be based on a dynamic task allocation rather than on a strategy of static engineering or of static human-centered automation. A system-centered automation strategy distributes functions to both the operator and the technical system so that the operator can flexibly decide whether to use automation to accomplish a task or to do it manually.

The influence of new automation strategies has moved the process of systems design from a linear-sequential design to a parallel-hierarchical and iterative procedure. Numerous process models for systems design have been developed (e.g., the V-model, the spiral model, the prototyping model). These models all incorporate empirical evaluation and phased correction into

the various stages of engineering human-machine systems. Figure 3 is a rough schematic representation of this principle.

This transition from a linear systems design to a parallel one that takes the user into consideration is a major step toward reaching the goals that engineering psychology hopes to achieve by influencing the nature of modern work. The presentation of the relevant psychological theories, concepts, and empirical data in a technically usable form is crucial to the success of such a procedure. Rapid and successful developments have been made in this field, especially in aviation, automotive, and production engineering. These developments, as compared with corrective design procedures, require considerably less effort and are more economical.

### 7. FUTURE DEVELOPMENTS

The future development of human-machine systems is closely related to technological progress, especially to

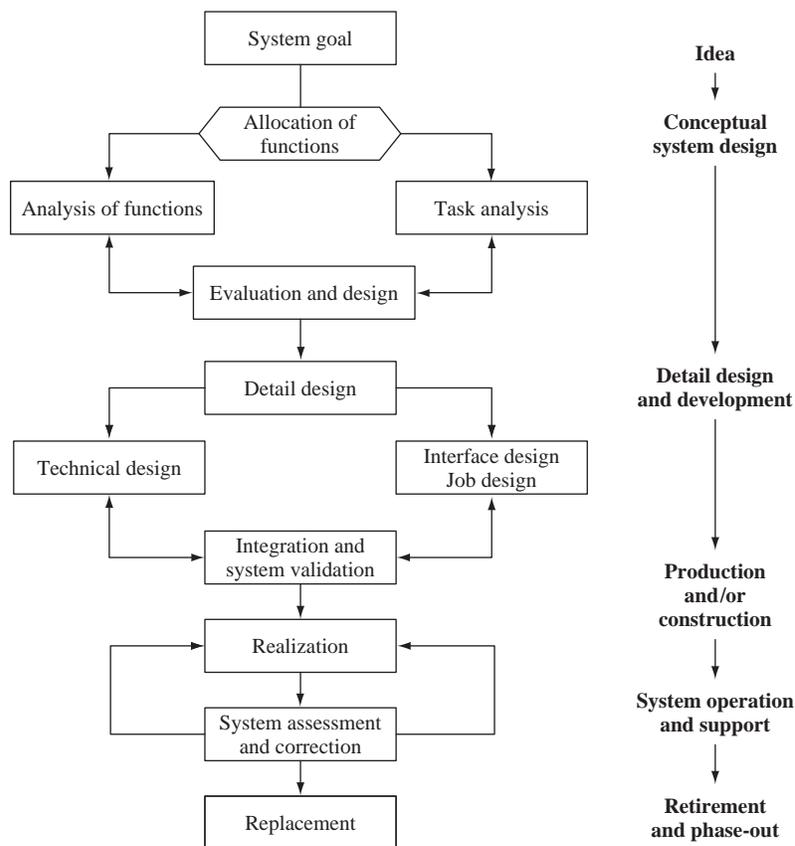


FIGURE 3 Parallel iterative systems design that includes the user.

information technology. The information exchange between human and machine will be further developed to include new interaction concepts and interaction technologies. Auto-stereoscopic perception, as well as the combined interaction of gaze, speech, and hand gestures, has been integrated into multimodal human–computer interfaces that could be applied in task environments with changing cognitive, sensomotoric, and/or attentional demands. New generations of assistance systems will observe the state of the human operator and collect safety- or task-relevant information from the environment to actively support the operator in achieving his or her aims. The development of multimodal interfaces has formulated new research questions for application and design in the field of engineering psychology, for example, about the transmission of concepts in natural communication to human–machine interaction. A special point of research in the design of interaction related to information technology is to provide access to disabled people for better living conditions.

An important concept for the supervision of complex human–machine systems is ecological interface design. This approach attempts to visualize complex processes in technical systems. The easily understandable representation of a system's states should support

operators in situations where safety is critical. Other sustained work domains of engineering psychology consider the increasing amount of distributed human–machine systems throughout nearly all branches, the replacement of leftover allocation with a humanized task approach, and/or the simulation of complex human–machine systems.

### See Also the Following Articles

Attention ■ Motives and Goals ■ Perception and Culture

### Further Reading

- Chapanis, A. (1965). *Man machine engineering*. Belmont, CA: Wadsworth.
- Harris, D. (Ed.). (1997, 1999, 2001). *Engineering and cognitive ergonomics* (Vols. 1–6). Aldershot, UK: Ashgate.
- Hoyos, C. G., & Zimolong, B. (Eds.). (1990). *Ingenieurpsychologie*. Göttingen, Germany: Hogrefe.
- Rasmussen, J., Petersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York: John Wiley.
- Salvendy, G. (Ed.). (1997). *Handbook of human factors and ergonomics*. New York: John Wiley.
- Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.



# Environmental Assessment

Rudolf H. Moos

Department of Veterans Affairs and Stanford  
University, Menlo Park, California, USA

Charles J. Holahan

University of Texas, Austin, Texas, USA

1. Major Domains of Environmental Features
  2. Linkages Among Environmental Features
  3. Influence of Environmental Factors
  4. Conclusion
- Further Reading

## GLOSSARY

*behavior settings* Units of the environment composed of interrelated physical, organizational, and behavioral properties.

*environmental assessment* The conceptualization and measurement of characteristics of environments that influence human behavior.

*institutional policies* Rules and regulations that organize everyday social interaction and the overall functioning of a setting.

*institutional structure* Objective characteristics of organizations such as their size, ownership, and centralization of authority.

*physical and architectural features* Objective characteristics of the natural and built environment, including the landscape, cities, and buildings.

*social climate* The social characteristics and expectations of a setting such as its level of support, task orientation, and structure.

*suprapersonal factors* The aggregate characteristics of the individuals in a setting, including their preferences, abilities, and behavior patterns.

An integrated perspective on human environments is essential for one of the central tasks of applied psychology, that is, to help create optimal and sustainable conditions that enhance individuals' quality of life. At its most basic level, environmental assessment involves the conceptualization and measurement of characteristics of environments that influence human behavior. More broadly, environmental assessment encompasses the linkages among aspects of the environment as well as the processes by which environmental factors influence human behavior.

## 1. MAJOR DOMAINS OF ENVIRONMENTAL FEATURES

The main characteristics of environments can be conceptualized in four interrelated domains: physical and architectural features, institutional structure and policies, suprapersonal factors, and social climate.

### 1.1. Physical and Architectural Features

The assessment of physical and architectural features typically involves direct observation. Four main sets of dimensions have been identified: (a) physical amenities that improve convenience and comfort, (b) physical features that provide support for individuals, (c) size

allowances for personal activities, and (d) physical integration with the surrounding community as assessed, for example, by the extent to which the community and its resources are convenient and accessible.

A related area of research focuses on landscapes and natural environments. The assessment of natural settings involves ratings of perceived environmental quality as well as cognitive mapping techniques that tap people's environmental comprehension. Natural settings have been described in terms of physical attributes (e.g., the presence and extent of hills, water, trees, and grass) and aesthetic qualities (e.g., beauty, harmony, complexity). Cognitive mapping, or the study of people's internal representation of the external world, is especially helpful for understanding the value of environmental cues for orientation and "way finding."

### **1.2. Institutional Structure and Policies**

There are four sets of characteristics in this domain. The first set encompasses structural factors such as ownership, size, and centralization. The second set of dimensions focuses on the balance between individual freedom and institutional order as measured by how much participants individualize their activities and routines and how much they are involved in managing the environment and influencing its policies. The third set of dimensions involves behavioral requirements such as expectations for individuals' functioning and acceptance of problem behavior. The fourth set of dimensions involves the provision of services and activities such as health services, daily living assistance, and social activities. These characteristics can be measured by direct observation or by ratings of participants or outside experts.

### **1.3. Suprapersonal Factors**

When individuals come together in a social group, they bring with them values, norms, and abilities. Because of selective mechanisms, groups draw their members from the general population in a nonrandom manner and produce distinctive blends of these individual characteristics. The aggregate of the individuals in a setting, or the suprapersonal environment, in part defines the subculture that develops in a group.

In the vocational or work domain, Holland showed that occupations can be categorized into six groups (realistic, investigative, artistic, social, enterprising, and conventional), each of which represents a different

personality type. Because people in a vocational group tend to have similar personalities, they are likely to create characteristic interpersonal environments. Thus, there are six types of environments that correspond to the six personality types. In this way, an average background characteristic of a group (i.e., its dominant vocational preference) creates a characteristic environment with unique demands, rewards, and opportunities.

### **1.4. Social Climate**

The social climate is the "personality" of a setting or environment such as a workplace, a classroom or school, a social group, or a neighborhood. Three sets of dimensions characterize the social climate of diverse settings. First, relationship dimensions measure how involved, supportive, and expressive people are in a setting. Second, personal growth dimensions tap ways in which an environment encourages personal growth such as by emphasizing autonomy, achievement or task orientation, competition, intellectual orientation, and moral-religious orientation. Third, system maintenance and change dimensions measure the settings' organization, clarity, control, and responsiveness to change. In general, social climate has been measured by the shared perceptions of the individuals in an environment and by outside observers' ratings.

## **2. LINKAGES AMONG ENVIRONMENTAL FEATURES**

The four domains of environmental features are interconnected. Physical and architectural features, institutional structure and policies, and suprapersonal factors shape each other and combine to create behavior settings with particular social climates. In turn, behavior settings and the social climate influence individuals' health and well-being both directly and indirectly through the other environmental domains.

### **2.1. Behavior Settings**

Behavior settings, which are composed of groups of individuals behaving together (e.g., working on a shared task, participating in a social group, learning in a classroom), have four essential attributes. First, behavior settings have one or more standing patterns of behavior such as sitting, listening, talking, and note taking during a union meeting. Second, behavior settings involve a

physical milieu that may include both natural and constructed features of the environment. Third, these settings have shared policies and expectations about acceptable behavior. Fourth, the physical milieu and standing patterns of behavior are interdependent so that, for example, chairs in a classroom typically face the teacher for listening and observing. The physical milieu and policies help to shape the standing patterns of behavior in a setting. These three characteristics, in turn, tend to create a particular social climate.

## **2.2. Physical Features and Social Climate**

Physical features that improve convenience and comfort, such as physical amenities and social-recreational aids, are associated with a more cohesive and well-organized social climate. Residential settings with more supportive physical features, such as prosthetic aids and safety features, tend to develop a social climate oriented toward independence. In contrast, crowding and lack of adequate space are associated with a lack of cohesion and goal direction as well as with ambiguity and disorganization.

## **2.3. Institutional Structure and Policies and Social Climate**

Ownership and size are related to social climate. Policies in nonprofit facilities are more accepting of problem behavior and more likely to promote residents' personal control. Accordingly, residents in nonprofit facilities see them as more cohesive, self-directed, and better organized than do residents in proprietary facilities. With some exceptions, large size is related to a less cohesive and less well-organized social climate with somewhat more conflict. Policies that enable individuals to determine their daily routine and to influence the operation of the setting, and those that have more provisions for privacy, tend to promote a cohesive social climate oriented toward autonomy and independence. When the rules and regulations in a setting are clear, the social climate is likely to be more cohesive and self-directed.

## **2.4. Suprapersonal Factors and Social Climate**

The modal personal characteristics of individuals in various occupations are associated with the work climates they create. For example, individuals in realistic

occupations, such as engineers and air traffic controllers, tend to see their work groups as high in control and lacking in cohesion, autonomy, and innovation. In contrast, employees in social-type work settings, such as teachers and social workers, tend to report more emphasis on supportive relationships, autonomy, clarity, and innovation.

Individuals' abilities, personal resources, and typical patterns of behavior help to determine the social climates in residential facilities. Facilities with more women and more socially privileged individuals tend to develop more cohesive, self-directed, and well-organized social climates. Facilities with more older and mentally impaired individuals are likely to be less supportive and goal-directed as well as higher on control.

## **3. INFLUENCE OF ENVIRONMENTAL FACTORS**

Each of the four sets of environmental factors has an impact on individuals' health and well-being. In addition, part of the influence of environmental factors depends on the personal competence, orientation, and preferences of the individuals in the environment. In turn, these individuals select and shape the environment they inhabit.

### **3.1. Environmental Impacts**

Physical and architectural features, institutional structure and policies, and suprapersonal factors tend to exert their effects through specific behavior settings and the social climate they create. Most basically, social systems tend to maintain and accentuate individual qualities that are congruent with their dominant aspects. Accordingly, youngsters in families that value independence, achievement, and intellectual and recreational pursuits are likely to develop more self-confidence and social competence. A workplace that values independent decision making, complex and challenging tasks, and high-performance expectations promotes intellectual flexibility and self-efficacy. More cohesive and socially integrated neighborhoods characterized by adult guardianship and monitoring tend to promote responsible parenting and youth behavior.

Some emphasis on each of the three domains enhances positive emotional and behavioral development, but too much focus on any one domain can lead

to distress and dysfunction. Highly achievement-oriented and structured families can create anxiety and erode youngsters' self-confidence. Similarly, high work demands can lessen employees' job morale and elicit depression and physical symptoms. Difficult and competitive learning environments encourage cognitive growth, but they also can erode students' self-confidence and produce high absenteeism and dropout rates.

### **3.2. Person–Environment Matching Models**

Part of the influence of contextual factors depends on the personal competence, orientation, and preferences of the individuals who experience them. The competence–demand model of person–environment congruence states that high environmental demands should have more positive consequences for individuals who are functioning well than for those who are functioning poorly, whereas environmental resources should have more benefit for less competent individuals. Specifically, a self-directed setting that has high performance expectations and relatively little structure should have the most benefit for functionally able individuals. In contrast, less competent individuals need more support and structure and often find high performance expectations and self-direction to be disruptive. As individuals' cognitive and psychosocial skills improve, they should be able to cope with more demanding and less structured environments.

The conceptual-level model describes people in terms of their cognitive complexity and describes environments in terms of their structure. It posits that individuals at a high conceptual level are able to organize their own environment, whereas individuals at a lower conceptual level need the stabilizing influence of a well-structured setting. The model also proposes that individuals with less cognitive complexity are more affected by changes in structure than are those at a more mature level.

### **3.3. Transactional Models**

Most of the research in this area reflects a social causation perspective in which the environment is seen as a causal factor that shapes individuals' development and maturation. However, there also is a social selection process by which individuals select and shape existing environments and create new environments. According to behavior setting theory, four transactional processes

explain how ongoing interactions among participants and other aspects of behavior settings produce a stable patterned set of behaviors and environment–behavior congruence. These processes involve goal circuits (i.e., how participants set goals and achieve them), program circuits (i.e., how participants control and organize the program or agenda of a setting), deviation-counteracting circuits (i.e., the ways in which unwanted modifications or problems in the setting are altered or corrected), and vetoing circuits (i.e., the ways in which deviant components of a setting are eliminated).

More generally, there are four main processes by which individuals and environments influence each other. In a passive process, individuals find themselves or are placed in an environment that shapes their behavior, for example, when a child is placed in a group care home. In an evocative process, individuals select specific responses that become part of their environment, for example, when a youngster with a sunny disposition elicits a supportive context. In an active process, individuals select social contexts that maintain and accentuate their dispositions or skills, for example, when a youngster joins a soccer team or a debating club. Finally, in a transcending process, individuals build entirely new physical and social contexts, for example, when individuals bereaved and maimed in a cataclysmic disaster rebuild their community.

## **4. CONCLUSION**

Environmental assessment provides conceptual and empirical tools that make environmental determinants of human behavior accessible to psychological inquiry. Researchers have identified four main domains of environmental features, assessed specific dimensions in these domains, and examined their interconnections. Conceptual models of how contexts influence individuals, of person–environment matching, and of how individuals select and create contexts have led to a better understanding of how the interplay of environmental and person factors shapes individuals' health and well-being. Environmental assessment complements the traditional person-centered focus in psychology by providing a more holistic and integrated understanding of human behavior as it occurs in real-world settings.

### **See Also the Following Articles**

Environmental Design and Planning, Public Participation in  
 ■ Person–Environment Fit

### Further Reading

- Barker, R. G., & Schoggen, P. (1973). *Qualities of community life*. San Francisco: Jossey-Bass.
- Bechtel, R. B., & Churchman, A. (Eds.) (2002). *Handbook of environmental psychology*. New York: John Wiley.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Friedman, S. L., & Wachs, T. D. (Eds.) (1999). *Measuring environment across the life span: Emerging methods and concepts*. Washington, DC: American Psychological Association.
- Golledge, R. G. (Ed.) (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes*. Baltimore, MD: Johns Hopkins University Press.
- Holland, J. (1992). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Moos, R., & Lemke, S. (1994). *Group residences for older adults: Physical features, policies, and social climate*. New York: Oxford University Press.
- Walsh, W. B., & Craik, K. H. & Price, R. H. (Eds.) (2000). *Person-environment psychology: New directions and perspectives* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.



# Environmental Design and Planning, Public Participation in

Arza Churchman

Technion-Israel Institute of Technology, Haifa, Israel

Elisheva Sadan

Hebrew University of Jerusalem, Jerusalem, Israel

1. Introduction
2. Issues Bedeviling Participation
3. The Role of Environmental Psychology in the Public Participation Arena  
Further Reading

## GLOSSARY

**civil society** The part of social life that lies beyond the immediate reach of the state and that must exist for a democratic state to flower; it is the society of households, family networks, civic and religious organizations, and communities that are bound to each other primarily by shared histories, collective memories, and cultural norms of reciprocity.

**community** A group or network of people who share something in common or interact with each other; it is a place where solidarity, participation, and coherence are found.

**distributive justice** The equitable distribution of resources and nuisances among groups within the society differentiated by factors such as power, social class, ethnicity, and so forth.

**empowerment** The process by which people achieve more control (perceived and actual) over their lives, destiny, and environment; it is the process of overcoming powerlessness and its outcomes (e.g., despair, marginality, estrangement).

**environment** An all-encompassing term that includes all aspects of the world—the physical, ecological, social,

economic, cultural, political, institutional, technological, and individual.

**participation/involvement** Situations in which individuals (or groups), who are not elected or appointed governmental officials and who are not professionals working for such officials, voluntarily take part in decision-making processes to influence these processes, change them, or improve them; participation is the activity of such individuals, whereas involvement is the activity of the officials and professionals who initiate such a process.

**participatory democracy** A democratic system in which citizens have opportunities for taking part in governmental decision-making processes relevant to their lives on a regular basis and not just in elections once every 4 years.

**participatory governance** The transfer of authority and responsibility from those who hold power by virtue of law, contract, or organizational role to those not so empowered; this is a step above participatory democracy because it makes explicit the authority and responsibility given to the participants within the governmental procedure.

**procedural justice** A system in which the decisions as to the distribution of resources and nuisances are made in an equitable manner.

**social capital** The bonds of the community that enrich people's lives in a variety of ways; it includes resources that inhere in social relationships such as mutual trust, a sense of reciprocal obligations, and civic participation aimed at benefiting the group or community as a whole.

**sustainability** The goal for development that recognizes the interwoven nature of economic, social, and ecological–environmental factors and that strives for intergroup and intergenerational equity.

**sustainable development** The process of balancing three development processes: economic, community, and ecological; it is development that meets the needs of the current generation without compromising the ability of future generations to meet their own needs.

Participation relates to people taking part in public decision-making processes or private initiatives with public implications. The recent movement in planning and design toward more participatory and empowering decision-making processes is a significant step in the direction of accommodating the needs and preferences of different groups of people. Introducing individuals into the decision-making process forces politicians, planners, and architects to consider the micro level and to meet the people, see their faces, and hear their voices. The interest in such processes within environmental psychology forces one to consider the macro level and to understand the decision-making system and the constraints and pressures within which planners operate. This article addresses various issues that reflect the complexity of the participatory process, the discussion of which also illustrates the interdependence of these issues. Finally, the article looks at the role of environmental psychology in the public participation arena.

## 1. INTRODUCTION

- “I don’t need to involve the public. I know what’s good for them.”
- “They didn’t like the Eiffel Tower at first either.”
- “Of course I’m in favor of public participation . . . but . . . the public isn’t ready, or they don’t understand, or it takes too long, or it costs too much.”
- “Who are you to claim that you represent the public?”
- “Why aren’t there more people here?”
- “We’ve been working on this plan for you for 3 years, and now when it’s ready you’re against it!”
- “You’re only interested in your own backyard and your own selfish interests, while I’m concerned with the larger picture and the common good.”

These are quotes of statements made by elected officials and professionals with regard to the principle of public

participation in decision making. They illustrate some of the problems encountered by those who are interested in facilitating or promoting the principles of public participation, whether they are professionals, elected politicians, or members of the public themselves.

Participation, in the sense that the term is used in this article, relates to people taking part in public decision-making processes or private initiatives with public implications, not those occurring within an organization or a group of people. Arguments over what should and should not be considered participation illustrate both the ambiguity and the political nature of the concept. For example, there are those who talk about participation in terms of voting in elections, and others who see it in terms of taking part in groups of various kinds. It is clearly a value-laden concept, which for some represents the very essence of democracy, a basic human right, and an ethical good in and of itself. Others value participation as a pragmatic means to various ends, some instrumental and some strategic. Participation is seen as having the potential for major benefits for the individual, for the community, and for society as a whole; as enabling decisions that are better able to fit the needs of the people; and as making support for policies and programs more probable.

Because participation has implications for the nature of the distribution of power among individuals and groups, it arouses tensions and emotions, and this makes it difficult to relate to participation dispassionately. On the whole, politicians, officials, and professionals see it as a threat to their power and as a bother. They argue that ordinary people are not capable of being part of the decision-making process due to the complexity of the issues at hand. They also argue that they see the larger picture, and make their decisions “objectively,” for the greater good. Residents, on the other hand, are “subjective” and concerned only with their own private good.

The concept of public participation appears in the writings of philosophers, theoreticians, and practitioners in the fields of political science, sociology, planning, architecture, landscape architecture, community psychology, community social work, and environmental psychology. Because one of the basic tenets of environmental psychology is the importance of identifying the context of a phenomenon for understanding it, the next few paragraphs are devoted to identifying the historical, political, intellectual/theoretical, practical/professional, cultural, and psychological contexts of public participation.

Participation is a multifaceted and multidisciplinary project, with theoreticians mainly to be found in political science and planning. The current discourse surrounding participation is rooted in the theories of Habermas and Giddens that have been further developed in the planning context by Healey and Friedmann. Freire has had an important influence on the thinking in this area as well. Recently, Putnam's theory of social capital has also found its way into the discussion of the role of the civil society and of participation in current democracies. There is also a body of knowledge about participation within community psychology, both theoretical and empirical. On the other hand, one can find professionals actively involving others within the fields of community organization, architecture, planning, and environment-behavior studies, including environmental psychologists. There are private design offices that specialize in participatory work (particularly in the United States), and there is growing interest in participatory research and participatory evaluation.

Attempts at involvement and at participation appear within rather varied contexts with different professions involved, depending on the context and often with different motivations and methods, with regard to environmental quality issues, health, housing, neighborhoods, schools, public spaces, and community development in general. United Nations (UN) agencies and documents, such as the Convention on the Rights of the Child, Agenda 21, Habitat 2, and the Healthy Cities Program, strongly recommend (and sometimes require) involvement of the public. In the European Union, the Aarhus Convention emphasizes public participation in decision making as a necessary component of governmental activity in the area of environmental policy. An interesting phenomenon during recent years is the promotion, or even requirement, of public participation by international agencies, such as the World Bank, operating in developing countries. Public participation appears to be the recent offspring of community development, which was a UN strategy for "developing" those that were then called Third World countries. That approach to development occasioned much intellectual and academic criticism due to its poor outcomes in those countries, and it would seem that a more empowering involvement approach is the new consensus. Because a democratic system is the basic necessity for the right to participation in that it requires inclusiveness, transparency, and accountability on the part of those in positions of power, attempts to apply its principles in nondemocratic systems are limited to topics that are perceived as not threatening the system as a whole.

The development over time in the approach to participation can be seen in the changes that have occurred in the language used to discuss the ideological bases for it—from democracy, participatory democracy, and power to justice (both distributional and procedural), sustainability, empowerment, social capital, and participatory governance. These changes in language reflect changes both in the theoretical underpinnings of the concept and in the degree to which it is embedded within public decision-making processes. These reflect the change from the 1960s "modern" and positivistic view of the professional and the needs-based participation process to the "postmodern" times of reflective practitioners and the Foucauldian thinking that knowledge is not where (and what) it was traditionally believed to be and that power must be shared or at least viewed with suspicion.

In many countries and cities, resident involvement now appears in the language of master plans as a basic requirement. For example, this is the case in Australia, virtually all of Europe (but particularly Great Britain and The Netherlands), Canada, the United States, and Israel. In these and other cities and countries (e.g., Brazil, Mexico, Venezuela, India, South Africa), there are examples of participatory processes of greater or lesser scope. Participatory processes have been conducted with adults and children, with men and women, with poor people and middle-class people, and with people with and without formal education.

Furthermore, "in the beginning" there was involvement, and even though the term used might have been participation, most of the discourse in the planning and architecture fields was about the activities of the professionals and how they should or could involve others. Whereas once the discourse of public participation was dominated by professionals who were "doing it" and the voice of the participants was never heard, there is much more recognition now that professionals are not responsible for the participation but rather are responsible for effectively involving people in processes that, until recently, professionals have controlled on their own. Thus, there is growing awareness today that the participants, their motivation, and their concerns should be more in focus.

To use an image suggested by Kidder and Fine in 1987, the lens used in planning is to a large extent a zooming-out one, whereas the lens used in environmental psychology is a zooming-in one. Public participation was usually discussed through the zooming-out lens, and this article zooms in on it. The movement in planning and design toward more participatory and

empowering decision-making processes is a significant step in the direction of accommodating the needs and preferences of various groups of people. Introducing individuals into the decision-making process forces politicians, planners, and architects to consider the micro level and to meet the people, see their faces, and hear their voices. The interest within environmental psychology in such processes forces one to consider the macro level and to understand the decision-making system and the constraints and pressures within which planners operate.

On a very basic level, many of the assumptions of environmental psychology should lead to the conclusion that participation is the only way, or the best way, in which to achieve the desired “fit” between the environment and its users. These assumptions include (a) that the physical environment has implications for people’s lives; (b) that people are different and have different needs and, therefore, that one cannot specify, identify, or posit one universal model of the person–environment relationship; (c) that people are active and not passive in the sense that they interpret, evaluate, and use their environment in ways that they desire or are able; (d) that freedom of choice is an important element in people’s behavior; (e) that aggregates (or groups) that have some needs and characteristics in common can be identified, for example, by age, gender, health status, socioeconomic status, and cultural background; and (f) that the purpose of the environment is to afford opportunities for achieving each individual’s own subjective definition of quality of life. The fact that participation is not yet seen as a *sine qua non* of environmental psychology is an unfortunate circumstance.

## **2. ISSUES BEDEVILING PARTICIPATION**

This section addresses various issues that reflect the complexity of the participatory process, the discussion of which also illustrates the interdependence of these issues.

### **2.1. Why Involve Others and Why Participate Oneself?**

If participation is seen as an end in itself, perhaps the question of what it is good for is unnecessary. However, if participation is seen only (or also) as a means to other ends, major questions arise as to whether it indeed can

achieve (or does achieve) these ends. Related to this question is the distinction between the process of involvement and/or participation and the products and/or results of this process. For example, one could look at how many people participated or who they were. One could look at whether participants improved their skills and abilities, whether there was individual or community empowerment, or whether there was a change in the level of trust between the involvers and the participants. Then, one could look at the product of the decision-making process. Is it a “better” one (and by whose criteria)? Did it lead to an improvement in the quality of the environment? There are a number of issues pertinent to this distinction. Which is more important, process or product? For whom is it more important, involvers or participants? Under what circumstances? How does one measure and evaluate the elements of the process or the results (e.g., quantitatively or qualitatively, by effectiveness or by efficiency, subjectively or objectively)? How does one determine that particular results are directly related to the process?

Another question that arises is whether the adherents of participation are too ambitious or unrealistic in their expectations and goals. Are their assumptions, theories, or hypotheses ones that can be achieved? Just as it has been argued that architects and planners should not make deterministic statements that promise specific results from their plans, it might not be appropriate to promise too much from a participatory process and raise expectations that might not be possible to achieve. When this happens, it may doom the project of participation to disappointment and rejection, both by the participants and by the involvers. It would seem that what is needed is (a) a careful analysis of what is possible to achieve under what circumstances, (b) a serious effort at implementing the process in a manner that has the potential for achieving those results, and (c) research and evaluation that will enable one to learn what does and does not work. On the other hand, the participants must understand and accept the fact that they might not be able to achieve all that they wish to achieve, that compromise might be necessary, and that consideration for the views and needs of others is an essential part of the process.

If one focuses the lens on the participants themselves, one can ask whether it is enough if the results are mainly on the psychological level, which is also the hardest level to demonstrate. If results happen on the behavioral level, such as attendance at meetings, how does one know that there are also results, such as personal empowerment, on the psychological level? To

be sure that the claims for the results of participation are valid, one must develop tools for evaluation that include indicators and “signs” of the psychological, educational, and environmental impacts or outputs of participation.

To paraphrase the words of Schorr, to be more assertive about what one knows, one must first be more systematic in assembling and trying to understand what one knows. One will not disappoint if one bases his or her recommendations on “what works” in its specific context. If one is systematic about documenting participatory processes—and the works of Fung, Chawla, Couto, Horelli, Sanoff, Schneckloth, Schorr, and others show that this is the trend now—one will be able to say more confidently, with accumulating evidence, what results are caused by the process. The accumulation of research results pointing in the same direction helps to build confidence in the hypotheses.

## **2.2. Why Is Involving Others So Difficult?**

Many of those who discuss participation or practice involvement are unaware of the complexity of the concept and of the variety of ways in which it can be implemented. A one-dimensional approach to involvement, such as public hearings as the sole medium, ignores the various possible goals for involvement and participation and the differences among the participants themselves. This article has already noted that some see participation as an end in itself, whereas others see it only (or also) as a means to other ends. Some of these other goals focus on the participants and on the manner in which they can achieve personal and group change through participation. Other goals focus on the decision-making process and the manner in which participation can change this process and its results. Different professions focus on different goals, although it is important that all professions are cognizant of all such goals. For example, environmental psychologists and architects and planners may focus mainly on the goal of achieving environments that better fit the needs and preferences of the various users, whereas community psychologists may focus more on the manner in which the participatory process leads to more community cohesiveness or personal skills. However, if the former ignore the fact that participants have personal and group motivations that may drive their willingness to participate, and if the latter ignore the fact that participants are also concerned with other kinds of tangible results, both

groups may fail to achieve their goals. Until now, those working in this area have not really examined systematically whether individual goals are realized. One reason is that they are too busy doing participation to do this. Another reason is that one faces the dilemma of how to “prove” that there is a cause-and-effect relationship. To gain the trust of the people, one must be honest about the constraints within which the process will be operating and be careful not to raise expectations beyond what is likely to be possible.

Part of the problem stems from a lack of understanding of the complexity of participatory processes, that is, that there is not one way that is appropriate for every context or situation. Furthermore, many lack knowledge as to how to work with people; working with people is perceived as something that anyone can do easily. If one really wishes to involve people, one must make sure that he or she has the necessary skills of listening, knowing how to talk to lay people, honesty, perseverance, and understanding of people’s needs and interests. It must be a process in which people will be able to participate and will want to participate and that will allow people to make decisions on questions that are meaningful to them. One must demystify the professional process, avoid jargon, and exemplify the essence of better communication, namely simplicity, clarity, and adaptability.

## **2.3. Who Is the Public?**

The issue of how to define the public concerns the involvers, who wish to have some assurance that those participating “represent” the public at large. Considering the fact that every “public” contains many different groups with diverse and sometimes conflicting interests, this is not a trivial demand. It raises the issue of the possible conflict between what is called the common good and the self-interest of individuals or groups. There are a number of answers to these questions. One is that there is no objectively defined common good; it is always a value judgment as to which interests to prefer over others. Another answer is that self-interest is not necessarily “selfish-interest.” Recognizing self-interest can be seen as helping to advance distributive justice. The more groups and interests participate in the deliberations, the more differences can be clarified and conflicts can be addressed. Still another answer is that those who participate can be viewed as informal representatives of those who do not attend, particularly if serious efforts

have been made to inform all of those relevant to the topic of the possibilities for participation.

Although the participants themselves might not be particularly concerned about this issue of representativeness, it does affect the degree of legitimacy of their claim to attention and their ability to influence the process. Some address the issue in quantitative terms (e.g., how many attend a particular event or process), whereas others do so in more qualitative terms (e.g., how varied or how serious those who participate are).

In a situation where there are professionals or decision makers who initiate the participatory process, it is incumbent on them to make sure that there are no initial stumbling blocks to the participation of as many people as possible (e.g., meetings at inconvenient times or in inconvenient places, notices that people cannot read or understand). In some situations, this requires preparatory work within the neighborhood or surroundings so as to raise people's awareness of and interest in the problems being addressed. The process itself must be a welcoming one that speaks to people in ways that they understand, that does not condescend or mislead, and that enables the empowerment of the participants. People must be provided with the information relevant to the questions being addressed, both those questions considered to be relevant by the involvers and those considered to be relevant by the participants.

On the other hand, those who are to be affected by the decisions taken must be willing to be part of the process, to understand that it is not only their right but also their responsibility to take on the role of participants. It is to their advantage to do so in terms of the personal and social capital that they can accrue and in terms of the manner in which they can work toward changing both the system and its decisions to better reflect their needs. However, such participation requires skills that not everyone has, and it may be necessary to help participants to acquire those skills and, in so doing, enhance their self-confidence and self-efficacy. Furthermore, it must be accepted that there are members of the public who are not interested in the topics or in the process and that this also is their right. Some people have such serious problems of various types that they have no time or energy to devote to other interests.

In other circumstances, there are cases where it is the public (or parts of it) whose members demand to be part of the decision-making process. They too have a duty to create a democratic process within their group and to work toward the inclusion of people with different characteristics and different ideas. In this way, they

will have a stronger case when they are confronted with the inevitable questions of "Who are you?" "Why should we listen to you?" and "Who chose you?"

#### **2.4. Whose Values Are Primary?**

Very often in participatory processes, there is conflict between the values of the professionals and decision makers and those of the public. The former, particularly in the environmental quality area where some of the issues seem to be more clear-cut, believe that they cannot leave the decisions up to the "ignorant" public. With these environmental issues, there is an agenda that they want to push where there are ostensibly more objective kinds of issues with which to deal, and the professionals are convinced that they have the answers. The fact that people are likely to mistrust the statements and pronouncements of politicians and professionals makes their task much more difficult. One way in which to gain (or regain) trust is to conduct a sincere participatory process.

However, there is a general value that should be recognized—one that agrees to disagree and that understands that the dialogue itself is important. The involvement/participation process should acknowledge the conflicts among various groups and, instead of denying them and deciding that the politicians and professionals will have their way each time, should encourage the democratic participation of others involved in the decision-making process and enable the expression of the differences of opinion that exist in the social arena. This is the whole purpose of what Couto called "making democracy work better." Rather than a question of whose values are primary, it is a question of representing the diversity of values involved in the issue. Sandercock called this working for a society in which differences can flourish based on principles of social justice, of multiple citizenships, of heterogeneous publics, and of coalitions building bridges across differences.

#### **2.5. How Does One Involve Others?**

Currently, there are a number of excellent sources that present many different techniques for involving the public, some more detailed than others but all extremely valuable for illustrating the variety of techniques available and thereby expanding the ways in which participation is viewed. These include, but are not limited to, the works of Chambers, Driskell, Horelli, Sanoff, Sarkissian and colleagues, and Wates. Many of

these publications are useful not only in presenting the techniques but also in spelling out the pitfalls to avoid in involvement processes. In addition, there is growing use of geographic information system (GIS) and Internet techniques, although limitations on their use with groups whose members might not have access to computers or the Internet must be taken into account, just as the use of plans or models raises questions as to whether or not nonprofessionals understand them.

### 3. THE ROLE OF ENVIRONMENTAL PSYCHOLOGY IN THE PUBLIC PARTICIPATION ARENA

The contribution of environmental psychology to the discourse and practice of public participation in decision making lies in a number of directions and stems from theoretical developments and practical experience. On the theoretical level, environmental psychology contributes a number of basic tenets, including (a) the definition of environment as an all-encompassing term, (b) the emphasis on context and on the relationship between sociophysical aspects of the environment and human behavior and attitudes, (c) the knowledge that both physical and social conditions can be acted on rather than accepted fatalistically as fixed and unchangeable, and (d) the recognition of the differences in behavior and attitudes among different groups of people and the importance of this for the basic philosophy of participation. The research is based on notions that are also basic to participation—that there is no one best solution to a design problem, that expert decisions are not necessarily better than lay decisions, and that a design or planning task can be made transparent. Environmental psychologists believe in the principle that the environment works better if it is designed and planned in such a way that it fits the needs of the people affected by its changes. Environmental psychologists involved in participation add to this the condition that these people are actively involved in its creation and management. Their role is to focus the spotlight on the people themselves—on participation rather than involvement and on the people and their lives rather than on abstract concepts.

In more specific terms, the research contributes an understanding of, for example, how people perceive environmental problems, what the determinants and predictors of environmental concern are, attitudes toward risk and perceived lack of control, and the relative importance

of different environments for different groups. The latter point is particularly relevant given the finding by many that people are more concerned with immediate and local issues than with wider and more abstract ones. Environmental psychologists have experience in research that focuses on real-world complex issues, the kind of research that is essential when it comes to researching participatory processes. This focus on research, if applied to the phenomena of participation, will enable the examination of what does and does not work and thereby further the paradigmatic shift that participatory processes entail. However, this will require that more environmental psychologists make that paradigmatic shift themselves, both in their research and in their work within design, planning, and policy decision-making processes.

### See Also the Following Articles

Decision Making ■ Environmental Assessment ■ Residential Preferences and Attachment across the Lifespan ■ Restorative Environments

### Further Reading

- Braydon-Miller, M., & Tolman, D. (1997). Transforming psychology: Interpretive and participatory research methods [special issue]. *Journal of Social Issues*, 52(4).
- Chambers, R. (2002). *Participatory workshops*. London: Earthscan.
- Chawla, L. (Ed.). (2002). *Growing up in an urbanising world*. London: Earthscan.
- Churchman, A. (2002). Environmental psychology and urban planning: Where can the twain meet? In R. Bechtel, & A. Churchman, (Eds.), *Handbook of environmental psychology* (pp. 191–200). New York: John Wiley.
- Couto, R. A. (1999). *Making democracy work better: Mediating structures, social capital, and the democratic prospect*. Chapel Hill: University of North Carolina Press.
- Driskell, D. (2002). *Creating better cities with children and youth: A manual for participation*. London: Earthscan and UNESCO.
- Freire, P. (1985). *The politics of education*. South Hadley, MA: Bergin & Garvey.
- Friedmann, J. (1992). *Empowerment: The politics of alternative development*. Cambridge, UK: Blackwell.
- Fung, A., & Wright, E. (2003). *Deepening democracy. Institutional innovations in empowered participatory governance*. London: Verso.
- Healey, P. (1997). *Collaborative planning*. London: Macmillan.
- Horelli, L. (2002). A methodology of participatory planning. In R. Bechtel, & A. Churchman (Eds.), *The handbook of environmental psychology* (pp. 607–628). New York: John Wiley.

- Kidder, L., & Fine, M. (1987). *Qualitative and quantitative methods: When stories converge* (New Directions for Program Evaluation, No. 35). San Francisco: Jossey-Bass.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.
- Sadan, E., & Churchman, A. (1997). Process focused and product focused community planning: Two variations of empowering professional practice. *Community Development Journal*, 32(1), 3–16.
- Sandercock, L. (1998). *Towards cosmopolis: Planning for multicultural cities*. Chichester, UK: Wiley.
- Sanoff, H. (2000). *Community participation methods in design and planning*. New York: John Wiley.
- Sarkissian, W., Cook, A., & Walsh, K. (1997). *Community participation in practice: A practical guide*. Murdoch, Australia: Murdoch University, Institute for Science and Technology Policy.
- Sarkissian, W., & Perlgut, D. (Eds.). (1988). *The community participation handbook: Resources for public involvement in the planning process* (2nd ed.). Murdoch, Australia: Murdoch University, Institute for Science and Technology Policy.
- Schorr, L. B. (1997). *Common purpose: Strengthening families and neighborhoods to rebuild America*. New York: Anchor Books.
- Wandersman, A., & Florin, P. (2000). Citizen participation and community organizations. In J. Rappaport, & E. Seidman (Eds.), *Handbook of community psychology* (pp. 247–272). New York: Kluwer Academic/Plenum.
- Wates, N. (Ed.). (2000). *The community planning handbook*. London: Earthscan.



# Environmental Psychology, Overview

Mirilia Bonnes and Giuseppe Carrus

University of Rome "La Sapienza," Rome, Italy

1. The Historical and Conceptual Backgrounds of Environmental Psychology
2. The Main Areas of Interest of Current Environmental Psychology
3. The Future of Environmental Psychology  
Further Reading

## GLOSSARY

**affordances** Characteristics of environmental objects that are related to both their physical properties and the possible actions that each perceiver can perform with them.

**conservation behaviors** Actions that contribute to environmental preservation and/or conservation.

**environmental stressors** Physical characteristics of the environment that produce stress.

**Hawthorne effect** An increase in work productivity deriving from a worker's awareness of being monitored by an external observer.

**participatory governance** The transfer of authority and responsibility from those who hold power by virtue of law, contract, or organizational role to those not so empowered.

**personal space** The emotionally tinged zone around the human body that people feel is "their" space.

**place** The product of physical properties of the environment, people's cognitions of the environment, and people's actions in the environment.

**place attachment** An affective bond between an individual and a particular place that is not interchangeable with another with the same functional quality.

**privacy regulation** The selective control of access to the self or to one's group.

**residential satisfaction** The experience of pleasure or gratification deriving from living in a specific place.

**restoration** The process of renewing, recovering, or re-establishing physical, psychological, and social resources or capabilities diminished in ongoing efforts to meet adaptive demands.

**social dilemma** A situation of intrinsic conflict between the pursuing of individual gains and the maximization of collective outcomes.

**spatial cognition** The internalized reflection and reconstruction of space in thought.

**sustainable development** Development that meets the needs of present generations without compromising those of the future.

**territory** A fixed geographical space marked and defended by an organism and used for life-sustaining activities.

Environmental psychology is that branch of psychology that studies the relationship between people and the sociophysical features of the built and natural environment, in order to enhance human well-being and to improve people–environment relations. It emerged as an autonomous field of scientific inquiry at the end of the 1950s and during the 1960s. The main

conceptual roots of environmental psychology can be found in various theoretical proposals within psychology, all of them underlining the ecological or transactional basis and the context-specificity of human psychological processes. The emergence of environmental psychology is also related to the need for responding to specific demands from other technical and disciplinary fields, such as architecture and natural-ecological science, that face the problem of designing and managing the physical features of people's everyday environments. Due to the influence of these different factors, present-day environmental psychology can be characterized by four distinctive aspects: (1) the attention paid to the physical characteristics of the environment where human behavior occurs; (2) the wide variety of the research methods adopted; (3) the interest in problems with a clear social relevance; and (4) the interdisciplinary collaboration with other environmental fields. The first part of this article briefly outlines the basic historical and conceptual backgrounds of environmental psychology as a scientific discipline. Then, the main areas of interest and the related outcomes of current environmental psychology are briefly reviewed. (These aspects will be more thoroughly discussed within each of the single articles included in the present section.) Finally, perspectives on the future developments of environmental psychology are summarized.

## **1. THE HISTORICAL AND CONCEPTUAL BACKGROUNDS OF ENVIRONMENTAL PSYCHOLOGY**

The starting date of environmental psychology as an autonomous disciplinary field within psychological science is commonly placed at the end of the 1950s. The environmental psychology that was formed during the 1950s and 1960s focused its attention on the physical features of the environment in which human behavior occurs. Its aim was to better understand the relationship between human behavior and the everyday physical or sociophysical environment. This environment was considered to be directly perceptible through the sensory organs, and was defined and considered in spatial and physical terms, whether built or natural, or on a small or large scale. At that time, there were two contrasting ideas about the relationship between human behavior and the physical environment. On the one hand, human behavior was conceived of as the "result" of the physical

environment, particularly when the built-up physical environment was considered. On the other hand, human behavior was conceived of as a "cause" of the physical environment, particularly when the natural environment was considered. However, this distinction was increasingly overcome by environmental psychology's growing emphasis on the necessity of adopting a transactional approach in the study of person-environment relationships. This approach led to the person-in-place as the main unit of analysis for environmental psychological inquiry.

Various converging factors, originating inside, around, and outside the psychological field, contributed to the emergence and development of environmental psychology. The origin, the past, and the present of this branch of psychology can be better understood by looking at all these factors in order to outline a disciplinary identity that goes beyond the general label of applied psychology. Because of the influence of these different factors, present-day environmental psychology can be characterized by four distinctive aspects: the attention paid to the physical characteristics of the environment in which human behavior occurs; the wide variety of the research methods adopted; the specific interest in problems with a clear social relevance, and the adoption of interdisciplinary collaboration practices with other environmental fields devoted to the design and management of our everyday life environment.

### **1.1. Environmental Psychology within and around the Psychological Tradition**

Psychology has been traditionally interested in environment-behavior interactions in a very general way. The major interest of environmental psychology rested on the discovery of the importance of the physical and spatial dimension of the environment as a constituting part of human actions and experience. This interest emerged within a wider ecological or transactional perspective coming from various domains of psychological inquiry.

Therefore, environmental psychology has always been concerned for the spatial-physical properties of the surroundings in which human behavior occurs, that is, of its physical setting. At the same time, it stressed the importance of considering these physical properties in a "molar" rather than "molecular" sense. It is also important to stress that the influence of the physical environment on human psychological processes often remains outside our individual and collective awareness.

Some pioneering studies in psychology highlighted the importance of the physical and spatial context in shaping human behavior, albeit often as an incidental outcome or a small part of other research aims. An example is the widely cited field experiments, conducted at the Western Electric Company by Elton Mayo in the 1930s, that investigated the effects of the lighting conditions in work settings upon workers' performance. Other pioneering works are the studies on the development of social influence networks by Leon Festinger and colleagues and the analysis of the "stream" of human behavior in natural settings by Roger Barker and colleagues. All of these earlier contributions were guided by a common methodological interest in studying human behavior in its natural setting, as researchers recognized the need for overcoming the usually low external validity of the traditional laboratory experiments. To this end, they preferred methods such as field experiments or non-obtrusive observation.

Other authors also played a crucial role in the emergence of environmental psychology, especially those who were more open to receive and develop ideas coming from areas that bordered on psychology, but that were traditionally interested in studying behavior in natural contexts. Areas such as cultural anthropology, about human proxemics, animal ethology, and microsociology are some examples. In other words, in order to be concerned with the spatial-physical environment, psychology had to get out from its habitual setting (i.e., the research laboratory), which was by definition a "non-environment." Just like as the earlier pioneers mentioned previously, these contributors also were generally opposed to the laboratory methods used by mainstream psychological research, and consequently were more willing to use other methodologies such as field experiments and field observations, both natural or systematic. R. Sommer and I. Altman's studies on personal space and social behavior, which remain as cornerstones of the early environmental psychology, are good examples.

The enthusiasm over the emergence of this new field of inquiry was then linked with the previously mentioned dissatisfaction for both the poor ecological validity and the low social relevance of much laboratory research. This resulted in a search for a "real world psychology." This frequent dissatisfaction can be traced to the various forms of ecological demand raised since the 1940s and 1950s by various authors and psychological schools, which later developed into what has been called "contextualism" or "contextual revolution." This revolution is certainly at the core of the development of environmental psychology, particularly in its

transactional-contextual approach, which has characterized environmental psychology since its beginning.

Initially, the increased awareness of the crucial effect played by the physical features of the everyday environment on human behavior was based on two main theoretical psychological traditions. The first is the psychology of perception, especially in its more ecological orientations, such as the lens model by E. Brunswik, the transactional school of the Princeton group, and the ecological approach to perception of J. Gibson. Gibson, for example, introduced the neologism affordances, which identifies the physical properties of environmental objects that are related to the possible actions that can be performed with them. The second psychological tradition is based on the social psychology approach, through the pioneering work of K. Lewin, E. Tolman, R. Barker, and U. Bronfenbrenner. It embraces a more "holistic" or "molar" perspective, which later developed into the transactional-contextual approach to person-environment relationships, as systematically outlined by many contributors to the first *Handbook of Environmental Psychology* edited in 1987 by D. Stokols and I. Altman. In this perspective, the physical environment or physical setting has been increasingly considered as a sociophysical environment, with a growing emphasis on the social aspects of both the physical environment and the psychological processes involved. The notion of place, and its related environmental-psychological processes, also became a central unit of analysis for many environmental psychologists. Typically, places were defined from an environmental psychology point of view as a product of three main dimensions: physical properties, people's cognitions, and people's actions.

## 1.2. Environmental Psychology and Other Environmental Fields

The problem-oriented demands rising in technical and disciplinary fields distant from psychology are also an important factor that contributed to the emerging of environmental psychology. Examples of these fields are architecture, engineering, urban planning, human geography, natural and ecological science. Typically, architects and engineers are concerned with problems regarding the relationships between people and the built or "human-made" environments. In contrast, geographers and natural scientists are more interested in the relationships between people and the natural features of the environment.

In architecture and urban planning, those who were dissatisfied with an egocentric approach to design desired to move toward a user-centered approach to design, as well as to move from “product” design and planning to “processes” design and planning. A seminal work in this area is that of the urban planner Kevin Lynch at the Massachusetts Institute of Technology. His famous book *The Image of the City*, published in 1960, is considered a cornerstone of the collaboration between urban planning and environmental psychology. Lynch proposed that the point of view or the image that users form of the urban environment should be kept in mind when designing and planning urban spaces. Likewise, in engineering and technology, many scholars and practitioners became interested in the human-use dimension of technological systems.

The collaboration between psychology and architecture was mainly guided by an interest in the influence of specific and localized spatial characteristics of the built-up environment on human behavior. A different orientation, however, characterizes the interest in human behavior that arose in the fields of natural and ecological science, which is the other main external domain contributing to the establishment of environmental psychology. Here, the focus was placed on the possible (and usually negative) impact that human behavior can have upon the natural environment, at both a local and a global or biospheric level. Therefore the importance of paying attention to the human dimension—considered at an individual, social and cultural level—of global environmental changes, such as the greenhouse effect and climate changes, the loss of biodiversity, the depletion of the ozone layer, and so forth, was recognized. In parallel with the increasing relevance assumed by these environmental problems, ecologists also became increasingly aware of the need to integrate the natural and social sciences when dealing with environmental phenomena. The growing interests of current environmental psychology in topics such as environmental concern, pro-environmental values and attitudes, ecologically relevant behaviors, sustainable lifestyles, and natural resource management is also a consequence of this increased awareness.

## 2. THE MAIN AREAS OF INTEREST OF CURRENT ENVIRONMENTAL PSYCHOLOGY

Contributions in environmental psychology can be grouped into five general domains: spatial behavior,

environmental cognition, environmental stress and restoration, environmental assessment, and environmental concern and resource management. In this section of the article, the main empirical results and outcomes of environmental psychology are briefly introduced for each of these five domains. The various single articles composing this section provide a more exhaustive treatment of the different topics related to these five domains.

The environmental psychological processes grouped into these five domains can be considered transverse to the variety of specific settings in everyday life. However, since environmental psychology has also been increasingly characterized by the adoption of a place-specific approach to people–environment interactions, many environmental psychologists have been increasingly interested in a number of specific places that have a particular impact on everyday life and well-being, such as offices, schools, houses, stores, hospitals, museums, prisons, cities, etc. Three articles in this section focus on three particularly important places in our everyday experience: the workplace, the school, and the city.

### 2.1. Spatial Behavior

The concept of spatial behavior relates to how individuals regulate and use (in terms of appropriation and defense) their spatial environments at different personal, interpersonal, and group levels. Environmental psychology interested in spatial behavior focuses on the role of the spatial properties of the environment in shaping and regulating social interaction in everyday situations. Four major environmental psychological concepts are considered in this field: territoriality, personal space, privacy, and crowding. The articles by R. Sommer and by C. Werner, B. Brown, and I. Altman review the relevant empirical findings of these domains of environmental psychology, as well as their implications for daily life.

These domains encompass the variety of strategies by which people and communities set up and regulate the spatial boundaries of their living environment (territoriality) and the very personal sphere of the amount of space we put between us and the others (personal space). The concept of privacy encompasses both of these aspects, as it refers more generally to the various personal and group-based processes through which people set and control their mutual closeness in daily interactions.

The article by R. Sommer thoroughly illustrates the concept of human territoriality. As Sommer outlines, territoriality is the spatial appropriation, marking, and defense of our living spaces. Following an evolutionary orientation, territories can be defined as those spatial

areas that deserve to be defended from external intruders. An intriguing issue connected with human territoriality is that of crime prevention, through the concept of defensible space. As the article describes, there is a widely documented negative correlation between the presence of physical and symbolic territorial markers (e.g., well-maintained dwellings, tended yards, signs of occupancy) and the likelihood of criminal intrusion and vandalism in residential areas. The relationship between the presence of markers and crime, however, may be mediated by social variables. Notably, clear territorial markers are related to stronger residential identification, place attachment, sense of community, and higher feelings of safety among residents. Thus, architectural defensive features might play an indirect role in reducing the likelihood of crime. They reflect and communicate strengthened community-based ties and greater social cohesion; these make potential intruders less willing to perpetrate criminal activities. Another important outcome of the concept of territoriality is related to urban planning. Sommer's article in fact points out how specific design features encouraging residents' identification with and control over a specific territory may promote its proper maintenance. Furthermore, common areas such as urban parks or playgrounds should not be placed between the territories of rival gangs, as they would then run the risk of being abandoned and vandalized. Other domains in which the study of human territoriality has offered useful insights are sport (as sport teams perform better on their home field), police investigations (as many criminal gangs mark their territories by the use of graffiti and other symbols), and environmental conflict resolution (as the creation of ad hoc territories reserved to different stakeholders may prevent the emergence of land use disputes).

A second article by R. Sommer illustrates the environmental psychological concept of personal space. As opposed to territoriality, which is a place-based or site-based concept, personal space is a person-related, and thus trans-place, concept. In his seminal studies in this field, Sommer defined personal space as that emotionally tinged area that people desire to maintain around themselves and that they feel is "their space." Sommer's article also provides a review of the various individual, situational, group, and cultural factors regulating personal space. As this article reveals, personal space has relevant implications for several domains of human-environment interaction. For example, the personal space concept has provided insights for the design of institutional settings with fixed seating, as well as in the domain of mass transportation. Likewise, issues

regarding personal space were incorporated in the design of public or semi-public settings (e.g., offices, stores, banks). These should be set up in order to provide spaces that leave the users as free as possible to shift among different degrees of desired mutual closeness. Many guidebooks describing the appropriate spacing in different social encounters are also available. For example, salesmen are trained about how their selling activities might profit from an appropriate spatial interaction with customers. Finally, the issue of personal space has serious implications for the legal field: for example, the invasion of personal space has been an issue in court cases concerning prison crowding and sexual harassment. It is also an aspect considered during jury selection in high-profile trials, where professional consultants advise lawyers in order to detect potentially biased jurors.

Personal space and territoriality are both involved in the more general concept of privacy regulation. In this sense, they can be seen as two different strategies that people use to regulate their privacy, and through which people strive for psychologically satisfactory level of openness or closedness to others. A seminal contribution to the definition of privacy in environmental psychology comes from I. Altman. Within this section, the reader will find a detailed presentation about the psychological mechanisms of privacy regulation in the article written by C. Werner, B. Brown, and I. Altman. As they report, privacy regulation is defined as the process of selectively controlling access to one's self or group. In their article, Werner et al. underline how environmental psychology treats privacy as a dialectic process: that is, people might avoid or seek social contact, depending on the specific situation they are in. An important point is that individuals feel more comfortable when allowed to be as open or as closed to the others as they desire. This is also related to another important issue in the environmental psychology of spatial behavior, that is, crowding. As reported by Werner et al. in their article, crowding can be defined as one's perception that there are too many people present in a given situation. The possibility of a proper privacy regulation therefore becomes very important when people experience crowding. Environmental psychology has also treated crowding as a specific environmental stressor.

The article of Werner, Brown, and Altman also refers to how environmental psychology has identified a number of physical features that either allow people to experience a satisfactory regulation of privacy or do not. As a consequence, the privacy concept has many practical implications for the design of various specific environments such as workplaces, schools, prisons,

hospitals, and public residences. For example, university dormitories designed as suites or apartments can promote the formation of small subgroups of occupants; this is meant to buffer the possible negative consequences of crowding upon the possibility of privacy regulation. When designing the layout of a home, architects should place rooms that are usually considered more private (e.g., the bedrooms, or the family bathrooms) far from the entrance. Likewise, privacy-related concepts have been applied to urban planning and community development. The article illustrates how, for example, public spaces provided with design features encouraging interpersonal contacts (e.g., streets that invite walking, tree-shaded streets, low speed limits, green areas) might be used to promote strengthened community bonds and social ties among neighbors.

Beyond the effect that specific design features can have in affording or impeding people's mutual interactions, a central factor in people's satisfaction or dissatisfaction with privacy conditions is the possibility of regulation and control. A general recommendation is that the physical environment should be designed in order to provide to its users the possibility of regulating privacy, by seeking or avoiding social contact according to their situational and personal needs and desires.

## 2.2. Environmental Cognition and Cognitive Mapping

Since its beginning, environmental psychology has focused on the relationship between cognitive processes and physical space. The article of R. Golledge on spatial cognition thoroughly reviews this area of study. As Golledge notes, spatial cognition is a multidisciplinary area involving psychological science (e.g., cognitive and developmental psychology, neuropsychology) as well as other disciplinary and technical fields (e.g., geography, anthropology, architecture and urban planning, computer science).

A core concept in the environmental psychology of spatial cognition has been cognitive mapping. The term cognitive mapping refers to the process of mentally acquiring, representing, storing, and using information about the spatial properties of the physical surroundings and the relations among its constitutive elements. The interest in cognitive maps in psychology can be traced back to the works of E. Tolman in 1948. Studies on cognitive mapping have focused on the different kinds of information people rely on when forming a map of a given spatial setting.

Studies on cognitive mapping have dealt with the representation of large-scale urban environments, as well as with the representation of the interior of buildings. One goal of this kind of research was to investigate how to aid people's orientation, memory of places, and wayfinding in complex environments.

Kevin Lynch's seminal book *The Image of the City* influenced several ensuing works on the formation of mental maps of urban settings. Urban planners can use principles highlighted by these studies in order to design urban spaces that are capable of facilitating people's orientation in the city. For example, some spatial features of the residential environment, such as spatial prominence, sharp contours of buildings, easy accessibility, and the presence of natural surroundings might facilitate the overall comprehension and use of neighborhoods. Likewise, principles drawn from the cognitive mapping literature have been used for aiding wayfinding in complex interior structures, such as university buildings, hospitals, or museums. The presence of proper signs (floor plans, large graphic aids) and the availability of visual accesses may be effective in enhancing users' proper orientation.

As Golledge points out in his article, there is now a particular interest in the study of spatial cognition in aged people. The increased trend of aging in western societies has in fact led to an increased concern for the quality of life of older people. Research in this field is therefore being devoted to trying to identify how the layout of the physical space can be designed and managed in order to buffer the declines in spatial competence associated with aging, and the negative consequences of this decline for the safety, effectiveness, and autonomy of the elderly.

## 2.3. Environmental Stress and Restorative Environments

The environment in which most of humankind currently lives is often characterized by the presence of several potentially adverse physical conditions that can be a chronic, powerful, and uncontrollable source of psychological distress. The article by G. Evans and S. Cohen on environmental stress, included in this section, provides an exhaustive review of this traditional area of study of environmental psychology.

Environmental stressors can be defined as those actual or perceived adverse properties of the physical environment that are capable of producing a negative (physiologically and psychologically costly) effect

upon a person. It is important to note that one's everyday environment is often a source of simultaneous and interdependent multiple stressors. The article by Evans and Cohen addresses this particular aspect of environmental stress. The effects of different environmental stressors frequently add or interact with each other. As Evans and Cohen outline, individuals must make a remarkable adaptive effort in order to cope with such adverse conditions. Because humans have great powers of adaptation in adopting strategies to cope with stressors, the coping activity may become a stressful condition itself, particularly in the long-term. The article focuses on the four main costs of adapting to poor environmental conditions: cumulative fatigue, learned helplessness, physiological mobilization, and overgeneralization. These adaptive costs are caused by chronic exposure to a number of different environmental stressors, such as crowding, noise, or pollution.

Some of these main environmental stressors can be a very common experience for a great number of people, in particular for residents of cities. Conversely, there are other critical environmental conditions that are very uncommon that are also of interest to environmental psychologists. The article by R. Bechtel on extreme environments and mental function elaborates on this issue. As Bechtel points out, there are two principal classifications of extreme environments. On the one hand, there are permanent extreme environments, such as high mountains, deserts, cold regions, and jungles, and on the other hand, there are temporary extreme environments, such as Antarctic polar stations, space shuttles, and simulation laboratories. The main goal of environmental psychology concerned with extreme environments is to better understand how human mental function and behavior can be affected by these kinds of environmental conditions in order to buffer their potentially negative impact on mental function: that is, how to make these environments less extreme for their users.

In recent decades, environmental psychology has also addressed stress-related issues from an opposite viewpoint: that is, how the physical environment can restore human mental functions, and therefore promote psychological well-being. The article by T. Hartig on restorative environments provides an exhaustive review of this promising literature. As the article explains, restorative environments are those that not only permit, but promote, restoration, where restoration is defined as the process of recovering physical, psychological, and social resources being diminished by efforts to adapt to external demands.

In his article, Hartig reports how a considerable amount of empirical research has demonstrated the

existence of specific environmental conditions capable of positively affecting people's feelings of well-being, although different theoretical elaborations have been proposed to explain it. Perhaps the most well-documented effect within this area of research is that natural settings appear to be more restorative than built-up ones, because they stimulate people's interest and provide people with the possibility of being away from their usual experience. Furthermore, natural environments are highly restorative because of various specific features, such as visual depth and visual complexity.

Research on restorative environments can have a high relevance for the design of health care structures. A widely cited study by R. Ulrich, published in *Science* in 1984, showed for example that providing hospital patients with the possibility of contact with nature (even just visual contact) might positively impact the effects of medical therapies. Restorativeness is a relevant issue for the domain of urban planning as well. Setting up urban green areas that are easily accessible to urban dwellers can provide them with more frequent opportunities for psychological restoration in a highly stressful environment such as the city.

#### 2.4. Environmental Assessment

Although a theoretical and conceptual distinction between environmental assessment and environmental appraisal has been proposed, the former being more place-centered, the latter being more person-centered, here both of these aspects of assessment are considered. Environmental psychology interested in environmental assessment has encompassed a large amount of empirical contributions and theoretical models aimed at answering very important questions. For example, how do people evaluate, judge, or express preference for different kinds of environments? How do the physical features of a setting interact with people's personal and social characteristics when they evaluate an environment? In particular, which physical properties of the environment are related to positive evaluations and which are related to negative ones? Furthermore, what are the individual and group differences that may lead to differing environmental evaluations? And also, how do cognitive and affective judgments interact when a certain environmental scene is evaluated?

Generally speaking, different affective qualities have been found to characterize different environments. These qualities vary according to two main bipolar axes, pleasure (pleasant/unpleasant) and arousal (high/low). The physical surroundings of people's everyday

lives can generate positive or negative emotions in terms of the pleasure they afford to the perceiver, or in terms of the amount of arousal they provide to the perceiver.

According to a more cognitive model proposed by R. and S. Kaplan, there are four main physical characteristics predicting the pleasantness or unpleasantness of an environmental setting: coherence, complexity, legibility, and mystery. Coherence and complexity are two features of environmental scenes that immediately strike the perceiver. Conversely, legibility and mystery are two features that can be inferred after viewing an environmental scene for more time. The Kaplans' model also offered interesting insights into the study of landscape preference.

Whether a setting will generate positive or negative emotions in its users/perceivers is also a function of the prior state of the perceiver him- or herself. People's affective appraisal of a given environment is an adaptive process, in the sense that it is a function of both the properties of the environment and the characteristics of the perceiver. For example, high sensation seekers prefer highly arousing environmental scenes, and vice versa.

Several studies have also showed that natural environments are usually preferred to built environments by a large number of subjects. This preference seems to be consistent across different cultures, genders, ages, and so forth. Regardless if they are evaluated in terms of general preference, aesthetic beauty, or perceived restorativeness, natural landscapes are judged more positively than built settings. The presence of water in an environment is also positively judged. Moreover, as previously mentioned, exposure to natural views has been proved to have positive effects upon physiological and cognitive indicators, such as the rapidity of healing after hospitalization and the relief of mental fatigue. Such a nature favoritism, or biophilia, has been explained in evolutionary terms: people are still attracted by nature because it provided sources of sustenance and shelter in the earlier stages of human evolution. The psychological benefits of contact with nature can be due to the fact that evolution favored those who could benefit more from exposure to nature.

Despite this apparently generalized nature favoritism, a limited degree of human intervention within a natural environment is associated with positive evaluations. That is, scenes in which the human presence is visible (although not too intrusive) are usually preferred to scenes of total wilderness. This leads to the importance of another factor in shaping environmental evaluations, perceived control. People tend to judge more positively those environments that they perceive as controllable, compared to those that they perceive as uncontrollable.

Typically, the perception of control is associated with higher perceived safety. Speaking more generally, environmental evaluation could be conceived of as a function of the perceived cognitive, affective, and behavioral fit between the characteristics of the person (e.g., needs, goals, values, expectations) and the actual physical properties of the environment. In other words, people will prefer those environments they see as matching their needs, goals, and expectancies in that specific context at that given time.

The processes driving environmental assessments are relevant for the study of residential preference and satisfaction. The article written by M.V. Giuliani on residential preference and attachment across the lifespan, and the article by M. Bonaiuto on residential satisfaction and perceived residential environment quality, both included in this section, specifically deal with these issues. As both Giuliani and Bonaiuto describe in their articles, residential preference and satisfaction are multidimensional processes, given the tendency of people to combine different aspects when judging the quality of a residential setting. At a macro-level (e.g., the neighborhood) these aspects are related to (1) the physical and aesthetic characteristics of residential areas, (2) the social relations that people can establish in a specific residential area, and (3) the actions and behaviors that people can perform in a specific residential area. At a micro-level (e.g., the home) these aspects are related to (1) the quality of the interior arrangement and decoration of the house, (2) the specific architectural style of the house, and (3) the location of the house, in terms of both proximity to functional services and distance from the city center.

An important outcome of studies on environmental assessment has been the development of psychometric tools for the measurement of environmental preference and satisfaction, such as perceived environmental quality indicators (PEQIs). Because environmental quality assessment is conceived as a multi-dimensional process, PEQIs should encompass a sufficient variety of dimensions: the most widely shared qualities are aesthetics, functionality, safety, social relations, noise and pollution, and green spaces.

The information gathered through these tools can often represent the first step in developing more inclusive and participatory methods in environmental design and planning. Why is including users' perspectives or letting people take part in decisions considered so important in the environmental domain, as well as in many other domains? The article by A. Churchman and E. Sadan on participation in environmental design and planning thoroughly illustrates this issue. As Churchman and

Sadan outline, the major argument in favor of public participation is that it enables decisions that better fit the needs of the people. The positive consequences of this better fit are twofold, as both public support for environmental policies and people's care for a proper maintenance of their environments will be more likely. Nonetheless, Churchman and Sadan also identify and discuss a number of factors that may bedevil the effective implementation of participatory processes in the environmental domain. For example, public leaders and politicians sometimes see participation as a threat to their power. It is important to outline here how this article presents various arguments that support the crucial role of environmental psychology in promoting and improving participatory processes in environmental design, planning, and management. This aspect is particularly relevant, as it relates to the issue of improving the processes of inclusive environmental governance, which is currently seen as a crucial goal by many international and intergovernmental authorities.

### **2.5. Environmental Concern, Environmentally Friendly Behaviors, and Natural Resources**

A shared belief among the scientific community and among public opinion is that the quality of our environment at the local and global level has rapidly decreased almost everywhere in the last decades. As a matter of fact, the major reason for that progressive and often dramatic deterioration is the impact of many human activities and rapid industrial growth on the ecosystems. Therefore, the healthy or unhealthy state of the environment primarily depends, and will depend in the future, on those human activities that can cause relevant and often irreversible environmental local and global changes.

In order to deal with these kind of problems, environmental psychology has increasingly addressed the issue of environmental concern in the last two decades. Two articles included in this section illustrate this important area of interest in present-day environmental psychology.

In his analysis of conservation behavior, F. Kaiser focuses on the psychological factors at the basis of the actions contributing to environmental preservation. As Kaiser suggests, people's conservation behaviors can be approached from two rather different perspectives. Conservation behavior, in fact, can be defined either from an observer's or from an actor's viewpoint. The first implies a focus on the consequences for the environment and considers apparently similar actions (e.g.,

recycling paper and recycling used batteries) as distinct behaviors. The second perspective focuses more directly on the human motivations driving conservation behaviors and groups even apparently different actions (e.g., recycling glass and owning solar panels) into the same behavioral category. Kaiser's article also illustrates how this second approach is more directly linked to a psychological perspective. On the contrary, a too strict focus on environmental consequences may lead to underestimating the psychological determinants of conservation actions.

But what are the psychological determinants of environmentally friendly behaviors? The article by H. Staats on pro-environmental attitudes and behavioral change deals more directly with this question. In particular, this article focuses on the relationship between environmentally friendly attitudes and behaviors: that is, the congruence or incongruence between what people believe and feel and how they behave toward the environment.

Pro-environmental attitudes are increasingly shared by many people in western societies. Some social structural variables (e.g., age, gender, level of education, place of residence, political orientation) appear to be related with this concern. Typically, young, female, highly educated, urban, and liberally oriented subjects express a stronger concern for the environment. The increased concern for environmental issues has been summed up by the concept, introduced by R. Dunlap and K. Van Liere at the end of the 1970s, of a new environmental paradigm (NEP). The core idea of NEP is that an increased number of people are developing a new perspective toward the environment, based on the belief that the state of the earth's ecosystems is becoming more and more precarious, therefore compromising human survival.

However, people often do not coherently translate such a positive concern into consequent pro-environmental behaviors. The reason for this lack of correspondence has been explained in different ways. Staats' article guides the reader through two important social psychological models of attitude-behavior relations: the theory of reasoned action (TRA) and the theory of planned behavior (TPB). As Staats points out, the basic tenet of these models is that the relation between attitudes and behaviors is mediated or moderated by other variables, such as behavioral intentions and perceived behavioral control. TRA and TPB have been applied, sometimes successfully and other times less successfully, to several environmentally relevant behavioral domains, such as waste recycling, travel modes, water use, energy use, and green consumerism.

At any rate, the role of variables other than attitudes seems to be crucial for explaining pro-environmental actions: among them there are past behaviors and habits, environmental knowledge, social and personal norms, and value orientations. Well-established habits are often difficult to quit, even when their performers are aware of the possible negative consequences for the environment. Likewise, a lack of specific knowledge about environmental issues, or about the environmental consequences of a specific behavior, may function as a barrier to pro-environmental actions: people are in fact frequently unaware of, or uncertain about, the actual state of the environment and the negative consequences of their behaviors on it. For example, a deeper knowledge of environmental issues seems to distinguish committed environmental activists from the general public, although it is difficult to state whether knowledge comes before activism, or vice versa. Both social and personal norms are also related to pro-environmental behaviors. Although some explicit pro-environmental norms are socially shared, people might not always behave according to these social norms, especially because several, sometimes conflicting, norms can be present at the same time in a given situation. In particular, the divergence between prescriptive (what is explicitly prescribed) and descriptive (what is observed in others' behaviors) social norms can hamper pro-environmental actions. A series of interesting experiments conducted by R. Cialdini and his colleagues showed that the context in which prescriptive and descriptive norms are framed and made salient can account for a particular environmental behavior such as littering. The role of values and ethical principles in shaping pro-environmental behaviors has been explored as well. A distinction has been proposed between two major value orientations about environmental issues: ecocentrism and anthropocentrism. Ecocentric people value environmental preservation for its own sake; conversely, anthropocentric people value environmental preservation because of the positive consequence that it can have upon human well-being. The former are more likely to behave coherently with their pro-environmental attitudes compared to the latter, even when this implies some personal or economic cost.

But what are better methods for the promotion of environmentally responsible behaviors and for the prevention of environmentally adverse behaviors? In his article, Staats outlines various possible strategies that seem to be effective to this end; they can be information-based (e.g., communication campaigns, education,

advertising), incentive-based (e.g., monetary rewards/punishments), prescription-based (e.g., laws, rules, regulations), or community-based (e.g., public involvement and participatory programs). Different kinds of behaviors in different contexts deserve different kinds of strategies. As P. Stern has recently pointed out, there are some basic principles to be followed for enhancing the likelihood of changing environmentally relevant behaviors: the use of mixed intervention strategies, the adoption of an actor's perspective, the constant monitoring of programs, and the promotion of public participation.

The models trying to explain the lack of correspondence between pro-environmental attitudes and behaviors that were briefly reviewed above rely on the role of various factors that could interfere with the supposed positive association between attitudes and behaviors. This theoretical position, however, may not adequately account for other social and societal processes involved in people-environment relations, in particular, the mutual interdependency of people's actions in the environment. A theoretical approach that delves more directly into this direction is the social dilemma paradigm.

Social dilemmas (SDs) are situations in which two or more people are faced with a choice between pursuing individual gains or maximizing collective benefits: if everyone (or too many) chooses the former option, everyone is worse off than they would be if everyone (or sufficiently many) chose the latter. In other words, in a SD, the interest of a single person is in conflict with a more general collective interest, and one is faced with a choice between a defective (or selfish) option and a cooperative (or altruistic) alternative. Such a conflicting situation often exists in the environmental field when individual and collective interests clash over the exploitation of a limited natural resource. These dilemmas are usually referred to as commons dilemmas.

As Staats suggest in his article, almost every environmental problem can be framed and approached as a commons dilemma situation. Most of the times the individual-collective conflict in the exploitation of a limited resource has a temporal dimension as well. That is, maximizing individual gains in the short term will result in a collective loss in the long run (the common resource will be extinguished); conversely, limiting the individual gains in the short term will result in a collective benefit in the long run (the common resource will be guaranteed). This process is at the core of the political concept of sustainable development, and has been well illustrated by G. Hardin in his famous 1968 article entitled "The Tragedy of the Commons."

On the basis of the SD paradigm, it is not surprising that people frequently find out that behaving in an environmentally negative way is in the short term more advantageous, easier, more comfortable, less costly, and so forth. To an extreme extent, in many daily situations, environmentally unfriendly behaviors are the rule and not the exception.

In addition to the Staats' article, two other contributions included in this section deal, more or less directly, with the SD paradigm. In particular, in his article on environmental versus individual risk taking, C. Vlek reviews the literature on individual risk taking in order to understand how to manage environmental risks as a collective risk. As Vlek illustrates, various traditional psychological theories on risk can be considered when dealing with environmental risk perception, such as classical decision theory, prospect theory, risk homeostasis, protection motivation theory, and emotional approaches to risk. Furthermore, it is shown that many collective risks for the environment are often the result of the summed external negative effects of numerous individually advantageous activities. To this extent, the SD paradigm is directly called into question by Vlek's article, as it shows how the minimization of environmental collective risks and the maximization of individual benefits may be incompatible goals in many everyday situations.

Likewise, the article on travel and the environment by T. Garling illustrates how concepts deriving from the SD paradigm can be relevant to the domain of motorized transportation. As Garling shows, the increasing trend of private car use in urban areas is starting to pose serious threats to the global environment and to the health and well-being of urban dwellers. These threats are overwhelming the individual benefits of the single car users. The choice of whether to shift from private to public motorized transportation systems or from motorized to un-motorized vehicles can be framed as a social dilemma for many car and motorbike drivers. Therefore, a number of policy measures aiming at reducing private car use among urban dwellers that are outlined by the Garling's article can be seen as attempts to solve that dilemma.

But which are the main factors orienting our choices in social dilemma situations? Three crucial components of SDs are information, communication, and social identification. The possibility of monitoring others' behaviors and the possibility of communication may promote cooperation. Likewise, high social identification may result in stronger cooperation, although the opposite may be true. In fact, when a dilemma involves two or more groups competing for the same resource,

identification with a specific group may result in diminished cooperation.

These aspects may also have relevant implications for legitimizing and facilitating the acceptance of authority structures regulating citizens' environmental behaviors in our societies. The way a situation is structured by the authorities and interpreted by the individuals acting in it is crucial for approaching environmental dilemmas. An example can be found in the domain of biodiversity conservation. Here, highly centralized environmental policies (e.g., land use regulations, designations of National Parks) can represent a dilemma in the perception of local stakeholders. A strong sense of local identification might then form a basis for in-group biased environmental perceptions, in which public authorities are seen as a conflicting out-group by local identifiers. This can lead to the emergence of local protests against specific environmental regulations.

In sum, the SD paradigm is increasingly being used to address many current environmental problems regarding the management of limited natural resources. The results from studies based on this paradigm suggest that strategies for solving collective environmental dilemmas in our societies should stress the common fate between individual group members, in particular because they show the crucial importance of communication and shared identities among individuals for the promotion of cooperative and pro-environmental choices.

## 2.6. Place-Specific Environmental Psychology

The articles in this section described above refer to environmental psychological processes that can be considered as transverse to the places where people's everyday experiences occur. However, current environmental psychology has been increasingly characterized by the adoption of a place-specific approach to people-environment interactions. This has led many environmental psychologists to concentrate their interest on specific places, ranging from the micro-scale of home interiors, to the meso-scale of workplaces, schools, and hospitals, to the macro-scale of neighborhoods and cities. There are three articles in this section that focus more directly on three specific places considered as particularly relevant to many people's daily lives: the workplace, schools, and the urban environment.

The article written by R. Walden on working environments reviews some of the main theoretical and empirical contributions addressing environment-behavior

relations in the workplace. In her contribution, Walden notes how the study of the physical characteristics of workplaces can be traced back to the beginnings of psychological inquiry. The studies by E. Mayo and his coworkers at the Western Electric Company on the effects of the physical environment upon work performance date back to the first half of the 20th century. These pioneering works clearly highlighted how the physical features of the work settings can interact with the interpersonal relations and organizational aspects at work in shaping work performance. Present-day models of environment–behavior relations in the workplace, such as one recently proposed by R. Gifford, still reflect the influence of Mayo’s earlier findings. The model by Gifford mentioned in Walden’s article places a particular emphasis on the different degrees of congruence or incongruence resulting from the interaction between the personal characteristics of workers (e.g., experience, job level, personality, skills, motivation) and the physical characteristics of the work setting (e.g., noise, temperature, light, spatial density, quality of materials). The higher the congruence, the better the outcomes of the working activities. These can be assessed in terms of workers’ stress or well-being, individual and collective performance, job satisfaction, or organizational commitment. In addition, various psychological processes (e.g., arousal, adaptation, overload, emotions, personal control) may act as mediators of the relationship between the environment–person fit and the outcomes of the working activities. As Walden notes in her article, future trends in the design of workplaces point to rapid development of communication technologies as well as to the role of these in matching the augmented concern for sustainability that characterizes human societies. One example is the possibility for off-site working, or telecommuting, through the use of remote cable connections or wireless systems. This possibility could partly offset the disadvantages of traffic congestions among commuters.

A second article by R. Walden, on school environments, focuses on another place that has also been widely investigated by environmental psychology. As with workplaces, the attention devoted to schools is a long-standing tradition in environmental psychology. In fact, the interest of environmental psychology for schools can be traced back to the pioneering studies of Barker and colleagues in the 1960s. Generally speaking, environmental psychology has clearly outlined how the physical features of schools have a considerable impact on the effectiveness of the educational

program. Different physical characteristics of schools can either promote or hamper the education process. As Walden reports, there is consistent evidence showing that small schools do a better job than large schools. Empirical research has showed how smaller schools are characterized by a better relationship between students and teachers, diminished rates of vandalism and violence, faster career advancement of children with disadvantaged backgrounds, increased participation in and commitment to school activities, and greater satisfaction of parents. Therefore, Walden states that schools with no more than 500–700 pupils appear to be a reasonable solution, while schools with more than 2100 students have shown the worst performances. Apart from school size, several other characteristics of the educational physical setting may affect the quality of the education process. The most important are the state of maintenance of the school buildings, the presence of design features affording proper privacy regulation, good acoustics in classrooms, the availability of daylight in classrooms, the presence of playgrounds and schoolyards furnished with nature and natural materials, and the presence of carpets and decorations. Finally, the design of new schools can benefit from the involvement of students, teachers, parents, and members of the local communities.

A third article focuses on another important location that encompasses the two previously mentioned places: the city. As G. Moser outlines in his contribution on urban environment and human behavior, about half of the world’s population spends the majority of their lifetime in big cities.

As reported by Moser, cities are more and more becoming places in which daily conditions are far from optimal for their inhabitants: cities are by and large considered as bad places to live. In line with this trend, environmental psychology has introduced the concept of urban stress to indicate the sources of stress that are typical of urban life, and therefore that affect a large percentage of the population. Living in a large-scale urban settlement can frequently become a source of multiple stressors, because of the simultaneous presence of several adverse environmental conditions. This aspect has also been discussed in the Evans and Cohen article on stress. For example, urban settings can amplify the relationships among crowding, social withdrawal, social support, and pro-social behaviors. Moreover, urban life presents several conditions in which people perceive a very low degree of control over events: this often results in diminished self-efficacy and helplessness.

On the other hand, despite their bad reputation, cities continue to attract a large part of the human population. Therefore, as the Moser's article reminds, it is not surprising that urban settlements have been a specific object of analyses for many scholars in environmental psychology. Some important psychological processes related to the everyday urban experience were actually pointed out by environmental psychology, such as place- and urban-related identity and urban attachment. In that sense, cities are a primary arena in which almost any environmental psychological theory and empirical finding can be considered. As a consequence, cities can be thought of as a wide field-laboratory for testing environmental psychological theories. But what are the main trends of environmental psychological research on urban settlements? The article by Moser describes the increased concern for issues of sustainability and sustainable development that have emerged in the scientific community in the last decade. It is worth stressing how cities, considered as ecosystems, were seen as having a crucial role by the many intergovernmental agencies concerned with the promotion of sustainability. As a consequence, the concept of sustainable cities has been proposed in order to define the major environmental transformations needed to improve urban systems in order to enhance their ecological "performance." These issues are the next major challenge for environmental psychology focused on urban issues. Embracing this perspective, Moser also shows how urban-related identity is an important environmental psychological factor that may account for the adoption of sustainable lifestyles among city dwellers. As an example, Moser mentions the city-identity-sustainability (CIS) model that has been recently proposed by E. Pol and others. According to the CIS model, a strong sense of identification with the city can lead urban inhabitants' to a stronger and more durable engagement in pro-environmental activities.

### 3. THE FUTURE OF ENVIRONMENTAL PSYCHOLOGY

Since its beginnings, environmental psychology has been a frontier field within psychological science. On the one hand, following the ecological demand coming from several fields of psychology, environmental psychology has tried to underlie the importance of a contextual/transcendental approach to psychological inquiry. Within this

approach, environmental psychology has focused on an aspect that was typically underestimated by mainstream psychology: the role of physical features of everyday life environments in affecting human psychological processes, behavior, and well-being. On the other hand, environmental psychology has tried to respond to specific and urgent demands coming from various technical fields devoted to the design, change, preservation, and management of these everyday environments.

In many aspects, environmental psychology is still a frontier field of psychological inquiry, as its primary aims have not substantially changed over the decades. The whole story about the context-related character of human behavior and mental function is still far from being known due to the wide variability and diversity of the physical contexts in which human behavior takes place. At the same time, many of the urgent environmental problems that generated an external demand for the development of environmental psychology have not been solved yet. Some of them have become even more serious, as the quality of the environment and the availability of life-supporting resources have diminished consistently in many parts of the world.

Thus, the research agenda of environmental psychology has broadened considerably in the last 20 years. For example, the collaboration between psychologists and architects has increasingly included people-natural environment relations together with people-built environment relations. The domain of political ecology with its concept of sustainable development has influenced the fields of architecture, engineering, and urban planning. The increased relevance of various new technological and architectural domains such as bio-architecture and landscape architecture is an example of such an influence. Also, an increasing focus has been placed on the human dimension of global environmental and biospheric changes.

To address these issues, many different psychological theories can be used, coming from a broad range of psychological research domains, such as perception psychology, social and developmental psychology, cognitive psychology, and behavioral decision theory. The task of environmental psychology is still to combine these existing theories as well as to develop new ones for better understanding specific people-environment relations, and for supporting environmental designers and managers in enhancing the well-being of present and future generations.

Overall, present-day environmental psychology can be characterized not only as a highly socially relevant field of applied psychology, but also as an opportunity

for developing psychology in a more context-related direction, that is, toward that real world psychology continuously envisaged in environmental psychology from its beginnings to the present.

### See Also the Following Articles

Environmental Stress ■ Intergroup Relations and Culture  
 ■ Personal Space ■ Stress ■ Territoriality

### Further Reading

- Altman, I. (1988). Process, transactional/contextual and outcome research: An alternative to the traditional distinction between basic and applied research. *Social Behaviour*, 3, 259–280.
- Barker, G. A. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.
- Bazerman, M. H., Messick, D. M., Tenbrunsel, A. E., & Wade-Benzoni, K. A. (Eds.) (1997). *Environment, ethics and behavior*. San Francisco: The New Lexington Press.
- Bechtel, R. (1997). *Environment and behavior: An introduction*. London, UK: Sage.
- Bechtel, R., & Churchman, A. (Eds.) (2002). *Handbook of environmental psychology*. New York: Wiley.
- Bonnes, M., Lee, T., & Bonaiuto, M. (Eds.) (2003). *Psychological theories for environmental issues*. Aldershot, UK: Ashgate.
- Bonnes, M., & Secchiaroli, G. (1995). *Environmental psychology. A psycho-social introduction*. London, UK: Sage.
- Canter, D., & Lee, T. (Eds.) (1974). *Psychology and the built environment*. London, UK: Architectural Press.
- Gifford, R. (2002). *Environmental psychology: Principles and practice* (3rd ed.). Collville, Canada: Optimal Books.
- Kaplan, R., Kaplan, S., & Ryan, R. L. (1998). *With people in mind: Design and management of everyday nature*. Washington, D.C.: Island Press.
- Levy-Leboyer, C. (1980). *Psychologie et environnement*. Paris: Presses Universitaires de France.
- Stern, P., Young, O. R., & Druckman, D. (Eds.) (1992). *Global environmental change: Understanding the human dimensions*. Washington, D.C.: The National Academies Press.
- Stokols, D., & Altman, I. (Eds.) (1987). *Handbook of environmental psychology*. New York: Wiley.
- Stokols, D. (1978). Environmental psychology. *Annual Review of Psychology*, 29, 253–295.
- Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science*, 224, 420–421.
- Uzzell, D. (2003). Our uncommon future. In R. Garcia-Mira, J. M. Sabucedo Cameselle, & J. Romay Martinez (Eds.), *Culture, environmental action and sustainability* (pp. 21–39). Gottingen: Hogrefe & Huber.
- Van Vugt, M., Snyder, M., Tyler, T. R., & Biel, A. (2000). *Cooperation in modern societies*. London, UK: Routledge.
- Wapner, S., Demick, J., & Yamamoto, T., Minami, H. (2000). *Theoretical perspectives in environment-behavior research*. New York: Kluwer Academic.



# Environmental Stress

Gary W. Evans

Cornell University, Ithaca, New York, USA

Sheldon Cohen

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

1. Theoretical Framework
  2. Conclusion
- Further Reading

## GLOSSARY

**adaptation level shifts in environmental perception** History of experience with an environmental stressor changes referential criteria for environmental perception of current environmental conditions.

**behavioral aftereffects** Performance deficits following exposure to environmental stressors.

**cognitive consequences of learned helplessness** Lower mastery, reduced self-efficacy, or greater externality in locus of control from chronic environmental stressor exposure.

**coping perseverance** The use of well-learned coping strategies when an environmental condition is no longer present.

**cumulative fatigue** The buildup of fatigue from expenditure of energy to cope with an environmental stressor; the three types of cumulative fatigue are behavioral aftereffects, spillover, and diminished coping with multiple stressors.

**diminished coping with multiple stressors** The reduced ability to cope with a subsequent stressor following exposure to an environmental stressor.

**effort by environmental stressor interaction** Physiological mobilization produced by cognitive or physical effort to maintain task performance during stressor exposure.

**environmental stressors** Physical characteristics of the environment that produce stress (e.g., commuting, crowding, noise, pollution).

**learned helplessness** The realization, caused by uncontrollable environmental events, that one cannot control one's environment. This may result in diminished motivation to assert control even when feasible to do so, the inability to learn that subsequent challenges that are objectively controllable can be controlled, and feelings of anxiety and depression.

**overgeneralization** Strategies for coping with chronic environmental stressors may become overlearned so that coping perseveration occurs in the absence of the stressor; adaptation level shifts in environmental perception may also occur.

**physiological mobilization** Elevated physiological stress produced by chronic exposure to environmental stressors; also caused by acute environmental stressor exposure in combination with the effort to maintain optimum task performance.

**spillover** The negative affect, strained interpersonal relationships, or fatigue produced by exposure to an environmental stressor in one setting that carries over into another setting.

Apart from the direct physical insults produced by toxins, mental health, psychophysiology, performance, and human motivation are all affected by environmental quality. In this article, we present a theoretical framework for understanding how environmental quality impacts human behavior. Human beings are not passive recipients of environmental conditions; instead, they work to optimize the balance between environmental conditions and human needs. Human beings thrive in a vast array of ecological niches.

However, we pay a price for our enormous adaptational capacities. The adaptive costs of coping with suboptimal physical conditions provide a theoretical framework for describing how the physical environment influences human behavior. We develop this theoretical framework and illustrate psychological impacts of poor environmental quality.

### 1. THEORETICAL FRAMEWORK

There are four categories of the adaptive costs of coping with poor environmental quality that have psychological consequences. These categories are cumulative fatigue, learned helplessness, physiological mobilization, and overgeneralization.

#### 1.1. Cumulative Fatigue

It takes energy to cope with suboptimal environmental conditions. Expenditure of energy, particularly when demands are high or prolonged, causes fatigue. Three types of cumulative fatigue are behavioral aftereffects, spillover, and diminished coping with multiple stressors.

##### 1.1.1. Behavioral Aftereffects

Behavioral aftereffects refer to impaired task performance following exposure to an environmental demand. In a typical aftereffects paradigm, an individual works on a task during a physical stressor such as noise. Following cessation of stressor exposure, the individual is asked to perform another task. Compromised performance on the second task, when the stressor is no longer present, is an index of behavioral aftereffects.

In the Glass and Singer behavioral aftereffects paradigm, subjects work on a simple task while exposed to an environmental stressor. Environmental stressors that have been studied include noise, crowding, temperature, and air pollution. During the environmental exposure period, subjects worked on simple tasks (e.g., anagrams). Performance during exposure is generally unaffected. Immediately following task performance under the environmental stressor, the stressor is terminated, and the subject is given a behavioral aftereffects task. Two commonly used aftereffect measures are persistence on a difficult or unsolvable puzzle and ability to detect errors in a proofreading task. Table I depicts a typical set of findings from one of Glass and Singer's studies, in which immediately after uncontrollable noise exposure was terminated, subjects were

TABLE I  
Glass and Singer Aftereffects Data Following Uncontrollable Noise

	Noise	Quiet
Number of attempts on two unsolvable puzzles	10.88	41.20

administered a task that assessed their persistence on unsolvable puzzles.

In addition to the large number of replications of the Glass and Singer aftereffects data across a wide array of environmental stressors, another aspect of this paradigm is noteworthy. Several investigators have examined behavioral aftereffects of chronic environmental stressors such as commuting, crowding, and noise. Bullinger *et al.* noted that the longer the duration of noise exposure in the case of the opening of a new airport, the stronger were the impacts of chronic noise exposure on behavioral aftereffects (see Fig. 1).

##### 1.1.2. Spillover

A second type of cumulative fatigue produced by poor environmental conditions is spillover. Spillover occurs when conditions in one setting influence a subject's well being in another location. A common example of spillover is when poor working conditions interfere with

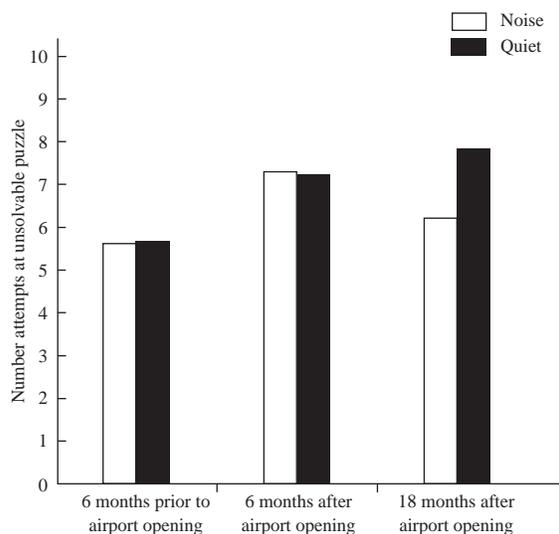


FIGURE 1 Task persistence among elementary aged school children in relation to airport noise exposure before and after the opening of a new airport. Controls matched on SES.

home life. A more congested commute creates interpersonal strains with family members at home. Epstein and Karlin reported that men but not women following a short term crowding experience in the laboratory felt less cohesive and more competitive. Immediately following exposure to crowded shopping conditions or high noise levels in a laboratory, individuals are less likely to behave altruistically when given a chance to provide help in a low stress setting. Table II provides illustrative data. A stranger (experimental confederate) appeared to be looking for a dropped contact lens on the floor of a shopping center in a low-density, inactive area of the center. Immediately prior to the assessment of spillover, subjects had been randomly assigned to perform a series of shopping tasks in noisy and crowded or relatively desolate areas of a mall.

**1.1.3. Diminished Coping with Multiple Stressors**

The effort applied in coping with environmental stressors can result in reduced capacity for coping with subsequent demands. Reduced coping capacity has been found following exposure to noise, crowding, high temperature, and pollutants. For example, the adverse effects of traffic noise on blood pressure in school children are accentuated if they also live in noisy homes. Residents of noisy areas respond more negatively to stressful events than residents with similar backgrounds (SES) who live in quiet areas. Crowding exacerbates the impact of high temperature on negative affect. Maxwell noted that the adverse influence of high-density daycare conditions on children’s levels of behavioral disturbances was accentuated by crowding levels in the child’s home. Researchers have also investigated crowding interactions with psychosocial rather than physical stressors. For example, elementary school aged children experienced greater physiological stress and psychological

**TABLE II**  
Altruistic Behavior Following a Simulated Shopping Task in a Crowded or Uncrowded Shopping Center

	<i>Crowded</i>	<i>Uncrowded</i>
Percent of subjects who helped	16.5	56.5
Seconds spent looking for contact lens	15	41

distress in relation to greater family turmoil if they lived in more crowded homes and the effects of daily hassles on psychological health in adults were accentuated in high density homes.

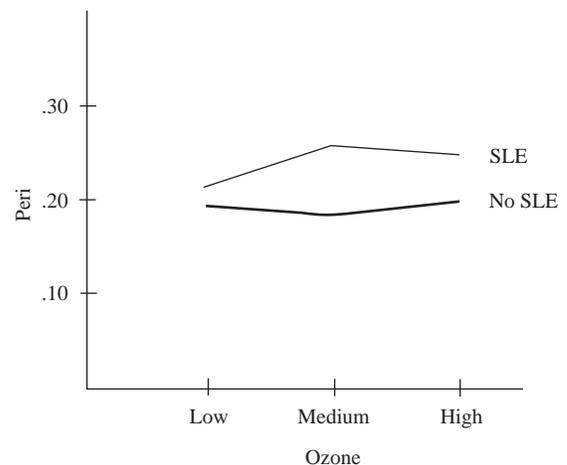
Air pollution also accentuates adverse responses to other stressors. The effects of polluted working conditions on respiratory and dermatological symptoms were greater for workers also contending with higher levels of job stress. Evans and colleagues found that ozone was linked to psychological distress, but only among individuals also experiencing stressful life events (see Fig. 2).

**1.2. Learned Helplessness**

Attempts to cope with uncontrollable environments can result in learned helplessness. Individuals who try to abate a negative environmental condition and are unable to do so eventually learn that their efforts to control their environment are fruitless. They become unmotivated to assert control (motivational consequences), even when it is feasible to do so, unable to learn that subsequent challenges they confront that are objectively controllable can be controlled (cognitive consequences), and become anxious and depressed (affective consequences).

**1.2.1. Cognitive Consequences**

Experiencing uncontrollable environments can lead to the inability to learn that controllable environments



**FIGURE 2** Psychological distress as a function of the interaction of exposure to pollution and recent stressful life events among a representative sample in Los Angeles.

can be managed or affected by personal effort. For example, individuals living under crowded conditions report less perceived control or feelings of mastery over their immediate surroundings compared to others living under less crowded conditions. Traffic congestion levels negatively impact feelings of control over commuting itself and in general among car drivers. Bus drivers facing more traffic congestion report significantly less control over their job.

Community airport noise exposure has also been linked to diminished feelings of personal control. For example, the typical maturational trajectory of increased internal locus of control was altered among elementary school children living for longer periods of time in airport noise impact zones (see Table III). SES controls were incorporated.

### 1.2.2. Motivational Consequences

As discussed previously, exposure to uncontrollable noise, crowding, or air pollution leads to deficits in behavioral aftereffects tasks. In the original aftereffects studies, Glass and Singer included an independent variable that points toward learned helplessness as a possible mechanism for behavioral aftereffects. When subjects were led to believe they could terminate the noise if it became too aversive by pushing a button, there were no behavioral aftereffects of noise. This occurred even though no one in this "perceived control" condition actually pushed the button. In short, aftereffects occurred only among those exposed to "uncontrollable" stressors. Table IV adds the noise with perceived control condition to our original behavioral aftereffects example shown before in Table I. Subjects were randomly assigned to conditions.

This salutary effect of perceived control over noise on behavioral aftereffects has been widely replicated. Parallel results have been uncovered in studies of crowding and pollution.

**TABLE III**  
Chronic Noise Exposure and Control Beliefs among Elementary School Children<sup>a</sup>

	Years in residence	
	<3	>3
Low noise	5.56	6.00
High noise	6.00	5.43

<sup>a</sup>Higher numbers reflect greater internality.

**TABLE IV**  
Aftereffects of Noise, Quiet, and Noise with Perceived Control

	Noise	Quiet	Noise with perceived control
Number of attempts on two unsolvable puzzles	10.88	41.20	41.67

More direct evidence that suboptimal environmental quality can produce helplessness comes from two paradigms. In one paradigm, participants performed a task in order to avoid a noxious stimulus. For some subjects, exposure to the noxious stimulus was contingent upon their performance—by performing well, they could minimize exposure to the noxious stimulus. Other subjects were "yoked" to the first group in that they were exposed to the same noxious stimulus for the same duration of time. For these subjects, however, exposure to the noxious stimulus was independent of their performance. These conditions (contingent vs noncontingent performance and punishment) were then followed by a second phase in which a performance contingent task was administered. For this task, avoidance of the noxious stimulus was possible for all subjects. Prior exposure to the inescapable stimulus condition induced learned helplessness in the second phase. Even when an instrumental response was available, subjects who had been in the initial inescapable condition took much longer to learn the escape response. Of particular interest is the fact that exposure to inescapable noise versus escapable noise reliably produces learned helplessness in adults.

In a second paradigm, vulnerability to the induction of helplessness is related to chronic environmental conditions. Cohen *et al.* gave elementary school children in either airport noise-impacted or quiet schools a jigsaw puzzle to work on. More children from high-noise areas failed to solve the puzzle. Of those children who failed, a larger percent of children from noise-exposed areas failed the puzzle because they simply gave up trying prior to the end of the testing period (see Table V). SES controls were included.

Crowded living conditions similarly affect learned helplessness in children. For example, the longer college students lived in a crowded dormitory, the greater their helplessness behaviors compared to students in

**TABLE V**  
**Percentage of Children Who Gave up on a Solvable Jigsaw Puzzle prior to the Total Time Available<sup>a</sup>**

	Noisy school	Quiet school
Percent giving up	31	7

<sup>a</sup>Total time available was 4 minutes.

uncrowded dormitories. Moreover, these helplessness behaviors closely mirrored the extent to which they felt powerless to regulate social interaction in their dormitory.

### 1.3. Physiological Mobilization

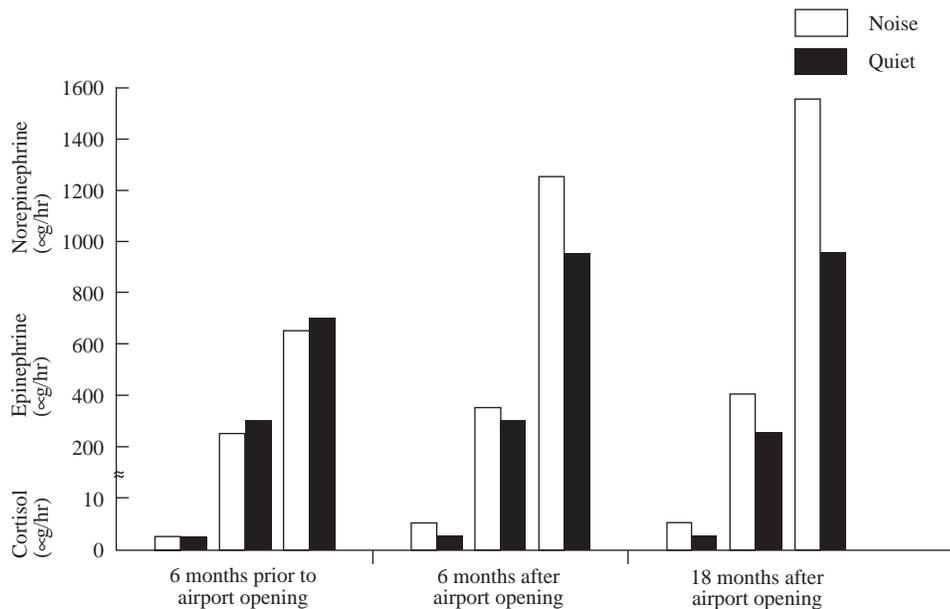
The marshalling of coping resources requires energy. This expenditure of energy requires mobilization of physiological resources, as reflected by elevated physiological stress. This occurs primarily in cardiovascular and neuroendocrine systems. Mobilization of energy in response to acute environmental demands is adaptive. The problem occurs with chronic physiological mobilization in order to meet repeated environmental demands. These elevations produce wear and tear on the body, alter the body's ability to turn off these systems when external demands cease, and can eventually lead to physical and psychological morbidity.

Effort and physiological mobilization are interrelated. When performing a task under a stressor, if effort is high, task performance is maintained but physiological mobilization escalates. If effort is low under a stressor, performance suffers, but physiological mobilization is minimal. Thus, there is an effort by stressor tradeoff.

#### 1.3.1. Physiological Stress

A large number of studies have been conducted on occupational noise exposure and blood pressure. The data are mixed and methodological flaws plague many of these studies. Community noise studies have used better designs and revealed consistent elevations in blood pressure in children exposed to airport noise. Fig. 3 shows data comparing school children's physiological stress levels before and after the opening of an airport. Children were tested 1 year before the airport opened, 1 year after, and then again 6 months later.

Crowding has similar effects on the physiological stress response. Crowding has been associated with elevated blood pressure in the laboratory and among institutional populations and in some but not all studies of crowded housing and neighborhoods. Traffic congestion has been associated with elevated blood pressure and stress hormones; more crowded or more difficult train commuting elevates stress hormones as well.



**FIGURE 3** Physiological stress responses in children in a prospective study of airport noise. The samples were well matches on SES.

### 1.3.2. Effort by Stressor Tradeoff

Some of the physiological stress accompanying environmental stressors is caused by increased expenditure of effort to maintain task performance. Cardiovascular and neuroendocrine responses to acute noise rapidly habituate, returning to baseline levels. This habituation does not occur, however, when subjects simultaneously work on cognitively demanding tasks during noisy conditions. More direct evidence for the role of effort emanates from studies showing a performance–physiological stress tradeoff. For example, Tafalla and Evans manipulated cognitive effort and noise in a laboratory experiment, finding that noise adversely affected performance but had little impact on cortisol when effort was low. When participants worked under high effort on a task during noise, the opposite profile emerged: no performance effects of noise but elevated physiological stress. See Table VI for an illustration of some of their findings.

Field studies have uncovered analogous job demand by occupational noise interactions on physiological stress. In one study, the impact of noise on blood pressure more than doubled under high- vs low-demand jobs. A similar effort by stressor interaction on performance and physiology was found in a laboratory investigation of crowding.

### 1.4. Overgeneralization

Humans respond to environmental insults with strategies to optimize the balance between demands and resources. If the demands are chronic, however, coping strategies can become habitual, persevering even when they are not appropriate. Adaptation level shifts in environmental perception may also occur. This is when the reference criterion one uses to judge a current stimulus is influenced by prior history with that

same stimulus. These shifts in referential criteria reflect habituation or reduce sensitivity to the stimulus.

### 1.4.1. Coping Perseveration

Children chronically exposed to noise appear to adopt a cognitive strategy of tuning or filtering out the noise. This strategy, while adaptive when noise is present, may be harmful if it becomes a characteristic strategy for processing auditory information, including meaningful stimuli such as speech. Cohen *et al.* showed that the louder the exposure to traffic noise at home, the more trouble children had in a task that assessed their abilities to discriminate between phonemes (e.g., the sound for b and d, or p and b). The children were prescreened for normal hearing thresholds. As shown in Table VII, noise exposure impairs phoneme perception, which in turn damages reading acquisition. SES controls were incorporated. Other noise studies have found similar relationships.

Perseveration of coping strategies has also been uncovered in studies of crowding. A common strategy for coping with crowding is social withdrawal. If people cope with overcrowded living conditions by socially withdrawing, an unintended consequence might be the breakdown of socially supportive relationships. Evans *et al.*, Lakey, and Lepore *et al.* found evidence for lower levels of perceived support in more crowded residences. The higher the ratio of people per room, independent of SES, the lower the levels of social support. These studies also showed that the link between residential crowding and poor psychological health was mediated by social support.

The adverse effects of residential crowding on social support appear to be caused by social withdrawal. In one study, college students who lived in high- or low-density apartments were brought into the laboratory. The lab was not crowded. The subject and a

TABLE VI  
Stressor by Effort Interaction on Physiological Stress and Task Performance

	Low effort	High effort
Quiet		
Reaction time	5138	4250
Cortisol	.022	.013
Noise		
Reaction time	8110	5987
Cortisol	.024	.026

TABLE VII  
Chronic Noise Exposure, Auditory Discrimination, and Reading Acquisition

Variable	Variance accounted for (%)
Noise	Phoneme perception = 19
Noise	Reading = 4
Noise, after statistically partialling out phoneme perception	Reading = 0

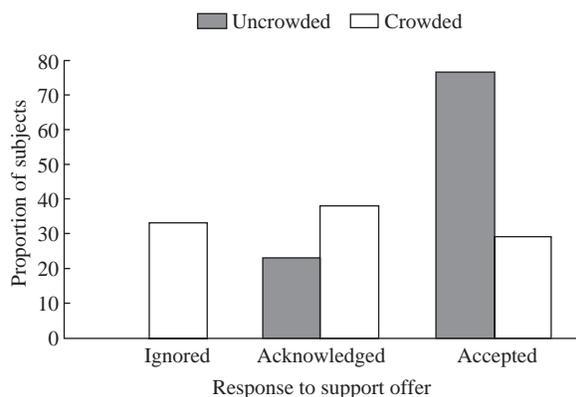
confederate were placed into situations in which they each needed social support at different times. As shown in Fig. 4, subjects who lived in crowded homes were significantly more unresponsive to the confederate's offers of support. The obverse occurred as well.

**1.4.2. Adaptation Level Shifts**

People may become so accustomed to suboptimal environmental conditions that they become desensitized to poor environmental quality. For example, an auditory stimulus will be perceived as softer if loud sounds have recently been experienced. Berglund *et al.* found that a given level of noise is less annoying if one has a history of living in a noisy area. Parallel results have been found for chronic residential crowding history and judgments of perceived crowding.

Wohlwill and Kohn examined residential histories and judgments of environmental quality. Adults who had recently migrated to a medium-sized community (Harrisburg, Pennsylvania) were asked to judge the attributes of their new community. Migrants from urban areas judged Harrisburg as significantly less noisy, less crowded, and less polluted than did their counterparts from small towns. In a second study, photographs of environmental qualities of communities varying in scale were shown. Adults from small towns consistently perceived greater levels of noise and crowding for a given photograph compared to individuals from large metropolitan areas.

People's judgments of perceived air quality habituate with continued experience of pollution. Flachsbart and Phillips showed that smaller changes in smog are



**FIGURE 4** Proportion of residents from crowded and uncrowded apartments who ignored, acknowledged, or accepted confederate's social support.

**TABLE VIII**  
Signal Detection Data for Recognition of Smog for People with Long-Term or Little Prior Experience with Air Pollution

History of air pollution	Response criterion	Detection sensitivity
Long-term experience		
Scene 1	.709	2.171
Scene 2	.637	2.052
Little prior experience		
Scene 1	.211	2.751
Scene 2	.276	2.340

required to alter judgments of perceived air quality for those unaccustomed to poor air quality compared to those who have prior familiarity with smog. Evans *et al.* analyzed pollution perception with signal detection theory. Signal detection theory enables one to separate perceptual processes from response criteria when judging stimuli. For example, if one is expecting an important phone call, one's threshold (response criteria) for responding to auditory stimuli shifts. It takes less of an indication that a bell has rung under these circumstances in order for someone to check the phone. Note that actual detection sensitivity to hear a bell is not changed, just response criterion.

College students who had recently migrated to southern California were tested. The students were well matched except with respect to prior histories of experience with air pollution. The students were shown two different, unfamiliar scenes of a distant vista. Each scene was shown multiple times with varying degrees of photochemical smog present. Subjects reported whether they saw smog and rated their confidence in their judgments. Students with prior histories of air pollution were less likely to report that pollution was present in a scene compared to people relatively unfamiliar with pollution (Table VIII). Actual visual thresholds of seeing air pollution (detection sensitivity) were equivalent.

**2. CONCLUSION**

The adaptive capabilities of human beings enables us to cope with a wide range of environmental conditions. Table IX provides a summary of some of the psychological consequences of human attempts to cope with suboptimal environmental quality.

**TABLE IX**  
**Adaptive Costs of Coping with Suboptimal Environmental Conditions**

Category of adaptive cost	Types of effects
Cumulative fatigue	<p><i>Behavioral aftereffects:</i> Performance deficits following exposure to environmental stressors</p> <p><i>Spillover:</i> Negative affect, strained interpersonal relationships, or fatigue produced by exposure to an environmental stressor in a different setting</p> <p><i>Diminished coping with multiple stressors:</i> Reduced ability to cope with a subsequent stressor following exposure to an environmental stressor</p>
Learned helplessness	<p><i>Cognitive consequences:</i> Lower mastery, reduced self-efficacy, greater externality in locus of control from chronic environmental stressor exposure</p> <p><i>Motivational consequences:</i> Diminished motivation in performance or inability to learn new tasks produced by exposure to uncontrollable environmental stressors</p>
Physiological mobilization	<p><i>Physiological stress:</i> Elevated cardiovascular and neuroendocrine activity following stressor exposure</p> <p><i>Effort by stressor tradeoff:</i> Physiological mobilization by cognitive or physical effort to maintain task performance during stressor exposure</p>
Overgeneralization	<p><i>Coping perseveration:</i> Use of well-learned coping strategies when environmental condition no longer present</p> <p><i>Adaptation level shifts:</i> History of experience with an environmental stressor changes referential criteria for environmental perception of current environmental conditions</p>

## Acknowledgments

Preparation of this article was partially supported by the John D. and Catherine T. Mac Arthur Foundation Network on Socioeconomic Status and Health, the W.T. Grant Foundation, the Robert Wood Johnson Foundation, the New Jersey Department of Transportation, the University Transportation Research Center-Region 2, and the Pittsburgh NIH Mind-Body Center (HL65111 and HL65112).

## See Also the Following Articles

Coping ■ Privacy ■ Stress

## Further Reading

- Baum, A., Gatchel, R. J., Aiello, J. R., & Thompson, D. E. (1981). Cognitive mediation of environmental stress. In J. H. Harvey (Ed.), *Cognition, social behavior, and the environment* (pp. 513–533). Hillsdale, NJ: Erlbaum.
- Berglund, B., Berglund, U., & Lindvall, T. (1975). A study of response criteria in populations exposed to aircraft noise. *Journal of Sound and Vibration*, 41, 33–39.
- Berglund, B., & Lindvall, T. (1995). Community noise. *Archives of the Center for Sensory Research*, 2, 1–195.
- Booth, A. (1976). *Urban crowding and its consequences*. New York: Praeger.
- Bullinger, M., Hygge, S., Evans, G. W., Meis, M., & van Mackensen, S. (1999). The psychological cost of aircraft noise for children. *Zentralblatt fur Hygiene und Umweltmedizin*, 202, 127–138.
- Cohen, S. (1980). Aftereffects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin*, 88, 82–108.
- Cohen, S., Evans, G. W., Stokols, D., & Krantz, D. S. (1986). *Behavior, health, and environmental stress*. NY: Plenum.
- Cohen, S., Glass, D. C., & Singer, J. E. (1973). Apartment noise, auditory discrimination, and reading ability in children. *Journal of Experimental Social Psychology*, 9, 407–422.
- Cohen, S., & Spacapan, S. (1978). The aftereffects of stress: An attentional interpretation. *Environmental Psychology and Nonverbal Behavior*, 3, 43–57.
- Cottingham, E. M., Matthews, K. A., Talbott, E., & Kuller, L. (1983). *Occupational stress and diastolic blood pressure in a blue collar population: The Pittsburgh Noise Hypertension Project*. Annual Meeting of the Society for Epidemiological Research. Winnipeg.
- D'Atri, D. A. (1975). Psychophysiological responses to crowding. *Environment and Behavior*, 7, 237–250.
- Dubos, R. (1965). *Man adapting*. New Haven, CT: Yale University Press.
- Eoyang, C. K. (1974). Effects of group size and privacy in residential crowding. *Journal of Personality and Social Psychology*, 30, 389–392.

- Epstein, Y. M., & Karlin, R. A. (1975). Effects of acute experimental crowding. *Journal of Applied Social Psychology*, 5, 34–53.
- Epstein, Y. M., Lehrer, P., & Woolfolk, R. (1978). *Physiological, cognitive, and behavioral effects of repeated exposure to crowding*. Washington, DC: American Psychological Association.
- Evans, G. W. (1979). Behavioral and physiological consequences of crowding in humans. *Journal of Applied Social Psychology*, 9, 27–46.
- Evans, G. W. (2001). Environmental stress and health. In A. Baum, T. A. Revenson, & J. E. Singer (Eds.), *Handbook of health psychology* (pp. 365–385). Mahwah, NJ: Erlbaum.
- Evans, G. W., Bullinger, M., & Hygge, S. (1998). Chronic noise exposure and physiological response: A prospective, longitudinal study of children under environmental stress. *Psychological Science*, 9, 75–77.
- Evans, G. W., & Carrere, S. (1991). Traffic congestion, perceived control, and psychophysiological stress among urban bus drivers. *Journal of Applied Psychology*, 76, 658–663.
- Evans, G. W., Hygge, S., & Bullinger, M. (1995). Chronic noise and psychological stress. *Psychological Science*, 6, 333–338.
- Evans, G. W., Jacobs, S. V., & Frager, N. B. (1982). Adaptation to air pollution. *Journal of Environmental Psychology*, 2, 99–108.
- Evans, G. W., Jacobs, S. V., Dooley, D., & Catalano, R. (1987). The interaction of stressful life events and chronic strains on community mental health. *American Journal of Community Psychology*, 15, 23–34.
- Evans, G. W., & Lepore, S. J. (1993). Household crowding and social support: A quasi-experimental analysis. *Journal of Personality and Social Psychology*, 65, 308–316.
- Evans, G. W., Lepore, S. J., Shejwal, B. F., & Palsane, M. N. (1998). Chronic residential crowding and children's well being: An ecological perspective. *Child Development*, 69, 1514–1523.
- Evans, G. W., & Maxwell, L. E. (1997). Chronic noise exposure and reading deficits: The mediating effects of language acquisition. *Environment and Behavior*, 29, 710–728.
- Evans, G. W., Palsane, M. N., Lepore, S. J., & Martin, J. (1989). Residential density and psychological health: The mediating effects of social support. *Journal of Personality and Social Psychology*, 57, 994–999.
- Evans, G. W., & Saegert, S. (2000). Residential crowding in the context of inner city poverty. In S. Wapner, J. Demick, H. Minami, & T. Yamamoto (Eds.), *Theoretical perspectives in environment behavior research: Underlying assumptions, research problems, and relationships* (pp. 247–268). New York: Plenum.
- Evans, G. W., Saegert, S., & Harris, R. (2001). Residential density and psychological health among children in low income families. *Environment and Behavior*, 33, 165–180.
- Evans, G. W., & Stecker, R. (2004). Motivational consequences of environmental stress. *Journal of Environmental Psychology*, in press.
- Flaschbart, P. G., & Phillips, S. (1980). An index and model of human response to air quality. *Journal of the Air Pollution Control Association*, 30, 759–768.
- Fleming, I., Baum, A., & Weiss, L. (1987). Social density and perceived control as mediators of crowding stress in high density residential neighborhoods. *Journal of Personality and Social Psychology*, 52, 899–906.
- Glass, D. C., & Singer, J. E. (1972). *Urban stress*. New York: Academic Press.
- Griffitt, W., & Veitch, R. (1971). Influences of population density on interpersonal affective behavior. *Journal of Personality and Social Psychology*, 17, 92–98.
- Guski, R., & Rohrman, B. (1981). Psychological aspects of environmental noise. *Journal of Environmental Policy*, 2, 183–212.
- Helson, H. (1964). *Adaptation level theory*. New York: Harper & Row.
- Herronkohl, R., & Egoff, B. (1974). *Perceived crowding, familiarity, and the choice of a residence*. American Psychological Association. New Orleans.
- House, J. S., Mc Michael, A., Wells, J., Kaplan, B., & Landerman, L. (1979). Occupational stress and health among factory workers. *Journal of Health and Social Behavior*, 20, 139–160.
- Kluger, A. (1998). Commute variability and strain. *Journal of Occupational Behavior*, 19, 147–165.
- Kryter, K. (1994). *The handbook of hearing and the effects of noise*. New York: Academic Press.
- Lakey, B. (1989). Personal and environmental antecedents of perceived social support developed at college. *American Journal of Community Psychology*, 17, 503–519.
- Lepore, S. J., Evans, G. W., & Palsane, M. N. (1991). Social hassles and psychological health in the context of crowding. *Journal of Health and Social Behavior*, 32, 357–367.
- Lepore, S. J., Evans, G. W., & Schneider, M. (1991). The dynamic role of social support in the link between chronic stress and psychological distress. *Journal of Personality and Social Psychology*, 61, 899–909.
- Lundberg, U. (1976). Urban commuting: Crowdedness and catecholamine excretion. *Journal of Human Stress*, 2, 26–34.
- National Research Council. (1991). *Environmental epidemiology*, Vol. 1. Washington, DC: National Academy of Sciences.
- Maxwell, L. E. (1996). Multiple effects of home and day care crowding. *Environment and Behavior*, 28, 494–511.
- Maxwell, L. E., & Evans, G. W. (2000). The effects of noise on preschool children's prereading skills. *Journal of Environmental Psychology*, 20, 91–97.
- Mc Callum, R., Rusbult, C., Hong, G., Walden, T., & Schopler, J. (1979). Effects of resource availability and importance of behavior on the experience of crowding. *Journal of Personality and Social Psychology*, 37, 1304–1313.
- Mc Ewen, B. S. (2000). The neurobiology of stress: From serendipity to clinical relevance. *Brain Research*, 886, 172–189.

- Medical Research Council. (1997). *The nonauditory effects of noise*. Report R10, Institute for Environment and Health. Leicester, UK: University of Leicester.
- Melamed, S., Fried, Y., & Froom, P. (2001). The interactive effect of chronic exposure to noise and job complexity on changes in blood pressure and job satisfaction: A longitudinal study of industrial employees. *Journal of Occupational Health Psychology, 6*, 182–195.
- Moch-Sibony, A. (1984). Study of the effects of noise on the personality and certain psychomotor and intellectual aspects of children after a prolonged exposure. *Travail Humain, 47*, 155–165.
- Novaco, R. W., Kliewer, W., & Broquet, A. (1991). Home environmental consequences of commute travel impedance. *American Journal of Community Psychology, 19*, 881–909.
- Novaco, R. W., Stokols, D., Campbell, J., & Stokols, J. (1979). Transportation stress, and community psychology. *American Journal of Community Psychology, 7*, 361–380.
- Novaco, R. W., Stokols, D., & Milanese, L. (1990). Objective and subjective dimensions of travel impedance as determinants of commuting stress. *American Journal of Community Psychology, 18*, 231–257.
- Paulus, P. B. (1988). *Prison crowding: A psychological perspective*. New York: Springer-Verlag.
- Peterson, C., Maier, S., & Seligman, M. E. P. (1993). *Learned helplessness*. New York: Oxford.
- Regoceva, V., & Kellcrova, E. (1995). Effects of urban noise pollution on blood pressure and heart rate in school children. *Journal of Hypertension, 13*, 408–412.
- Repetti, R. L. (1993). The effect of workload and the social environment at work on health. In L. Goldberger, & Breznitz, S. (Eds.), *Handbook of stress*, 2nd ed. (pp. 368–385). New York: Free Press.
- Rodin, J. (1976). Density, perceived choice, and response to controllable and uncontrollable outcomes. *Journal of Experimental Social Psychology, 12*, 564–578.
- Rotton, J. (1983). Affective and cognitive consequences of malodorous pollution. *Basic and Applied Social Psychology, 4*, 171–191.
- Ruback, R. B., & Pandey, J. (1991). Crowding, perceived control and relative power: An analysis of households in India. *Journal of Applied Social Psychology, 21*, 315–344.
- Ruback, R. B., & Pandey, J. (1992). Very hot and really crowded. *Environment and Behavior, 24*, 527–554.
- Rydstedt, L. W., Johansson, G., & Evans, G. W. (1998). The human side of the road: Improving the working conditions of urban bus drivers. *Journal of Occupational Health Psychology, 3*, 161–171.
- Saegert, S. (1978). High density environments: Their personal and social consequences. In A. Baum, & Y. M. Epstein (Eds.), *Human response to crowding* (pp. 259–282). Hillsdale, NJ: Erlbaum.
- Schaeffer, M. A., Street, S., Singer, J. E., & Baum, A. (1988a). Effects of control on the stress reactions of commuters. *Journal of Applied Social Psychology, 11*, 944–957.
- Sherrod, D. R. (1974). Crowding, perceived control, and behavioral aftereffects. *Journal of Applied Social Psychology, 4*, 171–186.
- Sherrod, D. R., & Downs, R. (1974). Environmental determinants of altruism: The effects of stimulus overload and perceived control on helping. *Journal of Experimental Social Psychology, 10*, 468–479.
- Stokols, D., Novaco, R. W., Campbell, J., & Stokols, J. (1978). Traffic congestion, Type A behavior, and stress. *Journal of Applied Psychology, 63*, 467–480.
- Tafalla, R. J., & Evans, G. W. (1997). Noise, physiology, and human performance: The potential role of effort. *Journal of Occupational Health Psychology, 2*, 148–155.
- van Kamp, I. (1990). *Coping with noise and its health consequences*. Groningen, Netherlands: Styx.
- Welch, B. L. (1979). *Extra auditory health effects of industrial noise: Survey of foreign literature*. AHRL-TR-79-41. Aerospace Medical Research Laboratory. Dayton, OH: Wright Patterson Air Force Base.
- Wener, R., Evans, G. W., Phillips, D., & Nadler, N. (2003). Running for the 7:45: The effects of public transit improvements in commuter stress. *Transportation, 30*, 203–220.
- White, S., & Rotton, J. (1998). Type of commute, behavioral aftereffects, and cardio-vascular activity. *Environment and Behavior, 30*, 763–780.
- Wohlwill, J. F., & Kohn, I. (1973). The environment as experienced by the migrant: An adaptation level view. *Representative Research in Social Psychology, 4*, 135–164.
- Wohlwill, J. F., & Kohn, I. (1976). Dimensionalizing the environmental manifold. In S. Wapner, S. Cohen, & B. Kaplan (Eds.), *Experiencing the environment* (pp. 19–53). New York: Plenum.



# Environmental Versus Individual Risk Taking: Perception, Decision, Behavior

Charles Vlek

University of Groningen, Groningen, The Netherlands

1. Risk Levels, Definitions, and Complications
  2. Models and Mechanisms of Individual Risk Taking
  3. Comparative Judgment of Multidimensional Risks
  4. External Risk Acceptance: Rules, Heuristics, and Procedures
  5. Collective Risks in Commons Dilemmas
  6. Conclusion
- Further Reading

## GLOSSARY

**commons dilemma** The (un)balance between individual interests and collective goods, depending on the discounting of temporally, spatially, and/or socially distant decision consequences of those involved.

**coping appraisal** Overall judgment of one's potential to handle an external threat, in terms of response efficacy, self-efficacy, and costs and barriers of action.

**external versus internal control** Events may occur through (external) "acts of god," (internal) human factors, or both. This underlies the difference between gambling tasks and control tasks.

**frequency versus scenario information** Historical relative frequencies versus future possible ways of accident occurrence, as bases for probability estimation.

**judgmental heuristics and biases** Simplified ways and tendencies to estimate probabilities and other uncertain quantities.

**perceived risk** Personal judgment, on various grounds, of the nature and seriousness of risk, often contrasted with expert-assessed or calculated risk based on scientific models and/or databases.

**precautionary principle** The willingness to decide for (costly) protective measures or against risky activity if there is reasonable suspicion of serious threat.

**risk** The possibility of a negative event (harm, loss, accident, disease, or disaster).

**risk attitude** The tendency either to venture something or not in view of the safe(r) status quo, depending on utilities of possible decision consequences.

**risk homeostasis** A risk taker's tendency to use externally increased safety for enhanced benefit achievement, thereby maintaining or returning to the original (target) risk level.

**threat appraisal** Overall judgment of an external threat, in terms of perceived seriousness, vulnerability, and associated benefits.

**utility function** The formal relationship between personal utility and the objective value of decision consequence variables (e.g., monetary outcome, driving speed, number of fatalities).

Managing risks is an everyday task of life, appearing in different scopes and scales. To understand environmental

risk management, we must be familiar with individual risk taking. In this article, basic concepts and theories about risk are reviewed, along with empirical findings about judgmental and behavioral mechanisms. For risk taking, perceived control often is more important than calculated gambling. The major dimensions of perceived technological risk are the undesirability of the consequences, perceived controllability, and importance of associated benefits. Collective risks often arise from negative external effects of numerous individual activities, such as driving a car. Several factors of collective rationality (or cooperation) are discussed, along with various intervention strategies for achieving a sustainable balance in such commons dilemmas. Conclusions focus on psychologically enlightened safety management, both small-scale and large-scale.

## **1. RISK LEVELS, DEFINITIONS, AND COMPLICATIONS**

On December 9, 2003, four Latvian mountaineers were climbing Mount Cook, the highest peak of New Zealand. To ensure each other's safety, they were all connected to the same rope. Suddenly, one of them slipped on the icy slope. The other three at that moment were not standing firm enough to absorb the shock of their falling colleague. Unfortunately, the four sportsmen fell 300 meters and died on the spot.

In November 2003, more than 200 people died, an entire village was destroyed, and much farmland was turned useless by abundant river floods in the Northern Sumatran district of Lankat (Indonesia) following days of heavy monsoon rainfall. Wholesale deforestation, undertaken for the benefit of logging companies, appeared to be the major cause; it had been unofficially tolerated for many years despite the warnings from environmental protection groups and local communities.

### **1.1. Research Origins**

Things can always go wrong. However promising or well-intended, any activity or situation may result in loss, harm, accident, or disaster (henceforth referred to as loss). Sometimes we are personally responsible and have to bear the damage ourselves. In other cases, risks are collective, arising through natural events, through the actions of others, or through both.

Statistics and economics have long dominated the thinking of researchers and policymakers about risk.

Mathematical statistics has laid the grounds for probabilistic risk analysis in the engineering sciences. Some 25 years ago, this elicited psychometric research on perceived risks of technologies. Neoclassical economic theory has evoked 40-year-old behavioral decision research, in which psychologists gradually developed authentic concepts and theories of their own. Social psychologists increasingly study commons dilemmas, capturing the tension between individual interests and collective goods, such as environmental sustainability.

Here, an overview of concepts, models, and findings on human risk judgment and decision making about safety is provided. The emphasis is on conceptual clarification and interrelation from different research directions. First considered is the concept of risk and individual risk taking, then the evaluation and management of collective environmental risks.

### **1.2. Different Levels of Risk**

In risk taking there seem to be two extremes. On the small-scale side, the quality of a challenging activity, such as mountain climbing, car driving, or tight-rope walking across Niagara Falls, may actually suffer from the risk awareness of the participant, since that awareness might distract attention and raise worries, whereas the self-confident subject tries to perform optimally and prevent accidents from occurring. On the large-scale side, people may feel uneasy or threatened by environmental conditions such as ambient air pollution, street crime, or the chance of flooding. Such situations stimulate explicit risk assessment and the design of coordinated safety strategies, often involving many people.

For the moment, loosely defining risk as the possibility of undesirable consequences of an activity or situation (henceforth: activity), risks may be categorized as occurring on seven different levels of complexity, as shown in Table I, together with illustrative examples.

Table I shows varieties of risk associated with single, low-level activities that are normally accompanied by a relative unawareness of risk; safety is warranted by due attention, motivation, and experienced behavioral control. Table I also gives examples of worrisome risks whose cause or (more often) effective control is not sufficiently clear. One may roughly distinguish three types of risks: clearly perceptible risks (as in car traffic), scientifically clear but not publicly well-known risks (as from soil pollution), and new, unknown, or ambiguous risks (as from genetic engineering) that may call for precautionary measures.

**TABLE I**  
Different Scale Levels of Risk and Risk Management

Level	Key words to illustrate risk (as "possible loss from ...")
Personal	Smoking, drinking, illness, sports
Indoor	Air quality, radon, do-it-yourself jobs
Local	Traffic, toxic waste, urban smog
Regional	Air, soil, and groundwater pollution; GMO crops
Fluvial	Pollution of rivers, lakes, coastal zones; flooding
Continental	Acid precipitation, deforestation, glacier melting
Global	Ozone-layer depletion, climate change, sea level rise

### 1.3. Definitions of Risk

Following traditional statistics, risk is the probability of loss, where probability is estimated on the basis of long-term observed relative frequency of comparable losses. In business economics, in decision theory, and in practical safety management, risk also has other, both extended and more subjective definitions. Table II lists six formal definitions of risk that may be used for quantitative estimation.

The psychological problem with any definition in Table II is that both probability and size of loss are essentially judgmental components. Probability may be objectified, but only under rather limiting conditions, such as, for example, the assumption that the relevant system will behave in the future as it did in the (observed) past and, obviously, that a sufficient

**TABLE II**  
Formal Definitions of Risk

<ul style="list-style-type: none"> <li>• Probability of specified loss</li> <li>• Size of possible loss</li> <li>• Product of probability and size of loss</li> <li>• Variance of total distribution of losses and benefits of decision alternative</li> <li>• Relative weight of possible loss versus comparable gain (following utility curve)</li> <li>• Two-dimensional set of points indicating probabilities of possible different losses<sup>a</sup></li> </ul>
--

<sup>a</sup>This may characterize, e.g., a factory, household, or transport on a multi-annual basis.

number of observations is available to compute relative frequencies. Size of loss similarly presupposes a fair specification of what such loss really means and how it eventually all works out (or not). Multi-attributable notions of loss (as a function of, e.g., people killed and wounded, property damage, environmental pollution) do not easily fit into classical risk definitions.

On an aggregate level (Table I), the international debate on climate change illustrates the scientific difficulties of environmental risk assessment and management.

### 1.4. Development of Risk Psychology

In critical response to technological risk analysis, social scientists, particularly psychologists, have explored people's risk perceptions and behaviors. Thus, a body of theoretical and empirical knowledge has developed that represents nontechnical and noncomputational ways of conceptualizing and grasping risk as an essential component of human performance, deliberation, and emotionality.

One result of this development is that nowadays the objectification of technological risk has limited (only "bandwidth") potential and that the actual assessment, decision making, and control of external risks is to a large extent a procedural and often multi-stakeholder matter involving attention, effort, competence, and communication with respect to technical as well as behavioral and social components of risky activities. Many environmental risks came to be seen as collective effects generated by numerous individual activities in commons dilemma situations (see Section 5). One challenge for psychology is to expand from existing theory and methodology about small-scale risk taking toward a better understanding of human judgment and behavior vis-à-vis large-scale environmental risks.

## 2. MODELS AND MECHANISMS OF INDIVIDUAL RISK TAKING

This section reviews a number of theoretical models and issues, giving various insights into the way we judge risks and deal with risky situations.

### 2.1. Classical Decision Theory

Following classical economic theory, a simple choice between two gambles may be portrayed as in Fig. 1,

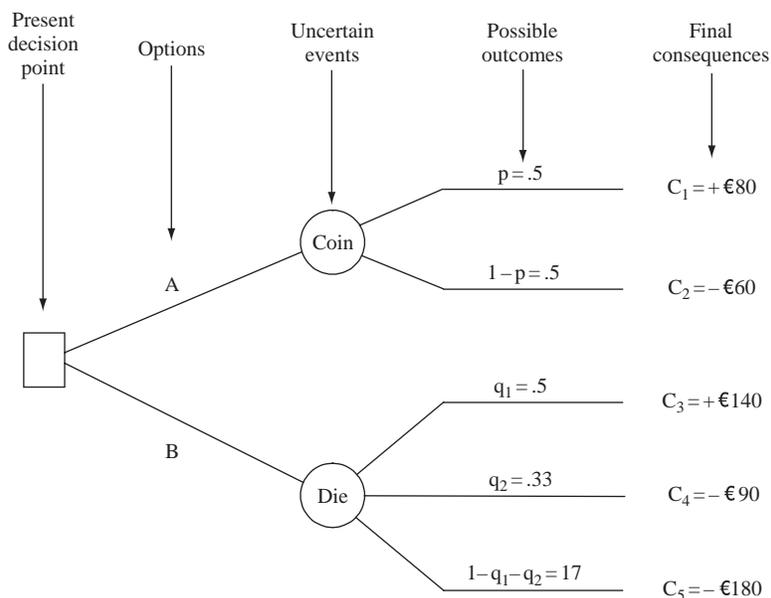


FIGURE 1 Decision tree representation of the choice between gambles A (flipping a coin) and B (throwing a die with possible outcomes 1–3, 4–5, or 6) and monetary consequences in euros.

which shows the perspective of a decision maker in terms of present options, possible outcomes of uncertain factors, and monetary consequences. Note that the expected value (EV) of both gambles A and B, that is, their long-term average return on repeated playing, amounts to €10. For a single play, which gamble would you prefer, and why? A rational EV maximizer’s answer would be, “Either, because both gambles are equally worthy.” EV theory, however, has been succeeded by expected utility (EU) theory, in which the subject’s utility function for money is crucial. One traditional shape of such a curve is drawn in the right half of Fig. 2, which reflects the subject’s supposed decreasing marginal utility of money. This would, for example, imply the subject’s refusal to play a 50–50 bet for a fixed (sizeable) amount of money, since the possible loss weighs heavier than the comparable gain. Thus, in the example of Fig. 1, both gambles would be unattractive, but actually gamble A has a higher EU than gamble B.

As regards risk, in view of Table II, we may note that A and B have an equal (total) probability of loss, B has the highest possible size of loss, B has the highest (summed) product of probability and size of loss, and B has the highest outcome variance. So, gamble B is the most risky. One should realize, however, that not risk but EU is the rational basis of choice.

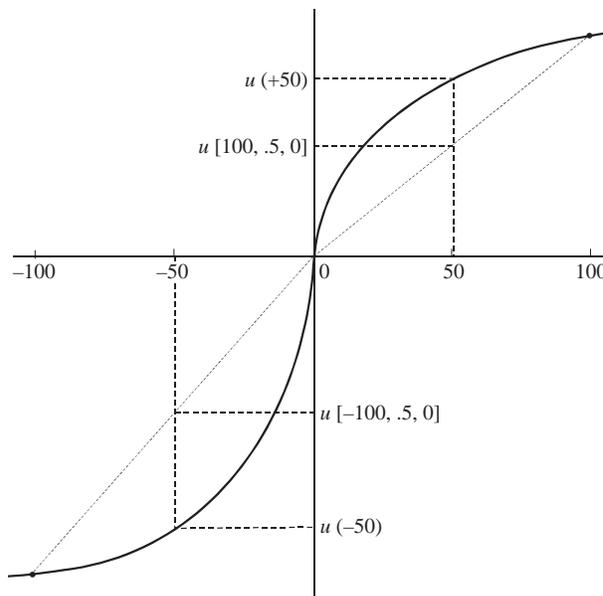


FIGURE 2 Two-component value function following prospect theory, indicating risk avoidance in the domain of gains and risk seeking in the domain of losses. Monetary amounts are indicated on the horizontal axis, utilities on the vertical one.

The negative acceleration of the utility curve in Fig. 2 fits the well-known Weber–Fechner law of psychophysics: as stimulus intensities increase, people become less sensitive to given intensity differences. For monetary amounts, the difference between 1020 and 1000 euro is psychologically less than the difference between 120 and 100 euro. This notion of decreasing marginal utility also applies in the negative (see Fig. 2). For example, the psychological difference between 120 and 100 (possible) accident fatalities is less than the difference between 20 and 0 fatalities.

## 2.2. Prospect Theory

The preceding reasoning is captured in a generalized value function following prospect theory as proposed by Kahneman and Tversky in 1979; see Fig. 2. Here, the utility function for losses is steeper than the utility function for gains: “losses loom larger than gains.” This double utility function has special implications for gambling in the domains of (only) losses and gains. Suppose the subject is offered the choice between receiving €50 for sure and a 50–50 bet to win either €100 or nothing (written as [100, .50, 0]). Following the utility ( $u$ ) curve for gains,  $u(+50)$  is significantly higher than  $u[100, .50, 0]$ ; see the “positive utility” axis in Fig. 2. Thus, the subject will avoid the risk of getting nothing and will prefer the sure amount of €50.

Now suppose that the same subject is offered the choice between a sure loss of €50 and a 50–50 bet to lose either €100 or €0 (written as [−100, .50, 0]). As the “negative utility” axis reveals,  $u(-50)$  is significantly lower (more negative) than  $u[-100, .50, 0]$ . Hence, the subject will prefer the bet over the sure loss, thus choosing the risk of losing as much as €100, in the hopes of actually obtaining the “loss” of €0.

The main implications for risk taking of the double utility function represented in Fig. 2 are the following:

- Following Weber–Fechner, the greater the absolute value of two possible outcomes (losses or gains), the smaller a given difference between them is perceived.
- Losses weigh more heavily than do comparable gains (this means, for example, that a 50–50 bet of losing or winning €100 will be rejected).
- Subjects are risk seeking in the domain of losses but are risk averse in the domain of gains.
- The perception of gains and losses depends on the subject’s reference point, which may be, for example, current assets, last year’s earnings, maximum outcome, or “worst case.”

From a rationalistic perspective, the last point is rather worrying. As Kahneman and Tversky and others have shown, smart reframing of a decision problem may induce a preference change in the subject, who thereby would behave inconsistently across different frames. Framing may also be conducted in other ways, as for example when any particular accident scenario or risk dimension is expressly presented in a conspicuous way in order to cause it to be more heavily weighted than other scenarios or dimensions.

## 2.3. Probability Estimation

Considering risk, probability is the likelihood of a future loss. Probability, however, is subject to different conditions, interpretations and estimation heuristics, so that numerical values may be of limited validity and reliability.

### 2.3.1. Probability Interpretations

Debates about risk may be complicated by three different interpretations of probability.

1. Frequentistic probability is the limit of a (past) relative frequency in a theoretically endless number of observations of an uncertain event (e.g., the flip of a coin or the crossing of a road junction) for a specified outcome (e.g., “tails” or a car accident) under substantially similar conditions.
2. Logical probability is a quantitative characteristic of a chance apparatus such as a coin, die, or wheel of fortune, which may be derived from its properties such as the number of possible outcomes, symmetry, and lack of bias.
3. Personalistic probability is someone’s degree of belief in the future occurrence of an uncertain outcome (e.g., a nearby forest fire next year), which may be expressed in the subject’s relative willingness, expressed in the odds proposed, to bet on it.

### 2.3.2. Estimation Heuristics

For practical risk assessment involving small likelihoods, probability estimation often is as crucial as it is difficult. Assessment difficulty may be due to

- incomplete or inconsistent information from various sources,
- unrepresentativeness of historical frequency data,

- impossibility or implausibility of logical system modeling (e.g., for industrial process failure), and
- lack of substantive knowledge, experience, and imagination as a basis for expert judgment (e.g., about system failure scenarios).

Moreover, in many practical situations, the probability of things going wrong is rather elusive, as, for example, in mountain climbing, road transport, and marine navigation. Thus, risk takers may use various mental tricks to estimate probabilities (when required) as a basis for risk comparisons, acceptance decisions, or safety maneuvers. Well-known probability heuristics are estimation by representativeness, cognitive availability, and anchoring and adjustment.

An event, for example, a particular series of sunny, cloudy, and rainy days, may be judged to be more probable when it is believed to be more representative of the underlying event-generating process, namely, the local climate.

Cognitive availability plays a role to the extent that examples may come to mind (e.g., financial losses, sad divorces, environmental accidents) that resemble the outcome whose probability is to be estimated. This availability heuristic also explains the great sensitivity and fear of car (or bus or train) passengers, especially non-drivers, who have actually experienced an unforgettable accident. "Never again" is often a motive for quickly enhanced safety measures in otherwise normal risk situations.

Anchoring and adjustment is an heuristic maneuver for coming up with a numerical probability estimate under a fair amount of second-order uncertainty. The idea here is that either a middle or an end scale value offers an anchor for an initial estimate, which is then conservatively adapted.

### 2.3.3. Frequency versus Scenario Information

Probability estimates are based on differing kinds of information: relative frequencies, logical system properties,

process similarities, remembered cases, and anchor values for adapted judgment. More-encompassing foundations for probability estimation are causal schemas, cognitive scripts, or mental images of future scenarios. The fundamental difference between frequency thinking and scenario thinking is indicated in the columns of Table III. Obviously, the frequentistic probability interpretation applies to the left column, while the personalistic (belief) interpretation applies to the right column. The logical probability interpretation may be applicable to both columns, in order to compute probability estimates without performing repeated observations (left column) or after some logical modeling of a multi-component system or process (right column), such as, for example, a petrochemical plant. Experimental subjects appeared to be sensitive to both frequency and scenario information. However, in a combined presentation of frequency and scenario information, the role of frequency information appears to be significantly suppressed. This suggests that for risk estimation, future orientation is more important than looking at performance history.

Remarkably enough, many great accidents or disasters have actually occurred because they had not occurred before (frequency information was lacking) and because responsible authorities did not believe them to be possible (there were no plausible scenarios) and thus failed to take the necessary precautions.

### 2.3.4. External versus Internal Control

The rows of Table III represent a second categorization of information about probabilities, namely, external versus internal control of the outcome-generating process. A simple example is the following: in flipping a coin, the outcome is externally determined by various physical factors. In contrast, in throwing a dart at a bull's eye, success depends very much on internal factors such as experience, skills, and effort. Estimating the probability of tails for the coin flip may be done using logical or

TABLE III  
Different Kinds of Uncertainty, Depending on Available Information

	<i>Repeated events (frequencies)</i>	<i>Unique events (scenarios)</i>
External, environmental control	Relative frequencies of fortuitous or random events, as in casino gambling	System properties, inclinations, strategic plans, policy scenarios, all outside the subject's control
Internal, human control	Relative frequencies of successes and failures, as in repetitive games of skill	Suppositions, ideas, arguments, feelings about the feasibility of one's own plan, task, or ambition

frequentistic data about the coin's likely behavior. Estimating the probability of success in throwing the dart is rather a matter of talent, self-confidence, and determination.

The most relevant contrast in Table III is between repetitive uncertain events under external control (upper left cell), as in pure gambling, and unique uncertain events under internal control (lower right cell), as in the execution of a rare performance task, such as tight-rope walking across Niagara Falls.

## 2.4. When Risks and Benefits Correlate: Risk Homeostasis

In many practical situations, greater benefit may be obtained when one is willing to take higher risk; per activity, benefits and risks often correlate. This was already realized by Coombs and Avrunin in 1977, when they tried to explain so-called single-peaked preference functions. These radically differ from the standard decreasing marginal utility functions adopted by neoclassical economists; see, for example, the positive-utility function in Fig. 2. Examples are the speed of transport, the bravado of mountain climbing, and the difficulty of medical operations.

### 2.4.1. Optimal Risk Taking

For the example of car driving, Fig. 3 shows positive and negative utility as a function of increasing speed. The upper dashed curve represents the (positive) utility of expected benefits: the faster you drive, the more mobile you are and the earlier you will arrive at your chosen destination; however, there is a decreasing marginal utility of increasing benefits. At the same time, however, the driving costs (e.g., through fuel use and tire wear) and accident risks are increasing disproportionately. The negative utility curve indicates the growing marginal disutility of increasing costs and risks. Thus, as Coombs and Avrunin wrote: "The good satiates and the bad escalates." The solid middle curve in Fig. 3 represents the sum of positive and negative utilities, that is, the overall net expected utility at the various driving speeds. As this sum curve shows, the optimal speed lies somewhere halfway along the speed axis; the subject has a single-peaked preference function for driving speed.

Now let us imagine (not drawn in Fig. 3) that by some external safety measure, for example, the installation of crash barriers along freeways or air bags in cars, the negative utility curve for costs and risks moves slightly

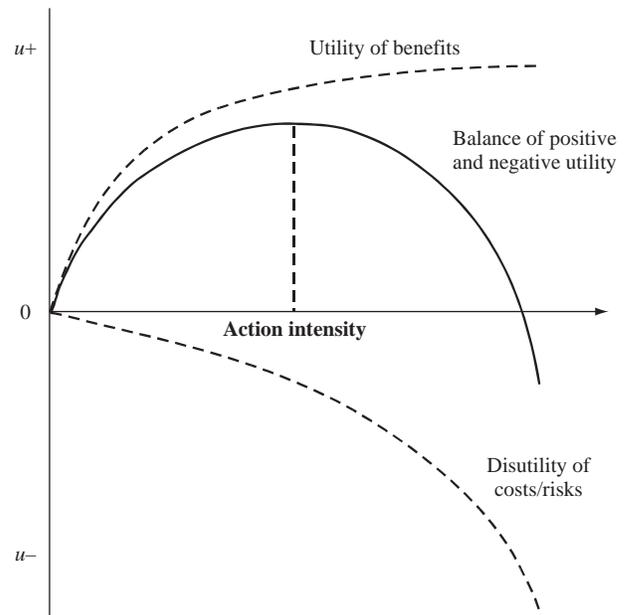


FIGURE 3 Single-peaked utility function (solid curve) for action intensity (e.g., driving speed) as the sum of positive utility of benefits and negative utility of costs and risks, both increasing as action intensifies.

upward. As a result, the summed (solid) expected utility curve would also move slightly upward, so that the optimal car speed is increased (toward the right). This, however, would put the car driver back at roughly the previous negative utility level for costs and risks, thus undermining the intended effects of the external safety measure. In fact, it appears that the subject uses the increased external safety for realizing increased mobility, while the average number of costly incidents and accidents remains the same.

### 2.4.2. Risk Homeostasis and Safety Regulation

The basic idea expressed by Coombs and Avrunin in 1977, as represented in Fig. 3, has independently been developed in Wilde's risk homeostasis theory. By now, the latter has been well validated, particularly in the domains of road transport, health care, and sports. The great lesson from this work for safety policymakers is that the effects of external safety measures may be canceled by subjects' desire for (even) greater performance benefits.

Risk homeostasis, however, would only apply to activities that leave subjects enough freedom to either intensify their performance or not, as in mountain

climbing, car driving, or performing medical operations. Putting an upper limit to the activity's intensity, for example, by declaring a prohibition on climbing certain mountain slopes or by setting a speed limit on highways, would prevent subjects from using increased safety for achieving greater benefit.

Therefore, increasing the safety of human endeavors seems to require the fulfillment of either one or both of two basic conditions:

1. The extent or intensity of the activity is limited.
2. The performer of the activity enhances his or her intrinsic safety motivation; that is, he or she will perceive the risks as higher than before.

For difficult-to-limit activities such as road transport, challenging sports, and competitive medical care, ultimate safety levels especially depend on the seriousness of perceived risks and the performer's self-control, thereby preventing serious accidents from occurring. On the other hand, when subjects persistently underestimate or belittle the risks, policymakers have no other recourse than to limit the relevant activity.

### 2.4.3. Risk-Taking Illusions

With reference to Fig. 3, two kinds of illusion may be identified. First, subjects may unrealistically believe that, on further behavior intensification (e.g., faster driving), significantly more benefit may be obtained. Second, subjects may erroneously believe that, on behavior intensification, the costs and risks of the activity will still be manageable. Both the illusion of achieving (still) greater benefit and the illusion of effective risk control may tempt subjects to perform at a higher than optimal level. The end result may be a serious accident involving great loss plus the loss of any benefit originally aspired for. As the saying goes, "It's better to arrive a little late than to end up in hospital."

## 2.5. Protection Motivation Theory

Used chiefly in health psychology, protection motivation theory (PMT) was developed by Rippetoe and Rogers in 1987 to explain individuals' risk taking and safety policies. PMT is strongly built on cognitive stress theory as developed by Lazarus, Folkman, and others prior to 1984. In 1996, PMT was correspondingly adapted by Gardner and Stern for application to situations of environmental risk. Like cognitive stress theory, PMT comprises a threat appraisal process and a

coping appraisal process. See Fig. 4, in which a motivational element has been added to each.

First, threat appraisal rests on the subject's values ("What do I cherish?") and involves perceived severity and vulnerability as well as a judgment of associated benefits; one result is the subject's feelings of fear and/or temptation. Second, coping appraisal pertains to perceived response efficacy ("Could anything be done?"), self-efficacy ("What could I do myself?") and costs and barriers of acting; one result here is the subject's perceived control and his or her motivation to act.

To the extent that the appraisal of coping potential measures up to the appraisal of threat severity, the subject may engage in problem-focused coping aimed at reducing the threat and/or increasing his or her own coping potential. If, however, severity of threat is perceived as overwhelming the subject's coping potential, the subject may (have to) resort to emotion-focused coping, aimed at living with the threat (which may amount to denying it).

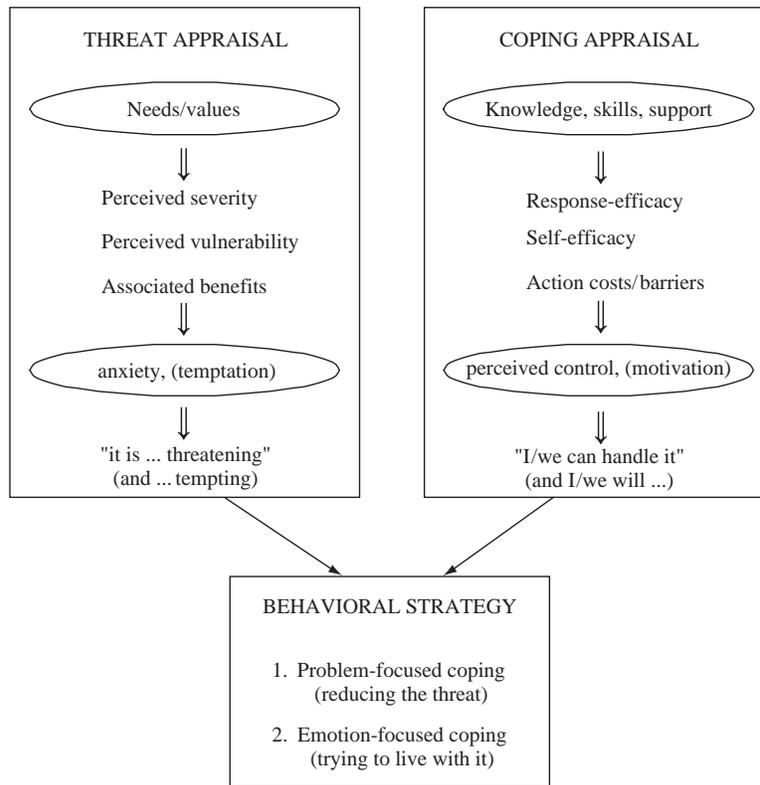
Perceived controllability appears to be so important that, psychologically, risk may well be defined as a lack of perceived control in view of significant possible (i.e., non-excludable) negative consequences of an activity or situation. When threat appraisal is low and so is coping appraisal ("Little could be done anyway"), engagement in any other behavior is unlikely. Alternatively, under low threat appraisal and high coping appraisal, subjects would easily take some safety precautions or adopt a safer behavior alternative.

Note that under PMT, subjects have to weigh their coping potential (right side of Fig. 4) up against the appraised threat (left side). This cognitive-affective balancing ("Can I handle the situation?") should somehow accompany the motivational balancing between the utility of associated benefits and the disutility of expected costs and risks, as discussed in Section 2.4 ("Am I willing to take the risks?").

Thus, the essential message of adapted PMT is that risk (and/or stress) management always has two sides. On the one hand, you may work on reducing the threat and the demands of the external environment, which, however, may tempt you via associated benefits. On the other hand, you may work on reducing the vulnerability and enhancing the self-efficacy of the (potential) victims, which, however, may be costly and difficult.

## 2.6. Emotional Significance of Risk

As marketing specialists very well know, emotions can deeply influence human judgment and decision making.



**FIGURE 4** Adapted version of protection motivation theory (PMT), with behavioral strategy as a result of the balance between threat appraisal and coping appraisal.

Emotions may be triggered by images of possible decision consequences, which themselves may be based on personal experiences, vivid presentation, proximity in time, and/or cultural conditioning. “Hot” anxiety-sensation thinking is the natural counterpart of “cool” risk-benefit thinking. Just as risks and benefits are balanced cognitively, so must anxiety and sensation be balanced emotionally. This may happen in mutual correspondence, so that cognitions and emotions accompany or even support each other. Not infrequently, however, emotional (“gut”) responses may precede and dominate cognitive responses, as may be easily seen in cases of extreme anxiety, fear or time pressure.

### 2.6.1. Negative Meaning of Risk

Anxiety and fear are prominent emotions in the face of risk. Fear may severely restrict or virtually suffocate deliberate problem solving. Fear of great loss may lead to gambling behavior aimed at trading the greater loss for a smaller one (see Section 2.2). Chronic fear is expressed in phobias, for example, for crowded

places, air travel, heights, or certain kinds of animal. Under great anxiety or fear, most people are mentally shortsighted and rather insensitive to rational, economic aspects of risky decision making, especially probability.

The strong priority effect of fear on information processing and (thus) decision quality may block subjects’ attention for reassuring information and arguments from more distant authorities. In cases of large-scale involuntary risk, this may cause serious problems of communication between risk-exposed lay subjects and non- (or less-) exposed experts and regulators. Debates about nuclear power generation, liquefied petroleum gas transport, and storage of toxic wastes may provide examples here.

As further elaborated in Section 3.3, subjects’ trust in technical experts and responsible authorities is another vital source of emotions. The building of trust not only requires the provision of essential, balanced, and credible information, but also demands that subjects’ emotions are recognized and appreciated, as a start toward more deliberative thinking.

### 2.6.2. Positive Meaning of Risk

Why do people rope-walk across Niagara Falls? Why do they hang-glide from the Eiffel Tower, bungee jump from a high railway bridge into a deep canyon, or start fights with other fans in European soccer stadiums? Why does the public flock to see motor races, air shows, or hazardous circus acts? What explains the increased popularity of so-called survival holiday tours?

As these examples suggest, risk taking can be fun. It can give a kick. Facing risks can challenge our capacities and self-confidence. Successfully completing a risky activity may be a true learning experience. Effectively managing risks may enhance our self-esteem and social status. Without risks, life would be dull and boring. Without a certain willingness to take risks, we would not be able to develop ourselves, to learn new things, and to innovate the physical, social, and economic environment on which future safety and well-being depend.

Also, in line with protection motivation theory (Section 2.5), when coping appraisal is high, that is, when subjects feel capable, skilled, and well-equipped, a given external demand is more easily perceived as a challenge than as a threat, and people's willingness to take risks should be greater.

Even large-scale disasters can have significant positive effects. Afflicted individuals may be forced to get creative and go beyond their everyday limits. Mutual solidarity may be enhanced and social coherence stimulated. After a catastrophe, many people feel that they have endured their ordeal in rewarding togetherness.

## 3. COMPARATIVE JUDGMENT OF MULTIDIMENSIONAL RISKS

In the past 25 years, a fair amount of research on perceived risk has followed the "psychometric paradigm." The purpose of this inductive approach was to discover the variables and dimensions underlying comparative judgments for a great variety of different risks from industrial activities, transportation, medical treatments, and technologies, as well as household and recreational behaviors. In a much-desired response to limited technical-statistical risk analyses from the engineering disciplines, this work was aimed at answering the following basic questions:

1. If it is not just probability of loss or accident, what (else) is it that causes a particular activity to be perceived as more or less risky and/or acceptable?

2. What might explain the differences in risk perceptions between different groups of people, for example, lay persons and experts, men and women, or people of different political inclination?

3. How meaningful and how feasible are risk comparisons as a basis for risk acceptance and/or the allocation of public funds for relevant scientific research?

### 3.1. Multidimensionality of Risk

The most powerful conclusion from psychometric research is that risk is a multi-attribute construct whose meaning goes far beyond the classical notions of risk as given in Table II. In the empirical studies underlying this conclusion, various tasks have been presented to respondents. Examples are ratings of risky activities on a set of explicit risk variables, rank ordering of different activities following their perceived riskiness and/or acceptability, and sorting a collection of activities into groups of similar items. The data were analyzed using factor analysis, multidimensional scaling analysis, and/or multiple regression analysis, all capitalizing on patterns of correlation among judgment variables. In addition, theoretical analyses of risky activities were carried out by various other researchers. Based on this varied work, Table IV presents a condensed list of dimensions underlying variations in perceived risk.

The message of Table IV is twofold. First, by rating an activity (insofar as possible) on the 11 risk dimensions, one may understand why subjects consider risks to be as high or low as they seem to believe (thus, the investigator also needs some overall risk criterion measure). Second, the same 11 dimensions hold suggestions of

TABLE IV  
Dimensions of Perceived Risk and Corresponding Safety Measures

1. Potential degree of harm or fatality
2. Physical extent of damage (area affected)
3. Social extent of damage (number of people involved)
4. Time course of immediate and delayed effects
5. Probability of undesired consequence
6. Controllability (by self or expert) of situation
7. Experience with, imaginability of consequences
8. Voluntariness of exposure (freedom of choice)
9. Clarity, importance of expected benefits
10. Social distribution of risks and benefits
11. Harmful intentionality

corresponding safety policies to be followed should any dimension be a prominent driver of perceived risk.

### 3.2. Group Differences in Risk Judgment

The reasons for perceived risk listed in Table IV may be weighted differently by different groups of people. For example, some groups may care less than others about delayed consequences, or they may put unusually high trust in relevant authorities. Major factors underlying intergroup differences in multidimensional risk judgments are the following:

- people's attitudes toward or their interest in the expected benefits involved (cf. dimension 9 in Table IV; see also Section 2.4); a general finding here is that between different activities, perceived risk is lower when associated benefits are larger;
- people's knowledge, capacities, and perceived control as regards the risk (cf. dimension 6 in Table IV; see also Section 2.5); it is in these respects that differences between experts and lay persons may occur, as well as differences between women and men;
- people's world view or general sociopolitical orientation, particularly with respect to authority, technology, and environmental resource use.

### 3.3. Importance of Trust

As already indicated, perceived controllability (dimension 6 in Table IV) is a crucial risk variable. For the lower level, everyday types of risk following Table I, it is the subject him- or herself who could do much to prevent or control risks and/or to mitigate any accident effects. However, for the higher level, large-scale risks following Table I, many people are dependent on scientific experts, policymakers, and safety inspectors for their information and personal security. Thus, in many cases personal trust in others who are responsible for the matter is a vital component of subjects' feelings of safety.

In 1997, Slovic noted—like others before him—that trust is slowly developed but may be quickly destroyed. Four reasons were listed for this “asymmetry principle”:

1. Negative (trust-destroying) events are more visible or noticeable than positive (trust-building) events.
2. When well defined and noticed, negative (trust-destroying) events carry much greater weight than positive events.

3. Sources of bad (trust-destroying) news tend to be seen as more credible than sources of good news.
4. Distrust, once initiated, tends to reinforce and perpetuate distrust.

### 3.4. Undermining of Self-Confidence

The preceding four reasons were given in connection with technological, social, and environmental risks involving industry, government, and segments of the general public. However, reasons 1, 2, and 4 also pertain to individual risk taking, for example, in traffic, sports, or gambling casinos. Personal losses or accidents stand out and remain cognitively available for a long time, and they thus undermine future self-confidence (reason 1). “Losses loom larger than gains,” Kahneman and Tversky noted in 1984, and their greater weight keeps many people from straight-forward gain-loss gambling (reason 2). And once in doubt about one's own performance capacities, one may lack sufficient self-confidence for future endeavors (reason 4). Taking this together, we may generally conclude that distrusting oneself takes away the fun or kick, the learning experience, the self-esteem, and the social status of risk taking; cf. Section 2.6.2.

### 3.5. Meaning and Feasibility of Risk Comparisons

Risk comparisons may be needed

- to decide on the safest way to carry out a given activity (e.g., a medical operation);
- to decide between two or more rather different ways to achieve the same goal (e.g., electric power generation by means of coal, wind, hydro, or nuclear energy);
- to decide on the allocation of investments in various risky activity domains (e.g., road traffic, aviation, food security, sports or medical care) in order to make them safer;
- (and, as we know from above, to find out why and in what respects respondents in psychometric studies find one hazardous activity, situation, or product more risky than another).

Statistical risk comparisons (e.g. between different occupations) have often been made by following a “probability of someone dying” definition, whereby probabilities were estimated on the basis of observed relative frequencies of past deaths. However, the

psychometric research previously discussed has made it clear that in many practical cases fair risk comparisons should be multidimensional. Indeed, multidimensional risk may be formalized (e.g., on the basis of Table IV) in terms of measurements, eventually yielding a bar diagram or a graphic semantic differential representing the risk profile of a particular activity or situation.

Quantitative risk comparisons based on such multidimensional representations are rather difficult to make. They can be completely formalized, and should then follow an explicit evaluation procedure involving attribute (or dimension) selection, activity rating, attribute weighting, and weighted summation of activity ratings across all attributes. Or multi-attribute risk profiles should just be qualitatively presented as such to responsible judges for their intuitive evaluation and rank ordering of activities.

Thus, making risk comparisons is a difficult and often tricky thing to do. In any case, one needs a clear and standardized definition of risk. The safest strategy is to limit risk comparisons to alternative options that fall into the same category (e.g. train versus car; one medical treatment versus another; coal- versus gas-powered electricity generation). Indeed, psychometric risk research itself is based on the somewhat paradoxical assumption that lay respondents can compare multidimensional risks of activities that do not meaningfully fit into a single decision problem.

#### **4. EXTERNAL RISK ACCEPTANCE: RULES, HEURISTICS, AND PROCEDURES**

Theoretically, risk acceptance is implied in the choice of a preferred course of (more or less risky) action, whereby expected benefits constitute the primary motivating factor (see Section 2.1). In many cases, however, risk-exposed people do not directly enjoy the benefits of the relevant activity. Under such involuntary (or nonbeneficial) risk exposure, explicit rules and procedures may be needed to determine the acceptability of a given risk situation.

##### **4.1. Risk–Benefit Trade-Offs**

Whether we deal with a static gambling situation (as in Section 2.1) or with a dynamic control situation

(Section 2.4), risk taking often requires a weighing of risks against benefits. The larger the benefit involved, the higher the risk accepted. So, if you see someone voluntarily take a high risk, always ask yourself which compensating (or seductive) benefit the subject expects to achieve. Of course, several other factors may play a significant role as well, for example, a possible unrealistically optimistic underestimation of the probability of loss (“It won’t happen to me”), time-discounting of long-term risk (“Who cares about later?”), and an exaggerated need for the benefit involved, for example the “kick” from a hazardous sporting event. For society at large, important economic benefits often justify the toleration of appreciable risks.

##### **4.2. Risk-Limiting Standards**

For the sake of cognitive convenience, in many practical situations risk–benefit trade-offs are “frozen” into simple risk acceptance standards. Examples are vehicle speed limits on highways, the degree of attentiveness in descending a staircase, or the limitation of preservatives in food products. In such cases, without giving up the intended benefits, one is willing (or forced) to adapt the activity, at some explicit or implicit cost, in order to make it safer. Acceptance standards often have a third-party protection function, too, as illustrated by speed limits on highways.

##### **4.3. Risk as Low as Reasonably Achievable**

Some consequences of a risky activity may sooner or later be fatal. Thus, one should naturally be willing to reduce such risks to a level that is “as low as reasonably achievable.” This ALARA principle, however, may lead to a kind of phobia. For example, some may refuse to eat shellfish, some would never take an airplane trip, and some, knowing they are a carrier of a genetic disease, would decide to never have children. The ALARA principle means that personal safety strategies, although primarily benefit-driven and perhaps somehow standardized in terms of risk limits, may further involve efforts to reduce the probability of some sad loss as much as possible, even if the subject considers the situation to be safe enough otherwise. The result, however, may be that an activity becomes very costly or no longer feasible.

#### 4.4. Justification, Risk Limits, and ALARA

Risk–benefit weighing, the testing of risk acceptance standards and subsequently trying to get risks ALARA, may constitute an effective combined strategy for risk management. For example, for involuntary risks to radiological workers (e.g., in hospitals and nuclear energy laboratories), the International Commission on Radiological Protection has recommended the use of such a joint set of safety principles. This strategy comes close to an empirical conclusion drawn by Hendrickx in 1991, that in a dynamic control task, risk acceptance consists of a deliberate choice of a safe-enough course of action followed by an allocation of effort to keep the risks under sufficient control. A triple heuristic conclusion on risk acceptance was also drawn on empirical (psychometric) grounds by Vlek and Stallen in 1981: for most lay people, risk acceptance hinges on: (1) the importance of expected benefits, (2) limitation of the maximum possible accident, and (3) the perceived controllability of the activity as a whole.

#### 4.5. Societal Risk Acceptance: Importance of Procedures

Involuntary exposure to industrial or environmental risks easily raises public concerns, protests, and feelings of being treated unfairly. To the extent that expected benefits are the main driving force of risk acceptance, we may expect involuntary and mostly nonbeneficial exposure to cause much trouble for responsible decision makers. “NIMBY,” or “not in my back yard” is one protest often heard regarding the question of where to build hazardous facilities such as municipal waste incinerators, electric power stations, and airport expansions.

The reactions of involuntarily risk-exposed persons toward responsible authorities have been widely studied, and various suggestions have been formulated to facilitate the processes of decision making and eventual risk acceptance. Table V lists six different diagnoses of public protests together with recommended policy approaches.

Specialists believe that in most practical cases, several diagnoses may be valid at the same time. Hence, they recommend that administrative procedures for planning, decision making, and undertaking of a hazardous activity involving involuntary risk should follow a carefully designed procedure incorporating all or most of the therapeutic ingredients mentioned in Table V.

TABLE V  
Six Possible Diagnoses and Corresponding Therapeutic Approaches Toward “NIMBY” (“Not in My Backyard”) Attitudes

<i>Diagnosis</i>	<i>Policy approach</i>	<i>Labelling name</i>
1. Ignorance	Communication	Technocratic approach
2. Unfairness	Compensation	Market approach
3. Exclusion	Participation	Participation approach
4. Social inequity	Redistribution	Social justice approach
5. Policy criticism	Strategic alternative	Transformation approach
6. Pure local egoism	Government coercion	Regulatory approach

### 5. COLLECTIVE RISKS IN COMMONS DILEMMAS

Involuntary exposure often occurs in connection with environmental risks at the higher scale levels following Table I. In quite a few cases, collective risks are generated in a process of accumulating small contributions from numerous individual actors.

#### 5.1. Definition of a Commons Dilemma

A commons dilemma is a situation of conflict between an aggregate collective interest and numerous individual interests; see Hardin’s inspiring 1968 analysis of the disastrous consequences of common pasture land (“the commons”) being overgrazed by ever more cattle owned by individual farmers, each trying to improve their individual income. A present-day example is the urban air pollution resulting from accumulation of exhaust fumes from many individual cars.

#### 5.2. Macro and Micro Level Perspectives

At the macro level of society, a commons dilemma presents itself as a permanent contrast between a collective risk, such as air pollution, and a large collection of individual benefits, such as the benefits of car driving, for example. Minimization of risk and maximization of benefits are incompatible social goals,

between which a trade-off must be made. In contrast, individual actors at the micro level of society are naturally focused on their own benefits “here and now”. Individuals may not recognize a commons dilemma as such

- because they are unaware of any collective damage;
- because they do not appreciate their own responsibility for collective problems;
- because they do not feel that the long-term collective risk is serious enough in relation to the numerous short-term benefits; and/or
- because they know there is a collective risk, but feel that little can be done about it.

In particular, the combination of short-sightedness, personal benefit, and perceived lack of control over the situation may readily lead to a denial or playing down of the risk and thus to a refusal to participate in risk control efforts.

### 5.3. Key Issues for Research and Policymaking

Understanding commons dilemmas and managing the collective (environmental and social) risks generated by individual activities revolve around problem diagnosis, policy decision-making, practical intervention, and evaluation of policy effectiveness. These four categories can be expanded into 12 focal issues, as listed in Table VI. The 12 issues listed in Table VI imply rather different tasks to be performed by experts from different backgrounds, not just psychologists. Also, the 12 activities would have to be carried out in reasonably good order, so that subjects of intervention strategies understand why they are asked to accept or cooperate, and to what future improvements this might lead.

### 5.4. Structural and Individual Factors of Social Cooperation and Defection

A multitude of experiments and field studies have clarified the main factors governing people’s tendency to either cooperate or defect (i.e., behave selfishly) when faced with a commons dilemma. Table VII reviews the structural and individual psychological factors governing cooperation.

In connection with this, seven general strategies may be used for the promotion of cooperation in managing

TABLE VI  
Twelve Key Issues for Research and Policymaking on Commons Dilemmas

---

I. Problem diagnosis
A. Analysis and assessment of collective risk, annoyance, and stress
B. Analysis and understanding of sociobehavioral factors and processes underlying risk generation
C. Assessing problem awareness, risk appraisal, and actors’ individual values and benefits
II. Policy decision making
A. Weighing of collective risk against total individual benefits (“need for change?”)
B. If “risk unacceptable”: setting objectives for reduction of environmental and/or social risks
C. Translation of risk reduction objectives into individual behavioral goals
III. Practical intervention
A. Focusing on individual target groups and considering essential conditions for policy effectiveness
B. Specifying feasible behavioral alternatives and selecting effective policy instruments
C. Targeting group-oriented applications of strategic programs of behavioral change
IV. Evaluation of effectiveness
A. Designing a monitoring and evaluation program to determine policy effectiveness
B. Systematic, comprehensive evaluation of observable effects and side effects
C. Intermittent and <i>post hoc</i> policy feedback and possible revision of policies

---

collective risks. These are briefly indicated in Table VIII, along with one default strategy.

Policymakers regularly use separate strategies in a piecemeal fashion. However, they rarely apply a comprehensive approach to managing collective risks in commons dilemmas, involving problem diagnosis, policy decision making, and practical intervention following various related strategies.

### 5.5. Precautionary Principle

Large-scale commons dilemmas (e.g., motorized transport or genetic engineering) may yield great environmental risks that may be hard to characterize. Not only is there a variety of threatening consequences whose time of occurrence is rather uncertain, but also there is great ambiguity about the probabilities of catastrophic

**TABLE VII**  
**Factors Facilitating Cooperation to Reduce Collective Risks in Commons Dilemmas**

<i>Structural factors</i>	<i>Individual psychological factors</i>
<ul style="list-style-type: none"> <li>• Smaller group</li> <li>• Stronger group identity</li> <li>• Higher visibility of others' behavior</li> <li>• Availability of cooperative alternatives</li> <li>• Lower cost of cooperative choice</li> <li>• Higher cost of defective choice</li> <li>• Legal obligation to cooperate</li> </ul>	<ul style="list-style-type: none"> <li>• Environmental problem awareness, insight into causal processes</li> <li>• Feeling of responsibility, moral obligation</li> <li>• Perceived (relative) appeal of cooperation</li> <li>• Self-efficacy in making cooperative contributions</li> <li>• Trust in others' cooperation (low social uncertainty)</li> <li>• Ability to resist bad social examples ("rotten apples")</li> <li>• Perceived environmental effectiveness of cooperation (outcome efficacy)</li> <li>• Altruistic and cooperative social value orientation</li> </ul>

events. What people do realize is that possible catastrophes are caused by humans and so should eventually be controlled by humans.

For environmental policymakers, a reasonable suspicion of a serious threat may offer sufficient grounds for applying the precautionary principle. This means that various risk avoidance, prevention, and/or reduction actions are implemented, aimed at reducing the threat without there being sufficient scientific evidence to conduct a justifiable risk–benefit analysis. Because this is a controversial practice, psychologists might contribute to the debate by investigating the extent to which everyday safety decisions involve precautionary

**TABLE VIII**  
**General Strategies for Behavioral Change in Managing Commons Resource Dilemmas**

<ul style="list-style-type: none"> <li>• Provision of physical alternatives, (re)arrangements</li> <li>• Regulation and enforcement</li> <li>• Financial-economic incentives</li> <li>• Information, education, communication</li> <li>• Social modelling and support</li> <li>• Structural, physical, and/or social change</li> <li>• Changing values and morality</li> </ul>
--

measures either preceding or falling outside well-informed rational decision making.

## 6. CONCLUSION

Both small- and large-scale levels of risk offer their own problems of analysis, judgment, decision, and control. Many risks are difficult to reliably quantify, due to lack of experience and information and because of the dynamic nature of the relevant activity. Thus, expert judgment, perhaps that of the subject assessing the risk, becomes an important factor, and total process control becomes more important than single gambling types of acceptance decision. Therefore, psychologists may define risk as a lack of perceived control (or of coping potential) in view of significant possible negative consequences.

In many cases, an activity or situation is considered to be risky for a variety of reasons (see Table IV). Multidimensionality implies that these activities or situations could be made (to appear) more safe in a corresponding variety of ways. Thus, classical definitions of risk as a function of probability and size of loss have limited practical significance. Multidimensionality of risk also implies that one should be very careful in making risk comparisons across different domains: comparisons variable(s) should be similar and well defined.

As regards large-scale social and environmental risks, sophisticated strategies are possible to limit such risks in view of the numerous social and economic benefits involved. The commons dilemma is a useful model that allows comprehensive problem diagnosis, policy decision making, and well-tuned practical interventions. For highly ambiguous environmental threats, the precautionary principle may be invoked, but this needs psychological investigation to support its practical applicability.

Finally, for future studies on risk, psychologists should familiarize themselves more deeply with specific domains such as industry, transportation, health care, and environmental protection to make their work less abstract and give it more practicality. This would also enable psychologists to work together with experts from other disciplines, whose concepts and tools may be instructive but will likely also need psychological enrichment.

### See Also the Following Articles

Environmental Design and Planning, Public Participation in  
 ■ Safety and Risks: Errors and Accidents in Different Occupations ■ Trust ■ Work Safety

### Further Reading

- Gardner, G. T., & Stern, P. C. (1996). *Environmental problems and human behavior*. Boston, MA: Allyn and Bacon.
- Gärling, T., Biel, A., & Gustafson, M. (2002). The new environmental psychology: the human interdependence paradigm. In R. B. Bechtel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 85–94). New York: Wiley.
- Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist*, 39, 341–350.
- Klinke, A., & Renn, O. (2002). A new approach to risk evaluation and management: risk-based, precaution-based and discourse-based. *Risk Analysis*, 22, 1071–1094.
- Lazarus, R., & Folkman, S. (1984). *Stress, appraisal and coping*. New York: Springer.
- Liebrand, W. B. G., & Messick, D. M. (Eds.) (1996). *Frontiers in social dilemmas research*. New York: Springer Verlag.
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127, 267–286.
- Rippetoe, P., & Rogers, R. (1987). Effects of components of protection-motivation theory on adaptive and maladaptive coping with a health threat. *Journal of Personality and Social Psychology*, 52, 596–604.
- Slovic, P. (1997). Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. In M. H. Bazerman, D. M. Messick, A. E. TenBrunsel, & K. A. Wade-Benzioni (Eds.), *Environment, ethics and behavior. The psychology of environmental valuation and degradation*. San Francisco, CA: The New Lexington Press.
- Vlek, C., & Keren, G. B. (1992). Behavioral decision theory and environmental risk management: assessment and resolution of four 'survival' dilemmas. *Acta Psychologica*, 80, 249–278.
- Wilde, G. J. S. (1994). *Target risk*. Toronto, Canada: PDE Publications.
- Yates, J. F. (1992). *Risk-taking behavior*. Chichester/New York: Wiley.



# Ethics and Social Responsibility

Pieter J. D. Drenth

Vrije University, Amsterdam, The Netherlands

1. Science: Value Free or Value Bound?
2. Ethical and Social Responsibility
3. “No-Go” or “Slow-Go” Decisions?  
Further Reading

## GLOSSARY

**“no-go” decisions** A term meaning that the research in question is wholly unacceptable with respect to specific research questions on the ground of social and ethical objections.

**“slow-go” decisions** A term that may include cases in which there is insufficient knowledge to properly appreciate and control the social or moral consequences of the research with respect to specific research questions on the ground of social and ethical objections; in such instances, results should be temporarily suspended until the ethical implications have been subjected to public discussion and consensus or at least a reasonable majority agreement has been reached.

**Wissenschaft als Inhalt** A morally neutral term referring to science as a product (i.e., the accumulated body of knowledge) solely subject to methodological–scientific norms.

**Wissenschaft als Tätigkeit** A term referring to science as the process of knowledge accumulation subject to ethical and political norms.

For quite some time, academic psychologists have incurred risks of accusations of “superbia” by developing an “ivory tower” attitude in which any accountability for human and social effects of their

research is repudiated: “Science is about how things are, not about they should be.” During the past few decades, however, the question of science and values has become a major subject of discussion again—certainly within psychology. Attacks on the autonomy and sovereignty of scientific psychology have come from various sources. First, there was the antiestablishment movement of the 1970s in which the political–scientific reflections of authors such as Marcuse, Adorno, Habermas, and Holzkamp became quite popular and were willingly embraced by student activists and (mostly junior) staff. Their criticism certainly contributed to the dismantling of the misconception that freedom of science would imply a negation of societal responsibility. A second attack resulted from the economic recession in many Western countries during the 1980s that led to a severe reduction of public and private funding for research. Acknowledged (and published) research output became an important condition for the survival of research personnel. This challenged another misconception of scientific freedom, namely, the license to freely think, meditate, and explore without caring much about output or quality. A third attack on the free and autonomous academic psychology came from the cry for “utilization” of scientific research at the end of the 1980s and during the 1990s that was heard in both governmental and industrial circles: Scientific goals should be seen as subordinate to those of economic and technological development, and criteria such as utility, applicability, and economic relevance should be given priority over pure scientific standards. Industrial, societal, and/

or technological relevance became an important consideration for funding research. It is ironic how much the neo-Marxist principles of the 1970s coincide with the industrial utility touchstones of the 1990s. Fourth, during recent years, the question of the autonomous and value-free character of science versus the relevant and value-bound character of science has received ample attention. A more detailed discussion of this issue is presented in this article.

### 1. SCIENCE: VALUE FREE OR VALUE BOUND?

The question phrased in the heading of this section has become an important issue during the past few decades, although opinions still diverge. On the one hand, many defend that scientific knowledge is value free and has no moral connotations. Science tells us how the world is, whether we like the story or not. Basic research is driven by scientific curiosity and not by the hope that it will be put to practical use. Ethical and moral issues can arise only when science is applied and is expected to produce usable objects. But then it has become technology and is not science anymore. Technological objects or processes can be used for better or for worse. Science produces insights, ideas, and pieces of knowledge that are in themselves neutral and can be corrupted only if they are mixed with political, social, economic, or other nonscientific aims.

On the other hand, there is a different view that does not accept the premise that science should be concerned only with producing reliable knowledge and, consequently, should be value free. The following arguments for the value-bound character of science can be brought forward.

- It is a basic obligation of all scientists and scholars to reflect on the paradigmatic presumptions and socio-historical entrenchment of their scientific activities. This reflection is, in itself, a metascientific and value-embedded phenomenon. Scientists must realize that their conceptualizations and models are always abstractions of reality and that only an approximation—or “reconstruction”—of reality can be achieved.

- Recently, the distinction between basic science and applied science has become less clear-cut. There is a good deal of overlap between the two spheres, and it is increasingly difficult to identify parts of science that do not affect, or are not affected by, technology. In this light, it will become clear that reserving “value-free

autonomy” for science and “value-bound heteronomy” for applied science and technology is no longer tenable.

- Scientists, and especially behavioral scientists, deal with a social, political, or psychological reality that is continuously affected and changed by the very scientific findings they produce. Health, safety, communication, privacy, mobility, welfare, economic development, and many other valuable goals of humankind are influenced radically by the advances of modern science. At the same time, many ethical or sociopolitical problems result directly from these same advances. Scientists should be aware of this and should anticipate the changes they bring about and the problems these changes raise.

- Even if scientists refrain from actually making political or ethical choices and restrict themselves to presenting probabilities and risks coupled with certain options, their reasoning is not value free. Risks do involve values and normative choices that scientists must face. Questions such as the following arise. Risks for whom? How far does the “right to know” go? What is the balance between self-determination of the individual and the interests of larger groups or the society as a whole? At what level of certainty does the scientist have to warn, especially in case of irreversible developments?

- Scientists cannot avoid the metascientific question of whether or not it is worth knowing what they pursue. They must justify—not only for themselves but also publicly if taxpayers’ or sponsors’ money is involved—why scientific issues must be addressed. This justification implies, in essence, nonscientific choices and decisions.

These arguments warrant the following conclusion in scientific research. On the one hand, objectivity must be maintained against the pressure from ideological movements, industrial lobbies, governmental strain, and/or political pressure groups. On the other hand, it has become ever more difficult to separate the functions of producing knowledge and making value-bound choices in extending research findings to the public or to the society at large. Here it may be helpful to distinguish between science as a product of knowledge (i.e., the accumulated body of knowledge) and science as a process of knowledge accumulation. The former, which was referred to as *Wissenschaft als Inhalt* by the Swiss philosopher Bochenski at the 1990 Engelberg Forum, is morally neutral and subject solely to methodological–scientific norms. The latter, which Bochenski described as *Wissenschaft als Tätigkeit*, is

subject to ethical and political norms that stipulate the choice of hypotheses to be researched, the manner in which data are gathered and experiments are conducted, and the question of what is ultimately done with research data.

Thus, research is embedded in the context of values, interests, and political objectives. Rather than denying this or retreating to the safety of the ivory tower, scientists would do well to realize this and take the appropriate responsibility seriously.

## 2. ETHICAL AND SOCIAL RESPONSIBILITY

This article does not discuss the ethical and social responsibilities of the professional psychologist in everyday practice. It is clear that in the application of psychology, be it in a clinical, pedagogical or social/organizational setting, the professional worker comes across many moral problems and dilemmas. Most national psychological associations have dealt with these issues, and many have developed codes of ethics for the professional application of psychology in various settings. A discussion of these codes falls outside the scope of this article.

But it has become clear that if one looks at the ethical and social responsibility of the psychologist as a scientist, his or her liabilities do not stop at the gate of the laboratory. This responsibility reaches further than the limits of purely scientific norms. It concerns the choice of the subject to be researched (or not), the proper conduct of experiments and surveys with due respect for human participants and/or animal subjects, and accountability for the application and use of the scientific results. More specifically, the following precepts are obtained. One can defend that the psychologist as a scientist should do the following:

- Promote and take an active part in the national and international discourse on ethical and social conditions and consequences of scientific research in psychology and its implementations.
- Be on the alert for selective and biased interpretation of the research findings by opponents or advocates of this research and deprecate such bias whenever discovered.
- Be observant of possible improper use or abuse of the research findings, warn against such possible misuse, and take corrective actions wherever possible.

- Provide careful, complete, and understandable information about the research results, not only in scientific journals but also in professional magazines and popular media (if possible). The latter should be considered important particularly if the research tends to have a substantial social, political, or economic impact.

- Promote the development and the acceptance of a set of rules and regulations for ethical conduct for professional practice in psychology, and train and educate students and young professionals to obey these rules and to develop conscientious attitudes and integrity in applying research results in their future professional work.

- Advocate and develop conditions that inhibit improper applications of scientific products (e.g., tests, instruments, treatments, intervention methods), including restrictive sales, extension of copyrights, and penal sanctions.

In other words, the responsibility of the psychologist as a scientist would have to be translated into very practical efforts and activities and specific rules of conduct.

## 3. “NO-GO” OR “SLOW-GO” DECISIONS?

An interesting question is whether scientific curiosity should be uninhibited and scientific research should be unrestrained, or whether social or ethical constraints should be imposed on science in some cases—in other words, whether there could be a need for “no-go” or “slow-go” decisions with respect to specific research questions on the ground of social and ethical objections. No-go means that the research in question is wholly unacceptable. At this time, there are cases in which such a constraint would find wide support, including attempts to make human–ape hybrids, applications of germline genetic manipulation to enhance intelligence, and human cloning for procreation. Slow-go would apply in instances where results should be temporarily suspended until the ethical implications have been subjected to public discussion and consensus or at least a reasonable majority agreement has been reached. This area might include cases in which there currently is insufficient knowledge to properly appreciate, and thus control, the social or moral consequences of one’s research.

It should be noted first that any discussion on constraints to be imposed on research is fraught with

danger. Throughout history, there have been too many examples of scientists who were repressed and whose research was stifled because the research outcomes were not consonant with the ruling ideologists, did not curry favor with powerful authorities, or were opposed to the interests of influential movements and action groups. Even if some of these movements have reputable objectives, such as peace and détente, equal rights for women, environmental protection, and non-discrimination, infringement of the right to investigate on the basis of political unacceptability of possible outcomes to external groups is highly precarious.

It would be equally wrong to refrain from doing research in a given area because it might possibly be abused or applied irresponsibly. This would virtually mean the end of all scientific research given that no scientific result is secured against willful abuse.

Therefore, it is not easy to identify incontestable ethical constraints to scientific research. Such constraints should refer to basic and peremptory values that would be imperative for scientists and responsible citizens alike. With respect to this issue, the European Federation of National Academies of Sciences and Humanities has discussed and accepted the following principles that could also be endorsed by researchers in psychology:

1. Research is not justifiable if unacceptable damage is inflicted on the object of research or its wider environment (e.g., social environment, society, nature) before, during, or after the experiment or the gathering of data. This applies to all objects of research, whether they be people, animals, nature, or culture.

2. Research is not justifiable in cases where the nature or the consequences of the research are in conflict

with basic human values. These values include, in any case, (a) respect for human dignity, which guarantees the autonomy and freedom of choice of all individuals, informed consent prior to participation in research, and the rejection of every intent to commercialize the human body; (b) solidarity with humankind, which guarantees regard and acceptance of fellow humans on the basis of equality; and (c) solidarity with posterity, which embodies the broader responsibility for sustained development of a planet that is to be left to future generations.

### See Also the Following Articles

Industrial/Organizational Psychology, Overview ■ Occupational Psychology, Overview

### Further Reading

- Berthoud, G., & Sitter-Liver, B. (1996). *The responsible scholar*. Canton, MA: Watson Publishing International.
- Drenth, P. J. D., Fenstad, J. E., & Schiereck, J. D. (1999). *Science and scientists between freedom and responsibility*. Luxembourg: Office for Official Publications of the European Community.
- Heller, F. A. (1986). *The use and abuse of social science*. London: Sage.
- Kimmel, A. J. (1988). *Ethics and values in applied social research*. Newbury Park, CA: Sage.
- Nagel, T. (1979). *Mortal questions*. Cambridge, UK: Cambridge University Press.
- Shea, W. R., & Sitter-Liver, B. (1989). *Scientists and their responsibility*. Canton, MA: Watson Publishing International.



# Exceptional Students

Jessica Singer-Dudek

St. John's University, Jamaica, New York, USA

1. Who Are Exceptional Students?
2. The Individuals with Disabilities Education Act
3. Labeling and Classification of Exceptional Students
4. Assessment of Exceptional Students
5. Characteristics of Exceptional Students
6. Implications for Professionals and Families  
Further Reading

## GLOSSARY

**applied behavior analysis** A scientific method for educating all students; its basic premise is that all behavior is controlled by its consequences, and so all behavior can be changed.

**continuum of services** The range of special education placement options, ranging from the least restrictive environment to the most restrictive environment.

**functional behavior assessment** A critical method for determining the function or purpose of behavior; it involves determining not only the consequences of behavior but also the antecedents of or precursors to behavior.

**high-incidence disabilities** Specific learning disabilities, speech and language impairments, mental retardation, and emotional disturbance.

**inclusion** When an exceptional student is a member of a regular class and receives instruction in the regular education setting for the majority of the school day; it may also refer to the philosophy of educating all students in a nonsegregated environment.

**individualized education plan (IEP)** A plan whereby a contract is devised among the school district, educational placement

(i.e., the school), and a parent(s) outlining a student's educational plan for the school year; it includes the type of educational placement, services to be provided, progress from the past school year, recent test scores and evaluation results, and goals and objectives for the school year.

**low-incidence disabilities** Nine disabilities identified under the Individuals with Disabilities Education Act (IDEA): other health impairments, multiple disabilities, autism, orthopedic impairments, hearing impairments, developmental delays, visual impairments, traumatic brain injury, and deafness/blindness.

**mainstreaming** When an exceptional student is a member of a special class but attends a regular class for certain periods of the day; a student with special needs can be mainstreamed for academic or nonacademic subjects.

**reinforcement** The principle of behavior stating that a desirable consequence delivered immediately after a behavior will serve to increase the likelihood that the behavior will occur again in the future.

**related services** Supplemental services that may include speech and language therapy, physical therapy, occupational therapy, nursing services, counseling, hearing services, and/or visual services.

**self-contained classroom** A special class that consists solely of students with disabilities; it has a smaller teacher-to-student ratio and may include teaching assistants or paraprofessionals.

**self-fulfilling prophecy** When low expectations are set and students only learn to meet these lower standards.

**self-injurious behavior** Repetitive or nonrepetitive behavior that results in bodily harm to the individual, either immediately or over time (e.g., scratching, hair pulling, face slapping, scab picking, head banging).

*stereotypy* Repetitive behaviors that appear to have no function other than the sensory enjoyment of the behaviors themselves; stereotypic behaviors are often emitted by students who are diagnosed with autism or mental retardation.

Exceptional students comprise nearly 13% of all American school-aged children. Approximately 10% of students in the United States are diagnosed with disabilities, whereas 2 to 3% of the population is characterized as gifted or talented. The education of students with disabilities in the United States is governed for the most part by Public Law 94-142, also known as the Individuals with Disabilities Education Act (IDEA). Although exceptional students can be separated into 13 disability categories under IDEA, each student is unique. The learning, social, and behavioral characteristics of exceptional students, including gifted and talented students, differ for each individual student. The individualized educational needs of exceptional students must often be determined using alternative types of assessments. According to these assessments, the most appropriate placement, services, and instructional modifications are put into place for each student. Exceptional students can be educated in a range of educational settings, including the regular classroom. Under IDEA, students with disabilities receive an education according to an individualized education plan (IEP) designed around their unique needs and goals. Special educational plans may also be arranged for gifted or talented students. Because of the careful identification of goals and individualization of instruction, collaboration between professionals and families is critical for the successful education of exceptional students.

## 1. WHO ARE EXCEPTIONAL STUDENTS?

Exceptional students comprise the population of students with disabilities who require specialized educational services as well as students who are characterized as gifted or talented. Approximately 10% of students in the United States are diagnosed with disabilities, whereas approximately 2 to 3% of the population is considered to be gifted or talented.

The term “exceptional” has often been used to describe unusual, unique, or outstanding qualities of people or objects. Consider the following phrases: “His artwork is exceptional” and “She is exceptionally bright.”

In this case, the term “exceptional” refers to students who learn and develop differently from most others or students who have exceptional learning styles, exceptional talents, or exceptional behaviors. Exceptional students are those who fall outside of the normal range of development. Some disabilities are identified at birth or soon after birth, whereas others go undetected until the students enter school, where learning problems first become apparent. In the case of gifted or talented students, their abilities may also go undetected until they reach school. Some exceptional students have learning problems, some have behavioral problems, some have communication difficulties, and some have social deficits. Many students with exceptionalities have two or more of these problems, although they may also excel in certain areas. Students who are gifted or talented may also display behavioral or social difficulties. For exceptional students, the most important distinctions are made in the area of educational needs.

## 2. THE INDIVIDUALS WITH DISABILITIES EDUCATION ACT

The education of exceptional students in the United States is governed by a federal law known as the Individuals with Disabilities Education Act (IDEA), which was first enacted in 1975 and was then known as the Education of All Handicapped Children Act. Prior to 1975, U.S. schools were not mandated to educate students with disabilities. Very often, students with disabilities were deemed to be uneducable and were barred from entering schools. Now, their educational rights are protected, and dictated, by IDEA. Students are protected under IDEA from 3 years of age until the end of the school years in which they reach their 21st birthdays unless the students graduate earlier. The education of students who are exceptionally gifted or talented is governed by individual states. However, not all states have legislation pertaining to the education of gifted or talented students.

According to the U.S. Department of Education, a total of 5,762,935 students with disabilities ages 6 to 21 years were served under IDEA during the 2000–2001 school year. IDEA recognizes 13 disability categories. The number of students in each disability category served under IDEA during the 2000–2001 school year is listed in [Table I](#). Each category of exceptionality has its own set of criteria for diagnosis. Students must meet all criteria to be diagnosed with a particular disability.

**TABLE I**  
**Number of Children Ages 6 to 21 Years Served under IDEA During the 2000–2001 School Year by Disability Category**

<i>Disability category</i>	<i>Number of students served</i>
Specific learning disabilities	2,879,445
Speech and language impairments	1,092,105
Mental retardation	611,878
Emotional disturbance	472,932
Other health impairments	291,474
Multiple disabilities	121,954
Autism	78,717
Orthopedic impairments	73,011
Hearing impairments	70,662
Developmental delay (ages 3–9 years)	28,683
Visual impairments	25,927
Traumatic brain injury	14,829
Deafness/Blindness	1,318
All disabilities	5,762,935

*Source.* From U.S. Department of Education, Office of Special Programs, Data Analysis System. (2002). *To assure the free appropriate public education of all children with disabilities: Twenty-fourth annual report to Congress on the implementation of the Individuals with Disabilities Education Act.* Washington, DC: U.S. Department of Education.

Sometimes, students are impaired in more than one area, and these students may receive a diagnosis of multiply handicapped. The first four disability categories—specific learning disabilities, speech and language impairments, mental retardation, and emotional disturbance—are considered to be high-incidence disabilities because they affect the majority of students with disabilities. Specific learning disabilities affect nearly half of all students with disabilities. The other disability categories are considered to be low-incidence disabilities.

## 2.1. The Six Principles of IDEA

Exceptional students often require specialized education to meet their unique needs or to accommodate and build on their special talents. The majority of students with disabilities and students who are gifted or talented require special education. IDEA ensures a free and appropriate public education (FAPE) to all students with disabilities. Under IDEA, there are six important principles that dictate how students with disabilities are diagnosed, evaluated, placed in educational settings, and educated. These principles are discussed in the following subsections.

### 2.1.1. Zero Reject

The zero-reject principle ensures that no student can be denied an education due to the presence or severity of a disability. A student with a severe learning difficulty or severe behavior problem may not be denied services or expelled from school due to his or her disabilities. Under this principle, a student who is terminally ill or has a contagious disease may not be denied a FAPE either.

### 2.1.2. Nondiscriminatory Evaluation

The second principle of IDEA ensures nondiscriminatory evaluation of any student who has, or is suspected of having, a disability. Several types of assessment are conducted by an interdisciplinary team of nonbiased members. No evaluations may be conducted without parental consent. All decisions pertaining to a disability diagnosis and to educational evaluation are made by a committee that includes the parent(s) or guardian(s) of the child. The committee seeks to answer three important questions: Does the student have a disability? If so, does the student require special education services? If so, what educational placement and which services will best serve the student's needs?

### 2.1.3. Appropriate and Beneficial Education

The third principle of IDEA ensures an appropriate and beneficial education. This means that the student must be placed in an educational setting that best suits his or her needs. The student must be making educational progress for the placement to be deemed beneficial. An individualized education plan (IEP) is devised for each student, outlining the progress, placement, services, and goals to be addressed for the student. Appropriate placement means that the student is placed in the least restrictive environment (LRE).

### 2.1.4. Least Restrictive Environment

IDEA requires that the student be educated with his or her nondisabled peers to the maximum extent possible. The LRE is the educational setting that is closest to the regular educational setting that can still meet the student's individual needs. Only if a student's needs are too great to be met in the regular education environment should a more restrictive placement be considered. The range of educational placement options is known as the continuum of services.

### 2.1.5. Procedural Due Process

The fifth principle of IDEA ensures procedural due process. This means that there are safeguards in place in case the parents and school district disagree about what is best for the student. Impartial officials preside over due process hearings and make decisions according to what is best for the student. The principle of procedural due process also dictates how parents are informed of all decisions regarding their child, including the child's educational progress.

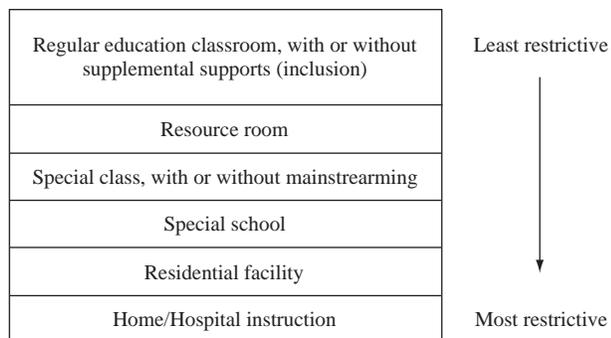
### 2.1.6. Parent and Student Participation

Parent and student participation is encouraged and mandated under the sixth principle of IDEA. This means that no decisions shall be made regarding the education of an exceptional student without parental consent and input. Student input is also encouraged whenever the student is capable of contributing. The student should have the opportunity to make decisions about his or her own future.

## 2.2. The Continuum of Services

Under IDEA, students with disabilities must be educated in the LRE, which is the educational environment that is closest to the general education classroom and that still meets the students' needs. There is a range of educational placements for students with disabilities, known as the continuum of services. Figure 1 illustrates the continuum of services.

At the top of the spectrum is the general education classroom. The student may receive supplemental services, known as related services, while in this placement (e.g., speech/language therapy, physical therapy), but



**FIGURE 1** The continuum of services from the least restrictive environment to the most restrictive environment.

the student is a member of the regular class. When a student with a disability is a member of a regular class and spends the majority of the school day in that classroom, this is known as inclusion. An inclusion setting is an appropriate and beneficial placement only if the student is making educational progress. The social benefits of the inclusion setting are also taken into account and weigh heavily when determining whether an inclusion setting is the best placement option.

The next option is a resource room. Here, the student is a member of the regular class but spends up to two periods each day in a special class taught by a special education teacher. Usually, particular goals are worked on in the resource room (e.g., math, reading, study skills). The next placement option on the continuum of services is a special class or self-contained setting. Here, the student is placed in a class consisting solely of students with special needs. Special classes have smaller student-to-teacher ratios and usually include one or two teaching assistants or paraprofessionals. There are a variety of student-to-teacher-to-teaching assistant ratio options available. Sometimes, students who are placed in self-contained classes actually excel in one or more subject areas. For example, a student may require specialized instruction for reading and all content-area subjects but be on grade level for math. This child may attend a regular math class during the school day. This is known as mainstreaming. In keeping with IDEA, the student is spending as much time as possible with nondisabled peers.

The next placement option is a special school. Here, the student attends a school composed solely of students with disabilities. This restrictive setting ensures that programs and accommodations are available consistently. Beyond a special day school is the next placement option on the continuum of services, that is, a residential facility. The student attending a residential facility requires 24-hour continuation of instructional programs and care. The final placement on the continuum of services is home or hospital instruction. Here, instruction must be provided by a special education teacher for at least some portion of the school day.

## 3. LABELING AND CLASSIFICATION OF EXCEPTIONAL STUDENTS

There are several advantages and disadvantages to labeling students with disabilities. The first advantage is that it allows these students to be served under IDEA.

Another advantage to labeling students with disabilities is that it ensures that these students receive the proper types of assessment and the proper types of educational placements. A third advantage to labeling students with disabilities is that it ensures that these students are provided protection under the law and that they continue to receive protection under the other laws that afford rights to individuals with disabilities (e.g., the Americans with Disabilities Act) after they leave school. Lastly, some students and their families actually welcome disability labels because they provide an explanation for issues that these children and their families may have been struggling with for a long time.

There are several disadvantages to labeling students with disabilities. First, labeling a child may cause segregation of that child from his or her peers without disability labels. Placing a label on a child leads others to treat that child according to the label rather than as an individual. Labels cause stigmas that can lead to decreased self-esteem. Labels can also lead to the self-fulfilling prophecy, which occurs when low expectations lead to lower performance. When lower expectations are placed on students, those students are never challenged and so only learn to perform to lower standards. On the other hand, setting high expectations and challenging students can lead to greater achievement.

Person-centered language refers to the way in which students with disabilities are described. Placing the disability label first leads to identification of the disability first and the individual second. Instead, it is correct to refer to the individual with a disability first and the disability label second. For example, one should say "student with a learning disability" instead of "learning-disabled student." Focusing on the disability label causes others to view the student only in terms of his or her deficits or areas of weakness. If people focus on the individual first, they are more likely to see ability first and disability second.

#### **4. ASSESSMENT OF EXCEPTIONAL STUDENTS**

Under IDEA, students are entitled to nondiscriminatory evaluation. This means that they are evaluated using a variety of measures, including both formal and informal assessments. Students with disabilities should participate in the same assessments as their nondisabled peers to the maximum extent possible. This means that many students with special needs are

required to participate in statewide assessments such as standardized achievement tests. Exceptional students may be easily identified by their scores on standardized tests such as intelligence quotient (IQ) tests. The bell-shaped curve refers to the distribution of scores from such tests. The majority of students fall in the middle of the distribution or the average range. Very few students fall at the extreme low or high ends of the spectrum or outside of the normal range. Students who fall at the extreme high end of the spectrum are typically students who are gifted or talented, whereas students who fall at the extreme low end of the scale are typically students with cognitive delays.

Not surprisingly, students with disabilities often perform poorly on standardized tests. There are several alternative assessment options available for students with disabilities. One of these is criterion-referenced assessment. Some commercial criterion-referenced assessments exist (e.g., Sparrow and colleagues' Vineland Adaptive Behavior Scales), whereas many are teacher-made. A criterion-referenced assessment consists of an inventory or hierarchy of repertoires across several domains, including social skills, communication skills, academic skills, leisure skills, self-help skills, and maladaptive behavior. These inventories provide a way of identifying which repertoires a student has mastered and which skills need to be taught. Identification of deficits leads to easier selection of goals. Criterion-referenced assessments are time-consuming because they may contain hundreds of items that need to be assessed with individual students, but the end result is a clear picture of the student's current skill levels.

Portfolio assessments include several different components, including samples of the student's work, progress reports and evaluations, and (sometimes) student self-evaluations. Portfolios demonstrate a student's progress over time in that samples of the student's work or tests are taken over the course of a school year or several years. For example, a writing sample taken in June may be substantially better than one taken in September. Videotapes or audiotapes may also be included in portfolio assessments.

Curriculum assessments involve teacher-made tests or assessments that come with commercially published curricula. Curricular assessments allow real measurement of student progress through items that have been taught directly over the course of the school year.

Other forms of assessment may be extremely valuable, although they are considered to be "informal" assessments. These may include reviews of records and past progress reports, observations of the student in the

classroom environment, anecdotal reports, and interviews with parents, teachers, and even the student himself or herself. Informal assessments can be valuable in obtaining information that could not be obtained from standardized or criterion-referenced tests.

## **5. CHARACTERISTICS OF EXCEPTIONAL STUDENTS**

Exceptional students include students across 13 disability categories and students who are gifted or talented. Although some disabilities are diagnosed based solely on behavioral characteristics, no two students are alike. However, it is possible to give some general overviews of student characteristics across three broad categories: learning, social, and behavior.

### **5.1. Learning Characteristics**

Some exceptional students excel in certain areas or learn better through certain modes of instruction. Some students, such as students with autism or mental retardation, excel in the areas of memory and fact retention but are severely deficient in the areas of language and social skills. Similarly, students with specific learning disabilities may experience difficulty in one area (e.g., verbal reasoning) but excel in another (e.g., math). Gifted and talented students and students with certain disabilities may possess one or more of what Gardner identified as multiple intelligences (e.g., bodily/kinesthetic, logical/mathematical, linguistic, spatial, musical, interpersonal, intrapersonal, naturalistic). Some exceptional students are visual learners, others learn kinesthetically, and still others learn through multiple modes of instruction. It is recommended that instruction always be presented in several ways and that students be exposed to various instructional methods, including hands-on activities, to enhance learning.

On the other hand, many exceptional students experience severe learning difficulties. Possible sources of learning problems may result from attention and memory difficulties, motivational difficulties, and/or organizational problems. Still other problems may result from a lack of generalization. Generalization can fail to occur across settings (e.g., from school to home), across individuals (e.g., from teachers to parents), and across stimuli (e.g., instructional materials presented in different formats). Teachers must plan for generalization and teach across settings, individuals, and stimuli to ensure

that responses learned under one condition are emitted under other conditions. There are several strategies available to remedy learning problems after the source of the problems has been identified.

Although some learning problems may result directly from a cognitive delay, as is the case with students with mental retardation and students with autism, other learning problems may occur due to other factors. Behavioral problems may impede learning but may also result from inappropriate instruction. If the task presented to a student is too difficult or too easy, if the pace of instruction is too fast or too slow, or if the size of the task is too large or too small, frustration or boredom may result. Frustration and boredom may lead to behavior problems. Gifted and talented students may also emit inappropriate behaviors due to boredom in their classes. Teachers must always ensure that these students have the prerequisite skills to perform the task under instruction and that they are properly motivated. It is important to figure out the learning styles of exceptional students to ensure that instruction will maximize their learning.

Because the learning styles and characteristics of exceptional students are so diverse, the IEP becomes a critical document for ensuring that students are receiving the specialized and individualized instruction they need. Certain instructional modifications that have been useful in improving learning will be outlined in the IEP to assist teachers in providing the best education possible to their exceptional students. Sometimes, however, exceptional students experience poor academic performance as a result of other factors such as social or behavioral problems that impede learning.

### **5.2. Social Characteristics**

Exceptional students may also experience difficulties with social interactions. They may appear awkward in social situations, have difficulty in making friends, and respond differently to social situations. Often, exceptional students may have the language skills to communicate but lack knowledge of the rules of conversation such as body language, proximity, turn taking, and staying on topic. The conversational skills of students with disabilities might seem egocentric or one-sided. These students might not listen to the responses of other students or might not play the role of both speaker and listener in conversations. Exceptional students often do not learn from appropriate modeling, do not pick up subtle cues, and do not interpret the feelings of

others appropriately. Gifted or talented students may also emit awkward social behavior, perhaps because their level of maturity is above that of their same-aged peers or because they have not had experiences similar to those of their peers. Difficulties with social skills may also lead to behavior problems.

### 5.3. Behavioral Characteristics

Behavior problems are probably most common among exceptional students. Behavior problems may range from assaultive behaviors (e.g., hitting, screaming, throwing objects) to severe withdrawal. Behavioral issues may also include inappropriate or bizarre behaviors (e.g., stereotyping, self-talk, self-injurious behaviors). The important thing for educators and parents to realize is that all behavior has a purpose or function. Sometimes, bad behavior is the result of an inability to communicate due to a lack of language skills. Other bad behaviors may result from attempts to gain attention or to escape or avoid an aversive task. Still other behaviors (e.g., stereotyping) are emitted because the behaviors themselves are reinforcing (e.g., sensory stimulation).

Attention-seeking and escape behaviors are often mistaken for willful disobedience or are seen as a result of the disability itself. Often, fictitious explanations of behavior result from placing the locus of control within the child (or within the parents) instead of looking to environmental variables to explain behavior. From a behavioral standpoint, all behavior is controlled by its consequences. This means that a student will continue to emit only behaviors that are reinforced or that are followed by desirable consequences. If the child screams in an attempt to gain attention and an adult responds to the child's screams, the child is more likely to scream to gain attention in the future. The consequences that educators and parents deliver, such as telling the screaming student to be quiet or asking the student what he or she wants, do not serve the purposes that were intended. Attention comes in many forms, including speaking to the child and making eye contact with the child. Escape behaviors are maintained by consequences that allow the child to get out of doing the task.

Functional behavior assessment involves determining what the function or purpose of the behavior is or what consequences are maintaining the behavior. Functional behavior assessments also seek to identify the antecedents or precursors for behavior as well as environmental and motivational variables. Behavior that continues to occur is behavior that is reinforced. The function of the behavior may be communication,

attention seeking, escape, or sensory reinforcement. Under IDEA, functional behavior assessments must be conducted for all aberrant behavior. Behavior plans then become part of the student's IEP.

## 6. IMPLICATIONS FOR PROFESSIONALS AND FAMILIES

It is very important for educators and parents to remember that undesirable behavior always has a function. If the function of the behavior can be determined, a solution to the problem behavior is usually suggested. Blame for the behavior should not be placed on the child or on the child's disability. Although the learning, social, and behavior difficulties of exceptional children might seem overwhelming, there is a great deal of research in the field, especially in applied behavior analysis, that demonstrates successful strategies and tactics for remedying bad behavior and teaching important skills. Sometimes, the most important skills for exceptional children to learn are how to manage their own behavior, how to seek help when necessary, and how to advocate for themselves. Fortunately, under IDEA, exceptional children have access to a wealth of pedagogical practices designed to prepare them to live as independently as possible.

### See Also the Following Articles

Educational Achievement and Culture ■ Gifted Students ■ Learning ■ Mathematics, Teaching of ■ Reading, Teaching of ■ Special Education ■ Teaching Effectiveness ■ Writing, Teaching of

### Further Reading

- Alberto, P. A., & Troutman, A. C. (2003). *Applied behavior analysis for teachers* (6th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Bos, C. S., & Vaughn, S. (2002). *Strategies for teaching students with learning and behavior problems* (5th ed.). Boston: Allyn & Bacon.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. New York: Macmillan.
- Friend, M., & Bursuck, W. D. (2002). *Including students with special needs* (3rd ed.). Boston: Allyn & Bacon.
- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.
- Heward, W. L. (2003). *Exceptional children: An introduction to special education* (7th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

- Hunt, N., & Marshall, K. (2002). *Exceptional children and youth*. Boston: Houghton Mifflin.
- Lewis, R. B., & Doorlag, D. H. (2003). *Teaching special students in the general education classroom* (6th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Lipsky, D., & Gartner, A. (Eds.). (1989). *Beyond separate education: Quality education for all*. Baltimore, MD: Brookes.
- Mastropieri, M. A., & Scruggs, T. E. (2004). *The inclusive classroom: Strategies for effective inclusion* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (2003). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.
- Turnbull, R., Turnbull, A., Shank, M., & Leal, S. J. (2004). *Exceptional lives: Special education in today's schools* (4th ed.). Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- U.S. Department of Education, and Office of Special Programs, Data Analysis System. (2002). *To assure the free appropriate public education of all children with disabilities: Twenty-fourth annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: U.S. Department of Education.
- Vaughn, S., Bos, C. S., & Schumm, J. S. (2003). *Teaching exceptional, diverse, and at-risk students in the general education classroom* (3rd ed.). Boston: Allyn & Bacon.



# Executive Development and Coaching

Rob Silzer

HR Assessment and Development Inc., New York, New York, USA

1. What Is the Historical Context?
  2. Who Is an Effective Executive?
  3. What Is Executive Development?
  4. What Are Typical Executive Development Approaches?
  5. What Is Executive Coaching?
  6. Which Performance Areas Are More Amenable to Development?
  7. Which Executives Are More Likely to Develop?
  8. What Are the Hallmarks of Development Success and Failure?
- Further Reading

*strategic thinking* Understanding and setting distinctive competitive goals and actions for an organization.

Executive development and coaching includes a variety of assessment tools and interventions that have been designed to help executives improve their knowledge, skills, abilities, and performance and to build the talent needed to meet strategic objectives and new business challenges in an organization.

## GLOSSARY

- coaching* Helping others to improve their knowledge, skills, abilities, and/or performance.
- development* Learning or strengthening knowledge, skills, abilities, or performance.
- executive* A general manager, head of a functional area, or corporate officer in an organization.
- general manager* A manager who has multiple organizational function responsibilities for running a business (e.g., production, sales and marketing, product development, human resources, finance).
- global executive* An executive who can work effectively and comfortably across cultural borders and who typically has lived and worked in at least two different countries.
- organizational culture* The shared values, beliefs, practices, rituals, and behavioral patterns in an organization.

## 1. WHAT IS THE HISTORICAL CONTEXT?

A greater focus on executive development and coaching emerged during the 1990s. During prior decades, executives were not given as much attention; instead, the emphasis was on selecting individuals who had management potential and on using basic management training to teach functional skills.

During the 1980s, the world of business changed significantly. There was a significant globalization of business markets and a destabilization of many companies and industries due to increased global competition, to a wave of business mergers and consolidations, and to an expanding acceptance of management talent moving across companies. The speed of business and decisions

greatly increased and forced people to work hard to keep up with the rapid changes. This led to a need for fungible managers and executives who can learn quickly and be effective in a range of business situations. Managers who can do this are highly valued, and they often leverage this demand for talent to aggressively negotiate for career opportunities that move them into executive roles so that they can help run a company.

During the 1990s, executives emerged as the key to effective corporations. They were seen as critical decision makers and were given a lot of personal visibility and credit in the press. In addition, many managers viewed being an executive and running their own businesses as viable career goals. However, it became evident that many managers and executives did not have the skills and abilities to effectively handle an executive role or to adequately adapt to major and rapid changes occurring in the business world. Because of the increased competition for the best talent, companies aggressively pursued initiatives to better train and develop their current internal managers and executives.

Initially, these executive development initiatives focused on sending current executives and managers with executive potential to graduate business schools for executive development classes and executive MBA programs. These efforts focused on providing individuals with additional business knowledge, particularly in corporate finance. Because of the cost and the limited number of participant openings in these external programs, companies began to conduct internal executive development classes and tailoring them to the specific needs and business situation of each company. Later, these programs expanded to also include some emphasis on the development of executive performance skills such as strategic thinking, media and press interviewing, and decision-making skills.

As corporations went through broad reorganizations and rapid changes, some executives had difficulty adapting to the changes. This led to the increasing turnover of executives and chief executive officers (CEOs) in companies, which in turn placed new performance demands on the remaining executives in a company. The existing executives were expected to retune their skills to fit the new expectations and strategic directions of the recently hired CEO. Human resource officers began to hire executive development experts, including retired business executives and organizational psychologists, to work in groups and individually with these executives in order to “fix their skills.” After some success in getting executives retrained and aligned with the new CEO, these development experts were then assigned to help well-

performing executives and high-potential managers who wanted to continue to build their executive skills. Slowly, the field of executive coaching emerged and evolved from fixing performance problems to further building the skills of strong performers.

This article reviews some of the current techniques used in executive development and coaching. When a company considers starting an executive development initiative, it typically must consider a few fundamental issues. How a company or development professional thinks about these issues has a direct bearing on the nature, approach, and success of the development effort.

## **2. WHO IS AN EFFECTIVE EXECUTIVE?**

There are many opinions on which knowledge, skills, and abilities (KSAs) an executive needs to perform effectively. Although some people use the term loosely, this article will reserve the term “executive” for general managers, functional area heads, and corporate officers in an organization. Many companies, human resource professionals, and industrial/organizational psychologists have identified their own lists of the necessary executive competencies. In 2002, Silzer and a range of respected experts reviewed executive skills and abilities from a variety of angles, while others, such as Fulkerson, McCall, and Hollenbeck, focused on the requirements to be an effective global executive. Silzer, in his chapter on selecting executives, outlined the challenge of selecting individuals who would be successful over the long term and across a variety of business challenges. He discussed the complexity of the selection process in terms of “executive fit” on four different levels:

- Person–position fit (matching the executive with the specific job requirements)
- Person–group fit (matching the executive with the existing executive team and direct reports)
- Person–organization fit (matching the executive with the organizational culture)
- Person–cultural fit (matching the executive with the country culture)

Some people argue that the executive job is so complicated, individually tailored, and ever changing that it is impossible or useless to identify one set of KSAs that apply across all executives and all situations. Unfortunately, this view leads down an empty path and provides little help when trying to structure

and implement executive selection and development approaches. Others would argue that while the executive role is very complex, there are some fundamental performance areas that do apply to a wide range of executives and business situations. A list of key executive KSAs that are relevant to most executive positions probably would include the following:

- Sound judgment and decision making
- Strategic thinking and strategy management
- Interpersonal influence
- Communication effectiveness
- Leadership confidence and impact
- Results orientation
- Business and financial acumen
- Customer focus
- Integrity
- Organizational insight and management

Executives are also typically expected to have some knowledge of general business practices, functional areas, organizations, financial management, and human resource management. Lately, knowledge of information technology, marketing, and customer service has become more important to executives, whereas knowledge of particular industries or products has become less important because executives now frequently switch companies and industries during their careers.

Of course, executives also need to handle their own, ongoing job responsibilities. These vary widely depending on the specific executive role, but there are some performance areas that apply to a broad range of executive positions, including the following:

- Accomplishing annual goals
- Managing the organization and financial budgets
- Building, motivating, and developing leadership talent
- Pursuing innovative ideas and distinctive competitive strategies
- Defining and supporting the organizational culture

### **3. WHAT IS EXECUTIVE DEVELOPMENT?**

In the past, there was a strong “selection bias” in human resource management. That is, there was a widespread belief that managers and executives had limited ability to grow or change, so executives, managers, and human resource professionals relied almost exclusively on selection to find the right person for a management or executive job. If someone was performing poorly, that

person was fired or moved aside so that another person could be selected to fill the position. Most psychological assessment tools and techniques in organizations were used to select the best person for the jobs.

However, during the 1980s, the field of training and development started to emerge as useful and relevant to management and executive effectiveness. For example, assessment centers that previously had been used exclusively for selection were being used to develop current and future incumbents to be more effective in these roles. Particularly in the fast-changing business environment of the 1980s and 1990s, companies had to figure out how to quickly redeploy and redevelop current employees for new roles and jobs because it often was infeasible to fire everyone and hire an entirely new workforce and executive team (although some companies did exactly that). As a result, there was a wave of interest in management, leadership, and executive development techniques.

The term “executive development” means learning or strengthening the KSAs needed to be effective in an executive position. There was significant interest by organizations in identifying and further developing those individuals who were seen as having significant potential to be effective in higher level executive positions. This led to the development of a range of new initiatives, including succession management systems, executive classes and programs based in graduate business schools, and the widespread use of management and executive competency models to specify the KSAs that a manager or an executive needs to be effective.

## **4. WHAT ARE TYPICAL EXECUTIVE DEVELOPMENT APPROACHES?**

Companies and development professionals have devoted a good deal of attention and creativity to designing management and executive development programs. In 1999, Conger and Benjamin outlined some development approaches used by different companies. In thinking about executive development, it is helpful to separate the approaches that focus on the assessment of executive KSAs and performance from those techniques that are used as development interventions.

### **4.1. Assessment Approaches**

Some executive development programs are designed so that all executives, regardless of their individual skills

and performance, go through the same program either as individuals or as groups. Another approach that gained support and acceptance during the 1980s is to first identify the development needs of individual executives and to provide a development experience that is appropriate for those specific needs. This leads to an expanding role for assessment in the process.

Assessment of an executive's development needs has typically used one of three general approaches:

- Performance reviews
- Multisource feedback
- Individual (psychological) assessment

#### **4.1.1. Performance Reviews**

The most common assessment approach is to use the annual performance review to identify an individual's performance/KSA strengths and development needs. Nearly every organization has designed an annual executive evaluation process that reviews an executive's behavior and performance in several areas that may include looking at results against annual goals, general executive performance, executive skills and abilities, and how well the executive demonstrated company values. Some review programs include specific behavioral standards of expected performance as review guidelines, whereas others are more open-ended and rely on the broad judgment of the reviewer.

#### **4.1.2. Multisource Feedback**

Another assessment technique that has gained widespread acceptance is a multisource feedback approach. This relies on collecting structured feedback on an executive's performance and KSAs from a range of people who have had an opportunity to interact with the executive and to observe the executive's behavior. This group might include immediate or higher level bosses, peers, direct and second-level reports, internal and external customers, members of the board of directors, outside community and financial contacts, vendors, and suppliers. The feedback is typically collected either through a structured written questionnaire (often referred to as a 360-degree feedback instrument), telephone interviews, or one-on-one meetings. A written questionnaire provides a more standardized approach that can cover quite a broad range of performance areas and KSAs. The interview and meeting approaches allow for a deeper discussion of key issues and the identification of numerous examples of behavior that illustrate the feedback.

#### **4.1.3. Individual Assessment**

A third approach that has been gaining significant support is an individual (psychological) assessment of the executive. This approach was initially used by companies to help select executives but is now also being used extensively for development. In 1998, Jeanneret and Silzer provided a good resource on how psychological assessment tools and approaches can be used effectively in organizations. When used for development, this approach is typically designed to assess current performance, KSAs, development needs, career direction, and potential for other roles. The approach usually includes a range of tools and techniques such as the following:

- Cognitive ability tests
- Personality and career interest inventories
- Structured behavioral interviews
- Behavioral exercises and simulations (e.g., role-play, "in" basket, and presentation exercises)
- Multisource feedback

Assessment techniques are usually used with individuals, although it is becoming more common for all executives to go through an assessment process at the same time. This group approach not only allows the company to understand the strengths and weaknesses of each individual executive but also provides a profile of the entire executive group. The CEO, and often the board of directors, can then see the shared strengths and shortcomings of the group. This approach can be very useful when the organization is going through a significant change such as a merger, downsizing, reorganization, or change in strategic direction. It can lead to a valuable reshuffling of executive talent across the organization to better match individuals to executive roles. In addition, the process can identify common performance weaknesses that can then lead to a group development initiative or even a modification in the company's strategic goals. For example, in some cases, organizations have used an executive group profile to determine the strategic direction of the companies by looking for opportunities that leverage the strengths of the existing executives and that minimize any reliance on areas that are group weaknesses.

## **4.2. Intervention Approaches**

Once the development needs or goals have been identified, then appropriate development intervention approaches can be identified that specifically address those needs. There are myriad intervention tools and

techniques that have been used to develop executives. They might be categorized into four general types:

- Learning through the experience of others as an individual.
- Learning through the experience of others with a group.
- Learning by personal experience as an individual.
- Learning by personal experience with a group.

#### ***4.2.1. Learning through the Experience of Others as an Individual***

Initially, this was probably the most common development approach, with individuals learning from their immediate bosses or the previous incumbents in their roles. When business started to go through widespread change, many bosses and previous incumbents no longer were the best sources for the executives on how to handle the new challenges. So, executives started turning to others with more experience for help. Some approaches to executive development in this category include the following:

- Working with an internal or external mentor
- Locating a tutor or an expert on a specific topic
- Establishing an internal or external network of resources and experts
- Taking a temporary assignment with a customer or vendor
- Initiating a personal reading program
- Conducting a study or writing a white paper on a critical business issue
- Visiting competitors, suppliers, and customers
- Attending conferences and seminars
- Teaching a business class

#### ***4.2.2. Learning through the Experience of Others with a Group***

This has been a frequently used approach, often relying on attending business classes or executive programs at graduate business schools. The high cost of these programs limited the number of executives who could attend and eventually led to a decrease in attendance but also to an increase in internal executive programs designed to meet the same goals. A group approach to development typically means that all executives go through the same program without regard for their individual development needs. However, in some cases, executives with similar needs have joined together to address those specific

common needs as a group. In 2002, Marcos described some innovative group approaches. Some examples of group development include the following:

- Attending a business school course or an executive program
- Participating in an internal leadership or executive training program
- Starting a reading discussion group
- Participating in a peer group shared learning program
- Initiating an internal speaker presentation series

#### ***4.2.3. Learning by Personal Experience as an Individual***

As the executive development field moves to assessing and identifying individual development needs, there is a growing focus on activities in this category. Many development action plans are being written that focus on the specific development needs of the executive. These plans initially and primarily consisted of taking courses, but the field soon adopted the now widely accepted idea that the best development occurs on the job in a “learn-by-doing” approach. This shifted quite a bit of attention to experiential learning at the executive level. One very visible technique is the “action learning” approach described by Dotlich and Noel in 1998. Other examples include the following:

- Starting an on-the-job skills building program
- Initiating an action learning assignment and project
- Rotating through different core businesses or functional areas in the company
- Becoming an executive assistant to an officer
- Leading a strategy review or a negotiating process
- Taking an overseas or global assignment
- Managing a start-up, turnaround, or high-growth business
- Managing a large, highly complex, or matrixed business
- Taking responsibility for a liaison role such as government lobbying or community affairs

#### ***4.2.4. Learning by Personal Experience with a Group***

Some of the preceding approaches evolved into group experiences. In some companies, “action learning” means group projects. A selected team of executives or high-potential managers is assigned to work on a

special project that addresses a major organizational issue or problem. The team learns a great deal about the issue and the organization, while the company gains some sound recommendations on how to address the issue. Examples of group learning approaches include the following:

- Joining an action learning project team
- Volunteering for a corporate task force
- Being part of a new business start-up team
- Pursuing an internal or external business partnership
- Joining an industry-wide committee
- Serving on a business review team

## **5. WHAT IS EXECUTIVE COACHING?**

During the past 10 years, executive coaching has become a widespread executive development tool. It can be defined as a one-on-one helping relationship between an executive and a coach that helps to improve the executive's performance and combines learning through the experience of others with learning from personal experience. Often, this process focuses on the executive's KSAs and behavior. It is different from mentoring in that coaching seeks to change the executive's on-the-job behavior and performance and not just to provide advice. Hollenbeck, in his enlightening 2002 chapter on executive coaching, distinguished it from psychotherapy by noting that the goal of coaching is "better job performance," whereas the goal of psychotherapy is "a more fully functioning person."

In some cases, every executive in a company is assigned a personal external coach. A number of companies later backed off of this sweeping approach due to the high cost. Executive coaching turned into something of a fad and attracted the interest of a wide range of coaching service providers. Because there are no credential standards, anyone can assume the title of executive coach. This includes human resource consultants, trainers, psychologists, retired business executives, health care providers, and MBA graduates. As a result, the field of executive coaching and executive development has been swamped with people who do not have the necessary coaching skills or business knowledge. Corporations are often fairly indiscriminate in selecting executive coaches. Many coaches are probably ineffective, but corporations do not spend much time in evaluating the effectiveness of coaching interventions. Under significant

cost reduction pressures, companies are now starting to better manage these assignments. This has resulted in a few emerging trends in executive coaching, including the following:

- A more discriminating use of external coaches
- Written standards for the selection and evaluation of coaches
- The training of current internal executives as peer coaches

Good coaches need to have knowledge of business and executive performance issues as well as skills in communicating, interpersonal influencing, and changing behavior. Many coaches fall far short of this. The coaching process typically lasts from 6 to 18 months, although many companies have started to put a 1-year time limit on these interventions. The process generally consists of five phases:

- Initial agreement and contract
- Assessment
- Feedback and planning
- Coaching intervention
- Follow-up

Moving through these phases depends on success at each step. Coaches need to be able to demonstrate progress in changing an executive's behavior. Too many interventions are just pleasant bimonthly conversations with little focus or impact. Coaches must also be trustworthy and be able to carefully manage the confidential information shared by the executive and the executive's immediate bosses. Many coaching interventions fail because of a poor coaching contract at the beginning of the process, a lack of confidentiality, or a lack of change in the executive's behavior.

Recent interest in executive coaching has focused on a few key issues:

- How can the process be made more cost-effective?
- How can the effectiveness of coaching interventions be rigorously evaluated?
- How can individual coaches be objectively selected, assigned, and evaluated?

## **6. WHICH PERFORMANCE AREAS ARE MORE AMENABLE TO DEVELOPMENT?**

Over the years, coaching experts have gained a good deal of insight into how to effectively change the

behavior and performance of executives. There is some agreement that interpersonal and communication skills are the most common areas that coaches focus on for development. They also are the areas where good coaches can have significant impact. Other areas that seem fairly amenable to development include personal organization, planning, administrative skills, business knowledge, presentation skills, financial knowledge, and project management.

On the other hand, there are a number of areas that are difficult to develop either due to the complexity of the area or because it requires specific preliminary abilities. These areas include complex decision making, innovative thinking, strategic planning, personal and career motivation, leadership versatility and influence, and personal integrity.

The field still needs to better define which development approaches are best suited for which performance areas. It is now known that interpersonal role-playing and behavioral exercises do seem to be useful approaches in changing an executive's interpersonal skills. However, there is little research matching development interventions with other performance areas. In general, an executive needs to go through four stages for changing behavior: understanding the behavioral issue, learning new effective behaviors, practicing the new behaviors, and getting feedback from others on those behavioral attempts.

## **7. WHICH EXECUTIVES ARE MORE LIKELY TO DEVELOP?**

There also has been growing agreement that some executives are more likely to develop than others. Although there is not total agreement on what separates these two groups, there are a few markers that seem to distinguish the developable executives. These executives generally have strong cognitive skills in order to adequately understand the development that is needed. They also need to be mentally flexible and behaviorally adaptive to try new behaviors, and they should have a clear motivation and willingness to change. That is, they should see the need to change their behavior. In addition, it helps if they have a learning orientation, that is, an intellectual curiosity and interest in learning new things. Executives who do not develop generally are inflexible, unwilling to change, or unable to learn new behaviors.

## **8. WHAT ARE THE HALLMARKS OF DEVELOPMENT SUCCESS AND FAILURE?**

Companies typically spend significant financial resources on executive development each year. There is a large group of internal and external development specialists and consultants who have been very successful in providing executive development services. Although many companies have recently worked to reduce development costs, many expensive development approaches, such as business school programs and executive coaching, are still widely used.

Executive development efforts often fail to have any impact on an executive, although companies rarely hold these programs accountable for producing measurable results. It may be some time before companies implement tough outcome standards. There are some known reasons why many of these efforts fail, including the following:

- Lack of clear objectives
- Inappropriate or irrelevant development activities
- Lack of follow through on a development plan
- Failure of the executive to learn the right behavior or lessons
- Inability of the executive to learn or change
- Lack of leadership support in the organization for the change
- Lack of business application for the change

At the same time, there are some hallmarks of successful development efforts, including the following:

- An executive who is able and motivated to learn
- Use of the right approaches and techniques for the development need
- Clear learning and behavioral objectives
- Hands-on learning experience
- Focus on specific behaviors and outcomes
- Clear connection to business applications
- Regular feedback and guidance from others
- Support from senior leaders

The emphasis on executive development is likely to continue and even increase in the future as the globalization of business increases. The demand for effective fungible executives will remain high, and corporations will see the benefit of having well-integrated executive selection and development efforts to meet new business challenges and to achieve difficult strategic objectives.

### See Also the Following Articles

Organizational Culture and Climate ■ Person–Environment Fit ■ Work Teams

### Further Reading

- Conger, J. A., & Benjamin, B. (1999). *Building leaders: How successful companies develop the next generation*. San Francisco: Jossey–Bass.
- Dotlich, D. L., & Noel, J. L. (1998). *Action learning: How the world's top companies are re-creating their leaders and themselves*. San Francisco: Jossey–Bass.
- Fulkerson, J. (2002). Growing global executives. In R. Silzer (Ed.), *The 21st century executive: Innovative practices for growing leaders at the top* (pp. 300–334). San Francisco: Jossey–Bass.
- Hollenbeck, G. (2002). Coaching executives. In R. Silzer (Ed.), *The 21st century executive: Innovative practices for growing leaders at the top* (pp. 137–167). San Francisco: Jossey–Bass.
- Jeanneret, R., & Silzer, R. (1998). *Individual psychological assessment: Predicting behavior in organizational settings*. San Francisco: Jossey–Bass.
- Marcos, V. (2002). Developing executives. In R. Silzer (Ed.), *The 21st century executive: Innovative practices for growing leaders at the top* (pp. 168–199). San Francisco: Jossey–Bass.
- McCall, M., & Hollenbeck, G. (2002). *Developing executives: The lessons of international experience*. Boston: Harvard Business School Press.
- Silzer, R. (2002). Selecting leaders at the top: Exploring the complexity of executive fit. In R. Silzer (Ed.), *The 21st century executive: Innovative practices for growing leaders at the top* (pp. 77–113). San Francisco: Jossey–Bass.
- Silzer, R. (Ed.). (2002). *The 21st century executive: Innovative practices for growing leaders at the top*. San Francisco: Jossey–Bass.



# Extreme Environments and Mental Function

Robert B. Bechtel

University of Arizona, Tucson, Arizona, USA

1. Definitions
  2. The Environments
  3. Conclusions
- Further Reading

## GLOSSARY

- Altiplano* The highlands of Peru.  
*chacheeko* Local name for a newcomer in Alaska.  
*cold soaked* Frozen.  
*dehydration* Loss of water in the body.  
*desertification* Land turning into a desert.  
*diurnal rhythms* The daily cycle of the sun.  
*head bolt* An electrical outlet for cars to plug into at a parking lot.  
*heat exhaustion* Collapse due to overheating and loss of water.  
*heatstroke* More serious condition due to body's failure to deal with heat.  
*hypothermia* Loss of body heat.  
*hypoxia* Lack of oxygen.  
*Monge's disease* Altitude sickness.  
*photophobia* fear of light.  
*puna* Local word in Bolivia for altitude sickness.  
*soroche* Local word in Peru for altitude sickness.  
*tacya* Ironwood digging stick used in the Peruvian Altiplano.  
*the 200 Club* Initiation rite at the U.S. polar station; one joins by running naked to the pole and back when the temperature is  $-100^{\circ}$  F.

Extreme environments are of two kinds: the permanent ones where people live year-round (e.g., high mountains, deserts, cold regions, jungles) and the temporary ones such as space flight, simulation laboratories, and the temporary incursions into the permanent extremes (e.g., mountain climbers, missionaries, explorers). Problems of function are most frequent in the temporary incursions into the permanent areas and places such as Antarctica, space flight, and the simulation laboratories. Extremes are being made more endurable by introducing communications technology and more creature comforts.

## 1. DEFINITIONS

The definition of extreme environments is very geocentered. People living in the heavily populated and industrialized latitudes around the tropics of Capricorn and Cancer assume that equatorial and arctic climates are extreme, whereas those living in these "extreme" latitudes think of them as normal. Similarly, those near sea level think of altitudes of a mile or more as extreme, whereas those living at these altitudes think of them as healthier. Thus, one person's extreme can be another person's normal surrounding. "Extreme" as a label acquires some credence when it involves environmental conditions that bring about physiological responses that

interfere with the normal functioning required to get the business of living done. In high altitudes, this includes hypoxia (i.e., lack of oxygen); in colder climates, it includes frostbite and hypothermia; and in hot climates, it includes heat stroke, heat exhaustion, and dehydration. In jungles, what it includes is less clear, but extreme humidity, exotic diseases, and a myriad of insects make it at least unusual compared with more temperate climes.

When it comes to environments such as space travel, balloon exploration, and experimental laboratories that simulate space travel, no one has any trouble calling these extreme. Similarly, when mountain climbers, explorers, and skiers climb to great heights, these are also seen as extreme.

A distinction must be made, however, between those latitudes and altitudes where people live year-round and have a chance to adapt through generations and those extremes where visitors enter for a brief time and then leave. The latitudes and altitudes where people live year-round under harsh conditions are a permanent situation, whereas those where there are only short incursions are best labeled “temporary.”

## 2. THE ENVIRONMENTS

### 2.1. The Temporary Environments

#### 2.1.1. Antarctica

Although Antarctica is now inhabited year-round by people from a variety of countries, all of these people are there on assignment and will rotate home at some point. No one has been born in Antarctica, although some have died there (largely by accident). Antarctica has been characterized as the highest, driest, and coldest desert on earth. No one thinks of it as a desert due to the miles of ice that have accumulated there in glaciers, but its average precipitation makes it comparable to deserts at more northern latitudes. Thus, Antarctica combines all of the climatic elements that make environments extreme—cold, altitude, unusual diurnal rhythms, and dryness—in addition to another element: isolation.

At an average altitude of 9000 feet (2743 meters), many people experience hypoxia (i.e., lack of oxygen). Hypoxia can cause sleep disturbances, and this can lead to loss of memory, hallucinations, and other mental aberrations. Therefore, sleep deprivation alone could account for failure to perform duties. The average person who goes there, however, experiences less

dramatic symptoms such as failure to attend properly at random times and the resultant mistakes.

The diurnal effects on sleep are also notorious. When Burgess Ledbetter and the author were doing their work in Alaska, they encountered a man who had not slept in 10 days. Thus, the effects of long daylight, as well as the long, dark, and cold nights, can be felt.

The cold of Antarctica has a dramatic effect. Jerri Nielson, the doctor who was rescued from Antarctica due to her cancer, described how the impact of the cold made her forget the orientation lecture. Rivolier, in his field studies, found that the impact of even minimal exposure to the cold was measurable. One reaction that affects performance is the extension of reaction time. This is especially troublesome in outside activities such as driving vehicles and repairing. Perhaps the most common malaise reported in Antarctica is the semicomatose state referred to by many names such as “cabin fever,” being “stir crazy,” having the “big eye,” and just “staring into space.” Not everyone experiences or reports this, but nearly everyone reports periods of wandering attention. The lack of attention can be serious when a schedule must be followed, and there are many reports of personnel who have let schedules slip.

Often overlooked in the concentration on negative experiences are the more positive outcomes reported by many inhabitants of the Antarctic stations. For some, the winter in Antarctica is the “experience of a lifetime.” Nielsen pointed out that Antarctica can be habit forming.

Finally, there is the form of madness called “the 200 Club.” This rite of initiation occurs at the pole station when the temperature drops to  $-100^{\circ}\text{F}$  ( $-73.3^{\circ}\text{C}$ ). The initiates are heated up in a room at  $+100^{\circ}\text{F}$  ( $+37.7^{\circ}\text{C}$ ) and then run naked (with shoes on) to the pole and back. The spectacle is commemorated with photos of icicles dripping from appendages.

#### 2.1.2. Space Travel

Most of the experience in space travel comes from the Russian space station *Mir*, the space shuttle, and the exploratory *Apollo* missions to the moon. The greatest danger in space travel, radiation from the sun, has largely been avoided by rotating the vessels, traveling at times of the quiet sun, and experiencing just plain good luck. The second most serious problem has not actually been dealt with in a remedial fashion. Weightlessness causes the body to dissolve calcium out of the bones. This can cause a serious disability such as osteoporosis. Russian cosmonauts who spent a

long time on *Mir* had to be carried out of the return capsule and took weeks to recover sufficiently to walk normally again. A program of 6 to 8 hours of strong exercise aboard *Mir* did not help alleviate the problem. There seem to be large individual differences in the loss of calcium. Some astronauts were able to walk after landing, whereas others were not. There remains a question as to whether the bone density ever returns to normal. Although bones have not been broken due to this problem so far, it would be a serious problem for longer voyages such as one to Mars.

The trip to Mars constitutes the most extreme of any environment to date. This is due to the communications lag between Earth and Mars of up to 40 minutes and the length of time spent traveling to, staying on, and returning from Mars that could take up to 3 years. Because it is not known whether there is any water on Mars, urine will have to be recycled through a purifying filter (Fig. 1). The “home” on Mars will consist of a reused rocket body (Fig. 2). A human factors problem is the need for windows that considerably weaken the structural integrity of the rocket body.

As in other isolated confined environments, space travel wreaks havoc on sleep patterns. Astronauts and cosmonauts can lose 2 hours of sleep every 24 hours and end up with serious sleep deprivation. Although this has not resulted in hallucinations or psychotic behavior, the lack of attention to detail and time can result in serious mistakes.

### 2.1.3. Simulation Laboratories

In September 1991, eight biospherians were confined in a Mars colony simulation and remained there for 2 years (Fig. 3). Before the 2 years were up, oxygen had to be pumped into the biosphere structure and higher calorie foods were introduced. The atmosphere was reduced to the equivalent of 13,000 feet (3962 meters), and the number of calories in the diet had fallen to 1800 per day. In addition, a conflict had developed among the crew members. This experience was similar to those of Russian attempts at simulation. Nevertheless, the crew finished the mission. Much earlier, the Tektite Lab, an underwater laboratory, had uncomfortable conditions of crowding, humidity, and disturbed sleep. But this crew also finished the mission and did even more work than is done under normal conditions.

The experiences with the simulation laboratories are mixed. Some result in conflict, whereas others do not. The simulations are often considered to be good training.

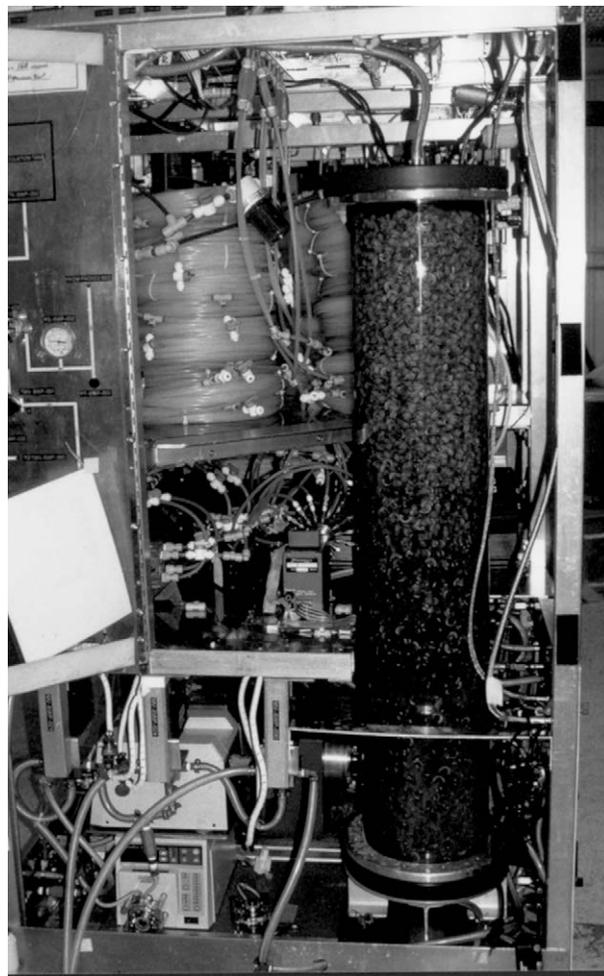


FIGURE 1 Urine recycling apparatus at the National Aeronautical and Space Administration (NASA).

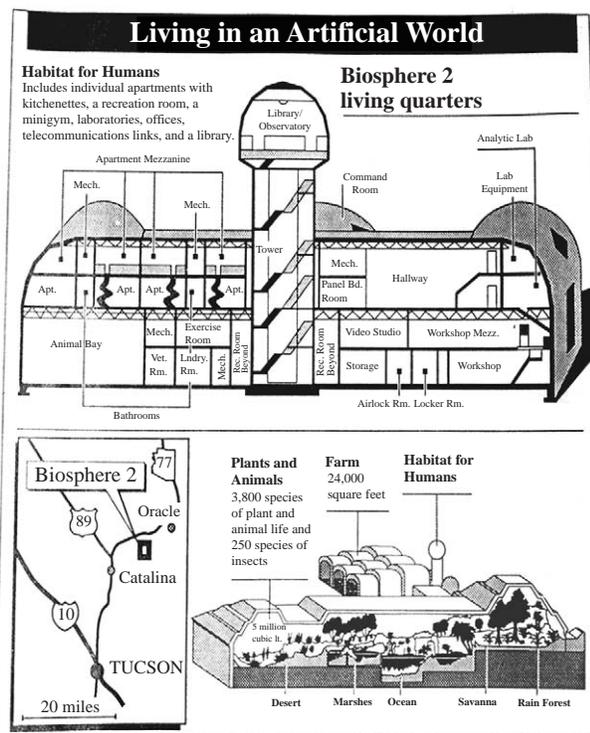
## 2.2. Permanent Extreme Environments

### 2.2.1. High Mountains

There are really only three areas on Earth where significant numbers of people live permanently at high altitudes: the Tibetan Plateau, the Andes, and the highlands of Ethiopia. These are also areas where mountain climbers and miners sometimes work. The continent of Antarctica does not qualify because all persons there are on temporary assignment. Only visitors to these places have the kind of symptoms that seriously interfere with mental function. Lack of oxygen is the primary environmental element that leads to interference with normal functioning (hypoxia) at high altitudes. There are two types of altitude sickness: acute mountain sickness and



FIGURE 2 Rocket body simulating crew quarters for Mars.



Source: Space Biospheres Ventures, Oracle, Arizona

FIGURE 3 Diagram of Biosphere 2. Apt., apartment; Bd, board; Mech, mechanical; Vet, veterinary.

chronic mountain sickness. Acute altitude sickness, which is what visitors experience, is also known as Monge's disease; it is also known locally as *soroche* in Peru and as *puna* in Bolivia. The person with hypoxia experiences restless nights, headache, vomiting (sometimes), tightness in the chest, irritability, and a fear of light (photophobia). Acute altitude sickness can also produce hallucinations and erratic behavior. There are many tales of this kind of behavior. One woman, a native of Peru, described her companion's constant efforts to turn off the car lights while driving on a mountain road. It was the most frightening ride of her life. Another traveler to the high mountains told of his wife's description of a blue horse. These symptoms usually disappear by the fifth day.

Chronic mountain sickness involves swelling of the brain and lungs. This can lead to death, and people who suffer from the symptoms must be quickly removed and taken to lower altitudes. The military has established a rule that oxygen masks should be used above 10,000 feet (3048 meters). Mount Everest is more than 29,000 feet high (8839 meters), yet in 1978 Reinhold Messner climbed to the top without supplemental oxygen. He went on to do it alone in 1980 and then, again without supplementary oxygen, climbed all 14 mountains in the world that are more

than 8000 meters high. This represented another monument to human adaptability.

### 2.2.2. Deserts

Deserts are characterized as being hot, dry, and devoid of the same kind of plant and animal life found in more temperate areas. Authorities disagree on definitions of deserts. Some prefer an extreme of 2 inches or less of rainfall per annum. Others' definitions go up to 15 inches of rain per annum. Some compromise and call these "arid lands." The definition of an arid land can be any place that requires irrigation. Areas on Earth that are seen as deserts include places such as the Sahara in Africa (the largest), the Namib in southern Africa, the Atacama in South America, and Death Valley in the southwestern United States. Areas such as the Sonoran Desert in the southwestern United States and northern Mexico, the Great Sandy in Australia, the Gobi in Mongolia, and nearly all of Saudi Arabia (Fig. 4) are also some of the more obvious ones. It should also be noted again that great parts of the Arctic and Antarctica are also considered desert if desert is defined by low precipitation. The main symptoms suffered in deserts seem to be chiefly from exposure to the sun with the result of heatstroke or heat exhaustion. Both involve many of the same symptoms: faintness, confusion, fainting, and vomiting. But the treatment is very different. Heatstroke, the most serious, requires immediate cooling, whereas heat exhaustion requires immediate water intake (salt is no longer recommended). Heatstroke is best recognized by a lack of perspiration. Thus, dragging the person into the shade and cooling with a water hose or any other method are critical.

It is estimated that the longest the average person could last if lost in the desert is 2 days. Death occurs when the body temperature rises above 113° F (45° C). Having said this, it should also be remembered that in 1905, Pablo Valencia got lost in the desert near Yuma, Arizona, and survived after 6½ days. He traveled only at night and rested during the days. Although many may regard the Grand Canyon as not in the desert category, in 1996 there were 482 search and rescue operations, the result of which 18 people died. It is estimated that more than 700 people have died in the Grand Canyon, with some of these dying from freezing during the winter. Most of the distress came from dehydration. Thus, the canyon seems to fit the temporary extreme category, even though people live at the rim and at the bottom year-round.



**FIGURE 4** The Rub Al Khali (The Hidden Quarter) desert in Saudi Arabia.

Recently, the Sonoran Desert claimed a large number of deaths: more than 30 in 2002 and more than 50 in 2003 due to the illegal migration from Mexico into the United States. All of these people died from dehydration.

### 2.2.3. Cold Regions

The far northern communities in the U.S. state of Alaska as well as those in Canada, Greenland, and the former Soviet Union are easily seen as cold regions, and for many reasons parts of Norway and Finland, all of Iceland, and even the U.S. New England states and Montana winters can seem extreme. Antarctica qualifies as well, but this continent is in a class by itself, as mentioned earlier.

Frostbite is the most immediate danger in northern climates. The military coined the phrase "30–30–30," meaning that flesh freezes in 30 seconds in a 30-mile per hour wind at –30° F (–34.4° C). Despite this, one

can see the black scabs hanging from the cheeks of many Inuit from riding their snowmobiles on the streets of many communities in Canada and Alaska. The average human body stops functioning below 77° F (25° C).

Hypothermia begins with shivering and is then replaced by muscular rigidity, confusion, and loss of consciousness. The confusion has caused many deaths, and to this day it is still sometimes reported in Alaskan newspapers that people, usually *chacheekos* (newcomers), have frozen to death while trying to walk back to gas stations that were within sight after their cars had stalled.

Having set the preceding limits by clinical data, it should be noted that in October 1992 at Point Hope, Alaska, Morris Sage and Sayers Tuzroyluck fell into a lake while fishing. Sage chose to stay in the water because it was +25° F (-3.9° C), whereas the air was -10° F (-23.3° C). Tuzroyluck chose to walk back to the town 7 miles away. Medical authorities say that the average person can last no more than approximately 30 minutes in such conditions. Both men spent 3½ hours in the cold, and Sage's body temperature was measured at 75° F (23.9° C) when he was found. Both men survived without ill effects. This is perhaps testimony to the adaptation of the Inuit people to their climate or to the tremendous variability of humans in response to extreme conditions.

Automobiles can be a hazard in cold regions. They require a complete overhaul, with all fluids (e.g., brake fluid, crankcase oil) being replaced with a thinner variety, the crankcase being wired with a heated dipstick, and the entire apparatus being wired to connect with an electric "head bolt" while parked. If not, the car will freeze—get "cold soaked" in local parlance—in approximately 45 minutes. Many choose to let the engine idle when head bolts are not available.

Diurnal rhythms are a great disrupter of sleep. Many residents of Alaska use blackout curtains and shades during the long summer daylight period to help them sleep.

#### 2.2.4. Jungles

Perhaps too much has been claimed for jungles as extreme climates. They do not have the extreme heat of the desert, nor do they have the heights of the mountains. Yet the humidity combined with a fairly high temperature makes activity uncomfortable and can produce heat exhaustion and heat stroke under certain conditions. Diseases such as malaria as well as exotic animals such as leeches, ticks, poisonous snakes, and

biting insects can test the limits of tolerance. Some even claim that the side effects of anti-malaria drugs are nearly as bad as malaria itself. Certainly, these distractions can produce the same kind of errors as are produced in deserts and cold regions. Also, the alertness required to keep a lookout for pests can often be as distracting as any malaise. Natives to the jungles show remarkable adaptations to their environment, and these cases are reported in many anthropology texts.

### 2.3. The Overlap

People who are assigned to extreme environments but are more than just visitors—who have a designated or indefinite length of stay—can often acquire some of the adaptations of persons living permanently in extreme environments. Peace Corps volunteers on the Altiplano in Peru have adapted to the lack of oxygen from which shorter term visitors may suffer. Likewise, Federal Aviation Agency personnel in Alaska have "cold adapted" after a certain length of stay. Military personnel also report being able to adapt to cold and hot regions. A caution that needs to be applied here is that many of these individuals can have high variability in their adaptation levels depending on age, physical condition, and personal habits.

In the Cold Regions Habitability Study, Burgess Ledbetter and the author encountered a few military wives who had not been out of their houses from October to April. Some of these were suffering from the symptoms of cabin fever, but others were taking the confinement in stride. The important point is that these longer term visitors have a longer time period during which to adapt.

### 2.4. Isolation: The Changing Extreme

With the advent of cell phones and the ability to carry sophisticated communications equipment literally anywhere in the world, the terms of isolation are being changed. During the investigation and remedial intervention in the Biosphere 2 Study, some of the biospherians stated that they had met more people by video and phone than they had during their entire lives prior to the confinement. If this trend continues, it raises a whole new question about the quality of isolation. Will it come to the point where any worker in Antarctica, any space station astronaut, or any missionary in the field can call home at will and see the

faces of his or her spouse and children? This alternative to complete isolation cannot be overlooked if it helps to alleviate the conditions that extreme environments otherwise impose.

### 3. CONCLUSIONS

#### 3.1. The Temporary Environments

What seems clear is that no extreme environment has just one variable affecting function. There are always combinations. In cold regions, it is the cold and the diurnal rhythms. In the deserts, it is the heat and the dryness. Sometimes, there are three or more variables such as cold, dryness, and diurnal rhythms. And these effects can be cumulative over time. Sleep deprivation builds over time and becomes more and more debilitating.

But extreme environments are slowly getting less extreme. Compared with the expeditions and studies done just decades ago, the communications and comforts have improved greatly. Even Antarctica can now get fresh vegetables.

Another element that has improved is the selection of personnel. After years of screening and experiments with various instruments, the selection process has virtually eliminated mental pathology from the remote stations in the world. This technology has been shared across nations, resulting in more stable environments in places such as Antarctica, the space shuttle, and other remote government stations. But is there a price to pay for this? Researchers such as Suedfeld and Palinkas have pointed out the positive benefits of isolation: the quiet, the close fellowship, and the sense of being in another world. Will this be lost in the clutter of modern communications?

Finally, will there be permanent space stations and colonies on the moon and Mars? Will people who stay there for a long time be able to return to Earth and readapt to its strong gravity? There seems to be some evidence that they will not. Stays on the space station, and even in the low-gravity colonies, might have to be limited.

#### 3.2. The Permanent Environments

Much is being revealed about the permanent extreme environments that raises questions about the variables that affect the occupants of these environments. Migration is evident in the Altiplano, the jungles, and the deserts. Economic forces may be the most relevant of all variables, and they are sending people to places where they can earn a better living. The use of a digging tool, the *tacya*, in Peru is being replaced with machinery, making much labor surplus. The Brazilian jungle is fast disappearing, and the deserts seem to be spreading, causing a new term to be coined: desertification. These new, largely economic variables are changing not only the populations of these environments but also the actual geography itself. Political aspects enter as well, especially in places such as Tibet. But even these are changing the character of the land. Only one thing is certain: change.

#### See Also the Following Articles

Environmental Psychology, Overview ■ Environmental Stress

#### Further Reading

- Ashcroft, F. (2000). *Life at the extremes: The science of survival*. Berkeley: University of California Press.
- Harrison, A. (2001). *Spacefaring: The human dimension*. Berkeley: University of California Press.
- Harrison, A., Clearwater, Y., & McKay, C. (Eds.). (1991). *From Antarctica to outer space*. New York: Springer-Verlag.
- Nielsen, J. (2001). *Ice bound*. New York: Hyperion.
- Rivolier, J., Goldsmith, R., Lugg, D., & Taylor, A. (Eds.). (1988). *Man in the Antarctic*. New York: Taylor & Francis.
- Steele, G. D. (2001). Polar moods: Third quarter phenomena in the Antarctic. *Environment & Behavior*, 33, 126–133.
- Stuster, J. (1996). *Bold endeavors*. Annapolis, MD: Naval Institute Press.
- Suedfeld, P., & Weiss, K. (Eds.) (2000). Antarctica. *Environment & Behavior*, 32, 5–156.
- Ward, M., Milledge, J., & West, J. (1989). *High altitude medicine and physiology*. Philadelphia: University of Pennsylvania Press.



# Extroversion–Introversion

Gerald Matthews

University of Cincinnati, Cincinnati, Ohio, USA

1. Psychometric Foundations
  2. Biological Bases
  3. Cognitive–Social Perspectives
  4. Clinical Applications
  5. Occupational Applications
  6. Conclusions
- Further Reading

## GLOSSARY

**arousal** Nonspecific excitation of the brain or one of its major subsystems (e.g., the cerebral cortex).

**factor analysis** A family of multivariate statistical techniques that aim to explain covariance among variables in terms of a smaller number of latent factors.

**interactionism** A popular contemporary approach to personality that sees traits as interacting with situational factors to generate behavior.

**personality disorder** An enduring pattern of abnormality in personality that is typically referred to in terms of various discrete disorder types described by the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

**personality trait** A stable personality characteristic that describes a continuum of individual differences in behavior expressed consistently across a variety of different situations.

**self-regulation** The cognitive processes and behaviors that support the pursuit of personal goals within a changing external environment.

Extroversion–introversion (E–I) is a major personality dimension that is an essential part of comprehensive trait models, including the Five-Factor Model. Various questionnaires provide valid and reliable measurement of E–I. The person's level of extroversion versus introversion depends on both biological and social–cognitive factors. The mechanisms that shape the development of E–I are not well understood. On the biological side, brain systems for arousal and reward have been implicated. On the social–cognitive side, biases in various aspects of self-regulation may be important. E–I is modestly associated with many criteria related to real-life functioning, including mental health, social behaviors, and work performance. E–I should be understood within an interactionist context, such that both person and situation factors are important as determinants of behavior. Empirical studies support clinical and occupational applications based on the assessment of E–I.

## 1. PSYCHOMETRIC FOUNDATIONS

### 1.1. Origins and Early Studies

The extroversion–introversion (E–I) trait refers to a dimension of personality that contrasts qualities such as sociability, assertiveness, and enthusiasm with qualities such as social reserve, quietness, and thoughtfulness. Its place as a central element of personality has been recognized since antiquity. Jung popularized the

idea of extroverts and introverts as distinct types characterized by their tendencies to channel “psychic energy” outward and inward, respectively. However, the typological approach is no longer representative of mainstream personality research. Current research assumes that E-I is a continuous dimension or trait, such that the level of E-I is approximately normally distributed in the general population.

Dimensional models of personality traits, including E-I, are typically based on factor analysis of questionnaire data. The first questionnaire measure of E-I developed on this principle, the Inventory of Factors STDCR, was published by Guilford in 1940 and included social introversion and thinking introversion factors as two of five basic personality traits. Eysenck conducted especially influential factor-analytic research that generated a series of reliable questionnaires, beginning with the Maudsley Personality Inventory in 1956. He also pioneered the use of experimental methods for validating questionnaire assessment and investigated the role of E-I in mental illness.

### 1.2. Assessment of Extroversion-Introversion

A parallel approach to assessment was based on analysis of the personality-descriptive terms that appear in natural language (i.e., the lexical approach). Fiske identified five factors in personality ratings of this type, including a factor of “confident self-expression” that has been subsequently identified with extroversion. Lexical studies have converged with research based on personality questionnaires in suggesting that there are five fundamental personality traits (the “Big Five”), each of which is defined by more narrowly defined lower order traits, within a hierarchical model.

Costa and McCrae’s Revised NEO Personality Inventory (NEO-PI-R) questionnaire assesses extroversion as one of the Big Five. The Five-Factor Model (FFM) has gained support from the identification of the Big Five, through factor analysis, in various questionnaires developed for other purposes. Furthermore, E-I meets four criteria listed by Costa and McCrae for a trait to be considered fundamental: E-I appears to be biologically based, it is culturally universal, it refers to important psychological concepts, and measures of E-I have good criterion and predictive validity.

There has been some controversy over the lower order traits that define extroversion as a broad trait. Eysenck distinguished sociability and impulsivity as

important constituents of the trait, whereas the NEO-PI-R assesses warmth, gregariousness, assertiveness, activity, excitement seeking, and positive emotionality as primary facets. Sociability and outgoing self-confidence are universally seen as defining features of extroversion, but the place of traits such as impulsivity, activity, and dominance is still being debated. Thus, although various assessments of E-I converge, the composition of the E-I factor differs somewhat across various psychometric models. A more radical proposal was advanced by Hogan, who suggested dividing extroversion into separate factors of sociability and ambition, with the latter being seen as more predictive of job performance.

The NEO-PI-R and other questionnaire scales for E-I are internally consistent and substantially intercorrelated with one another, showing convergent validity. Measures are also stable over time, showing test-retest correlations of 0.6 to 0.7 over periods of several years. Published measures of E-I do not appear to be strongly biased by response sets such as social desirability, although such concerns should not be ignored in applied research. In addition, self-ratings of E-I converge quite well with ratings made by other persons, perhaps because E-I is one of the most salient and visible aspects of personality.

## 2. BIOLOGICAL BASES

### 2.1. Behavior and Molecular Genetics

The heritability of E-I has been investigated using twin and adoption studies. Structural equation models fitted to behavior genetic data suggest that broad heritability of E-I is approximately .50. The remaining variance is attributed to the environment, gene-environment interaction, and measurement error. Models suggest that the influence of the unshared environment, unique to each child, is considerably stronger than that of the shared environment related to the family. Molecular genetic studies are beginning to investigate specific alleles that may correlate with E-I. For example, variations at the D4 dopamine receptor gene have been found to correlate with some traits related to E-I, such as excitement seeking and positive emotionality, although the variance explained by any single allele is likely to be small.

### 2.2. Psychophysiological Studies

The two leading biological theories of E-I are arousal theory and reinforcement sensitivity theory (RST).

Zuckerman conducted important work on possible biochemical bases of E–I, but that is beyond the scope of this article. Eysenck’s arousal theory identified arousability of a reticulo–cortical circuit controlling alertness as the key system that underpins E–I. Introverts are said to be more easily aroused, so that they regulate arousal by avoiding intense stimuli, whereas less readily aroused extroverts are prone to seeking stimulation. Numerous psychophysiological studies provide mixed support for the greater arousability of introverts, with studies of evoked potentials and phasic electrodermal responses providing the most consistent results. Replicable results are obtained only with careful control of experimental parameters.

Gray’s RST links E–I to sensitivity of primarily dopaminergic brain systems controlling reward and sensitivity to positive reinforcement signals. This system may be more sensitive in extroverts than in introverts, so that extroverts are more readily conditioned by reward signals than are introverts. It may also influence positive emotion. Some studies suggest that extroverts are more psychophysiological responsive than introverts in rewarding situations, but the data are often inconsistent. It seems likely that multiple neural mechanisms may influence E–I. Both arousal theory and RST may have some validity, but neither fully explains the data currently available.

### 3. COGNITIVE–SOCIAL PERSPECTIVES

#### 3.1. Information-Processing Studies

E–I correlates modestly with a wide variety of measures of information processing. These relationships vary with task demands. Typically, extroverts show superior performance on verbal short-term memory, divided attention, speech production, and speeded motor response tasks, whereas introverts do better on sustained attention, long-term memory, and reflective problem-solving tasks. Extroverts also tend to show a lower response criterion, and favor speed over accuracy, in some task paradigms. In addition, extroverts tend to outperform introverts in arousing conditions, although this effect may reverse during evening hours.

Biological theories have difficulty in explaining the dependence of E–I effects on the information-processing demands of tasks. Instead, it seems that extroversion is associated with a complex pattern of multiple processing advantages and disadvantages relative to introversion.

Matthews proposed that the information-processing correlates of E–I have an adaptive function in preparing extroverts and introverts for the cognitive–social environments they prefer. For example, the superior verbal short-term memory and speech production skills of extroverts may support the development of conversational skills that facilitate adaptation to demanding social environments.

#### 3.2. Social Cognition

Extroverts and introverts tend to differ in various high-level aspects of cognition, including self-cognitions. Extroverts are more likely than introverts to appraise demanding situations as challenging and to cope through use of proactive problem-focused strategies. Extroverts also report greater levels of social participation and affiliative needs than do introverts. Although extroversion is associated with superior social skills, the trait may also be associated with some maladaptive aspects of social cognition such as a tendency toward narcissism. In addition, extroversion has been linked to likelihood of divorce, perhaps due to impulsivity in forming and dissolving intimate relationships. Extroverts and introverts may differ in self-regulation, pursuing somewhat different goals in life supported by different styles of social cognition. However, much of the variance in social cognition is unrelated to personality traits and should be understood with reference to the individual’s life circumstances and social learning.

### 4. CLINICAL APPLICATIONS

#### 4.1. Emotional Disorders

The trait most strongly linked to anxiety and mood disorders is neuroticism, but extroversion may also be implicated. Typically, patients tend to be somewhat introverted, and introversion predicts various symptoms of emotional pathology in longitudinal data to a modest degree. Introverts are also somewhat more vulnerable to subclinical stress symptoms than are extroverts. Introversion may be especially linked to anhedonia (i.e., lack of positive affect) in depression. Many non-clinical studies suggest that introverts experience less positive affect and happiness than do extroverts, although the magnitudes of the association between E–I and affect vary considerably across studies. Some authors propose that extroverts are temperamentally disposed to experience positive affect irrespective of

circumstances, whereas others emphasize the role of situational modifiers such as social involvement.

Longitudinal studies suggest that social introversion may be a risk factor for depression, although the emotional disorder itself may raise introversion as a consequence of the patient's difficulties in social functioning. Assessment of extroversion may be useful to the clinician in understanding the patient, selecting a treatment, and anticipating the course of therapy. For example, extroverted patients may respond better to interpersonal therapy than do introverted patients.

#### 4.2. Personality Disorders and Criminality

Traditional diagnoses of discrete personality disorders may be complemented with dimensional approaches that describe abnormal personality in terms of continuous dimensions. Dimensional studies have related several features of abnormal personality to E-I. Extroversion has been linked to excessive stimulus seeking and impulsivity, whereas introversion overlaps with social avoidance and inhibition, including schizoid personality. Thus, extremes of E-I may create a potential for personality disorders linked to these qualities. Eysenck's theory predicted that extroversion, especially in combination with neuroticism, should be related to a higher incidence of criminality and deviant behaviors because underarousal prevents effective conditioning. Although some studies have supported this prediction, the data are inconsistent. For example, persons incarcerated for long periods of time tend to show rather low levels of extroversion. It seems likely that impulsivity, rather than extroversion per se, is linked to criminality, and the more harmful aspects of impulsivity may be related more to traits such as psychoticism and low conscientiousness than to extroversion.

### 5. OCCUPATIONAL APPLICATIONS

#### 5.1. Vocational Choice

Extroverts and introverts exhibit clear differences in their occupation choices. Extroverts tend to predominate in jobs that are people oriented (e.g., sales) and/or stressful (e.g., financial dealing), whereas introverts appear to be attracted to jobs that are more solitary and reflective in nature (e.g., scientific research). Similarly, studies of vocational interests show that extroverts have

stronger social and enterprising preferences. In general, E-I is only weakly related to job satisfaction, but the nature of the job seems to influence whether extroverts or introverts are better adjusted. For example, introverts show better tolerance for repetitive monotonous work than do extroverts.

#### 5.2. Job Performance

There is an extensive literature on E-I and job performance. Often, correlations are small, as demonstrated by meta-analyses. However, such analyses may be misleading in that they neglect the likely role of the nature of the job as a factor moderating the extroversion-performance association. Research driven by a priori hypotheses has been more successful in demonstrating the occupational relevance of E-I. Extroversion relates to better performance in more jobs requiring social skills such as sales. Extroversion also relates to superior performance during training, perhaps because extroverts handle novelty better than do introverts. Conversely, introverts may perform better than extroverts in monotonous work environments. Introverts may also be less accident prone than extroverts in transportation and industrial settings, although this relationship may be mediated by impulsivity rather than by E-I per se.

### 6. CONCLUSIONS

E-I is a robust trait that is an essential component of any comprehensive account of human personality. There is good convergence among various questionnaire measures of the trait, although there is still debate over its constituent traits. Behavior genetic studies show that both genotype and environmental influences are important influences on phenotypic E-I. Psychophysiological and cognitive-social studies have made some progress in revealing the sources and consequences of an individual's level of E-I, but current theories of the trait are not fully satisfactory. Nevertheless, assessment of E-I may be important in a variety of applied contexts, including clinical and occupational psychology.

#### See Also the Following Articles

Psychometric Tests ■ Self-Control ■ Traits

**Further Reading**

- Costa, P. T., Jr., & Widiger, A. (Eds.). (2002). *Personality disorders and the Five-Factor Model of personality* (2nd ed.). Washington, DC: American Psychological Association.
- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural sciences approach*. New York: Plenum.
- Furnham, A., & Heaven, P. (1999). *Personality and social behaviour*. London: Edward Arnold.
- Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Nyborg, H. (Ed.). (1997). *The scientific study of human nature: Tribute to Hans J. Eysenck at eighty*. Amsterdam: Pergamon/Elsevier Science.



# Eyewitness Identification

**Colin G. Tredoux**

University of Cape Town, Rosebank,  
South Africa

**Christian A. Meissner**

Florida International University,  
Miami, Florida, USA

**Roy S. Malpass and  
Laura A. Zimmerman**

University of Texas at El Paso,  
El Paso, Texas, USA

1. Introduction
  2. The Event
  3. The Witness
  4. The Perpetrator
  5. Obtaining Information from the Witness
  6. The Witness Identification
  7. The Prosecution
  8. Conclusion
- Further Reading

## GLOSSARY

**blank lineup (target absent lineup)** A lineup in which all the members are known to be innocent.

**cognitive interview** An interview technique designed to enhance memory and communication of events through the use of guided retrieval techniques.

**effective size** The number of plausible members in a lineup.

**encoding specificity principle** If cues that were present at encoding are also present at retrieval, recall of stored information is more likely.

**estimator variables** Factors in eyewitness situations whose influence can only be estimated and are not under the control of the justice system.

**eyewitness** A person who observes an event and can provide information about that event.

**lineup bias** A lineup constructed in a manner that leads the suspect to stand out from the other lineup members.

**mock witness** A person who is not a witness to a crime but is asked to identify a perpetrator from a lineup based on another source of information, such as a verbal description.

**postevent (mis)information** Events that occur after a crime that influence eyewitness memories of the crime.

**sequential lineup** A lineup presentation procedure in which one lineup member is presented to the witness at a time.

**showup** The presentation of a single suspect to a witness in order to determine if that suspect is the perpetrator.

**simultaneous lineup** A lineup presentation procedure in which all lineup members are presented to the witness at the same time (i.e., in a line).

**system variables** Factors that may influence eyewitness memory and are under the control of the criminal justice system.

**verbal overshadowing effect** When the act of verbally describing a perpetrator decreases the witness's ability to accurately identify a perpetrator from a lineup.

An eyewitness's identification of a perpetrator is a prevalent form of incriminating evidence presented in a criminal case. However, eyewitnesses are frequently mistaken, and their errors have led to the conviction of innocent individuals. For more than a century, psychologists have sought to understand the variety of factors that may influence the accuracy of eyewitness identification. This article reviews the current state of

knowledge in the science of eyewitness identification according to the sequence of events that is likely to unfold from the time that an eyewitness makes an observation to the prosecution of the perpetrator of the crime.

## 1. INTRODUCTION

The testimony of eyewitnesses is an important factor in many criminal cases. Cases that hinge on eyewitness testimony can bring perpetrators of crimes to justice, set innocent people free, and enable the police and courts to carry out their prime societal responsibilities of preserving law and order. However, eyewitnesses are frequently mistaken, leading to the conviction and imprisonment of innocent people for crimes they did not commit. The extent of this problem is becoming clearer with the exoneration of innocent people through DNA evidence. According to The Innocence Project, a New York-based legal team that reviews case records of incarcerated individuals who claim their innocence, there are currently more than 140 documented DNA-exoneration cases in the United States. Mistaken eyewitness identification is generally the most frequent source of evidence used to convict an innocent defendant. Of the current database of DNA exonerations, approximately 85% are estimated to have involved faulty eyewitness identification evidence.

Acknowledgment of the dangers of eyewitness identification is not new. In England, for example, the Criminal Law Revision Committee of 1972 gave explicit recognition to this fact: "We regard mistaken identification as by far the greatest cause of actual or possible wrong convictions." Psychology and law researchers have long recognized the vagaries of eyewitness identification, conducting original empirical research from as far back as the late 19th century. This research has made a significant contribution to police and legal practice, partly because psycholegal researchers bring an empirical and scientific perspective to a legal problem, which is an approach that is quite unnatural for lawyers and jurists to take. In the current discussion of eyewitness identification, we structure our review, as far as possible, according to the sequence of events that is likely to unfold when an eyewitness makes an observation. The processes that eyewitnesses go through at each stage of the process can vary greatly, and the interests of researchers are quite different at each stage, as are the challenges they face.

### 1.1. Stages in the Eyewitness Process

There are three main stages of the eyewitness process: (i) perception of an event and the persons central to it, (ii) storage and assimilation of the information extracted from that experience, and (iii) remembering aspects of the experience and acting on it. Time delays between the first and third stages can vary greatly. Some information may be requested almost immediately, as when a witness at the crime scene is asked to describe an event or an offender. Other information may be requested much later, such as when a suspect is presented for identification and the witness is asked to make a complex set of judgments and decisions.

### 1.2. Categorization of Variables

Pertinent psychological variables differ at the various stages of the eyewitness process, and different bodies of theory and research may apply to each variable to understand them. Two important categories of variables are system variables and estimator variables, distinguished by Gary Wells in 1978. Estimator variables arise during the first stage of the eyewitness process and are those factors over which the criminal justice system has no control, whose degree and magnitude can only be estimated after the fact. By definition, these include attributes of the witness, the event, and the perpetrator, and they may involve such things as the opportunity the witness had to view the event and perpetrator, the state of mind of the witness at the time, the attentional constraints that have impinged upon the witness as a result of the demands made on the witness's attention, or any person variables (such as race or gender) that may have influenced memory performance. System variables, on the other hand, are under the control of the criminal justice system. These variables focus on the treatment of the witness, the manner in which law enforcement officers attempt to obtain information from the witness, and the interaction of person attributes with these processes. System variables occur at later stages of the process (e.g., the second and third stages), at the points where information is requested of the witness.

The demands on research vary depending on the stage of the witnessing process and the type of variable under study. For early stage estimator variables (such as the quality of memory encoding), an effective understanding requires one to simulate them in the laboratory or to study them *in situ*, in collaboration with law enforcement. The latter is rarely done. Law

enforcement generally has a very limited interest in and capability for studying these matters, although worldwide there are a few major law enforcement agencies that engage in research in this field. Many other estimator variables cannot feasibly be studied in laboratory simulations for ethical or logistical reasons, such as high levels of stress, fear, and violence. We review, in general terms, what is known about major estimator variables in this article.

System variables are quite different from estimator variables in that they are implemented in similar ways by law enforcement and researchers, and they act upon persons who have already been through the first stages of a criminal event. Thus, witnesses will naturally vary in the amount of information they have encoded about the event and the perpetrator. Witnesses in laboratory studies of system variables will often participate in crimelike events, but these are not intended as precise simulations of criminal events. Rather, they are merely ways of giving witnesses a temporal event, with some degree of complexity, and that includes information about an event and a perpetrator about which they might later be tested. Of most interest is the effect of various system variables (such as the structure of a lineup or the instructions provided to a witness) on the accuracy of information obtained from the witness.

### **1.3. General Research Assumptions**

Generally speaking, the following assumptions underlie research in this area. First, all individuals function according to the same general set of psychological processes. For example, perception and memory do not work differently for different people in any fundamental sense, and the psychological processes that witnesses engage in do not fundamentally change in different contexts or situations. Second, information must be encoded at the time of the witnessed event for it to be reported later. If the witness does not attend to the event or to information within the event, then information about that event will not be available for subsequent recall or recognition. Any distractions of attention away from portions of an event will result in decreased encoding and therefore an absence of subsequent memory for such information. Likewise, attraction of attention to some elements will result in superior encoding of such information. Finally, the willingness of the witnesses to identify someone from a lineup involves a decision process that can be somewhat controlled by those administering the identification procedure. Although the specific factors leading to identification

may differ from person to person, the decision processes utilized are assumed to be highly consistent.

## **2. THE EVENT**

An eyewitness is defined by an event: This is something he or she directly observes and will later be asked to provide information about. In order to study eyewitnesses, then, we need to study the kind of events to which they bear witness. However, witnessed events range from the mundane to the terrifying and are potentially infinite in their diversity. Eyewitness researchers are thus faced with the daunting task of understanding how eyewitness memory and related processes function across a wide range of possibilities. It is important to understand that eyewitness recognition and identification of the criminal offender is based (for the most part) on comparing a face (person or photo) shown to the witness with an image of a face stored in memory. In general, the lower the quality of the image stored in memory, the less likely an eyewitness is to make an accurate recognition decision. Any event or process that degrades perception (i.e., acquisition of the original image) also reduces the likelihood of a correct recognition or identification. The following factors contribute to a witness's opportunity to encode information about an event or an individual.

### **2.1. Time to View the Event/ Perpetrator**

The duration of the witness's exposure to the offender is related to later recognition performance, such that limiting exposure time generally reduces eyewitness accuracy. This has been shown in both laboratory and archival studies. Although a range of specific times has been employed across studies, it is difficult to calibrate specific time durations to specific levels of identification accuracy, particularly given the many other factors involved. Furthermore, it is difficult to accurately assess the time a witness had to view the perpetrator *in situ*, as studies have shown that witnesses' post hoc recollection of time estimates can be very inaccurate.

### **2.2. Distance from the Event/ Perpetrator**

The physical distance between the witness and the offender is also related to later recognition performance,

with longer distances leading to poorer rates of identification. Again, it is difficult to calibrate specific distances with specific levels of identification accuracy. In addition, witnesses can be very inaccurate at estimating the distance between themselves and the offender, so the information may not be helpful or reliable even if we were able to calibrate accuracy in a laboratory.

### 2.3. Visibility

Light levels (e.g., time of day) and obscured illumination (e.g., sun shining in the witness's face) have been studied by eyewitness researchers and shown to influence both perceptual and identification processes. Specifically, poor lighting and obscured illumination result in lower rates of accurate identification.

### 2.4. Stress, Fear, and the Presence of a Weapon

A witness's attention may be impaired or distracted if he or she focuses on the psychological stress or fear accompanying a criminal or otherwise traumatic event. Even if a witness tries to be attentive, high fear or stress may hinder the accuracy of subsequent identifications. Psychologists have often described the effects of stress or fear according to the Yerkes–Dodson law, which posits that very low or very high amounts of arousal are most likely to impede encoding and recognition. The presence of a weapon, which may be accompanied by stress or fear, has also been shown to divert a witness's attention away from the face of the offender. In general, the presence of a weapon reduces both the subsequent quality of the description provided by the witness and his or her attempts at perceptual identification of the perpetrator.

## 3. THE WITNESS

Individual eyewitnesses may differ along a number of dimensions or attributes, and it is natural to ask whether any of these attributes are relevant to an eyewitness' performance. Research has identified several important characteristics.

### 3.1. Witness Age

Although eyewitness memory remains relatively stable across a wide age range, it is less reliable for individuals

in certain age groups. For example, very early in life, children demonstrate a limited ability to recall information and show lower accuracy in the identification of faces. They also have a proclivity toward choosing (or guessing) during a lineup identification task, leading to an increased likelihood of mistaken identification of an innocent person. By early teen years, their performance is generally indistinguishable, on average, from that of adults. Although children tend to recall less information when compared with adults, the overall proportion of correct information recalled does not typically differ. Older adults also become somewhat variable, but in the absence of a disease process (e.g., dementia), older adults are indistinguishable, on average, from younger adults. Sensory changes, such as cataracts or changes in the contrast or sharpness of vision, may affect an elderly witness's ability to gather and encode information about a situation or offender. Elderly adults also appear to be more susceptible to the effects of suggestive questioning or postevent misinformation than the average adult. Although children may also be susceptible to suggestion, this is presumed to be due to their unwillingness to challenge an adult's authority. The elderly are more likely to forget the sources of their information.

### 3.2. Alcohol and Other Drugs

There is very little research specifically on the effects of alcohol and other drugs on eyewitness memory for faces. Although research shows that alcohol somewhat impairs memory for verbal materials, its influence on face recognition or person identification tasks has varied across studies. Whereas some studies have shown no effect of alcohol on lineup identification, others have demonstrated impairment in both recall and lineup identification performance, particularly when witnesses consumed alcohol prior to viewing the event and arousal was limited. Of course, drugs that impair or disrupt perception necessarily impair encoding and memory process as well.

### 3.3. Witness Race, Gender, and Occupation

No reliable evidence demonstrates any general advantage or disadvantage in eyewitness memory or identification associated with the race, gender, or occupation of the eyewitness [although, it will be noted later that the race(s) or gender(s) of the eyewitness and perpetrator may interact]. However, men and women appear to differ in the type of information they recall about an

event. For example, women tend to recall more items such as clothing, whereas men focus on items such as the type of vehicle or weapon.

Law enforcement officers are commonly assumed to be superior at identifying faces and remembering the details of events. Although research has generally found that the level of experience police officers have, either in years on the job or through training, does not significantly enhance their recognition ability, it does have a positive influence on the quality of the descriptions they provide. Officers are generally able to provide more detailed accounts of an event, and they appear to be less susceptible to the effects of postevent (mis)information when compared to laypersons. Officers also provide more correct descriptive information about the perpetrator than do laypeople, without an increase in incorrect information, and tend to elaborate on action details more than laypeople. This elaboration on action details, however, must be weighed against officers' greater tendency to misperceive innocent actions as criminal in nature. Empirical studies attempting to train individuals to remember events and faces have demonstrated a similar pattern; namely, individuals' recall accuracy can be improved for details of an event, but it is difficult to improve recognition accuracy for faces.

### 3.4. Witness Confidence

Eyewitnesses will often claim to have great confidence in their ability to identify a perpetrator, but for more than 30 years, research examining the utility of confidence as a predictor of accuracy has generally demonstrated a weak relationship between degree of certainty and identification accuracy. However, there are some conditions in which a stronger relationship may exist. Initial judgments made with very high confidence, for instance, have been shown to be quite diagnostic of witness accuracy.

Even though a witness's confidence may initially be related to the accuracy of the memory, that relationship can be changed dramatically by events that occur after the eyewitness makes a positive identification. If eyewitnesses are told immediately following a positive identification that they have correctly identified a suspect, not only does their degree of confidence increase but also their memory for the crime itself may change. For example, they may become more likely to report that they saw the criminal longer and under better viewing conditions than they had previously reported. In summary, although evidence exists for a relationship between confidence and accuracy under some conditions, confidence has been

shown to be quite malleable. Thus, a witness's statement of confidence, in general, may prove to be a poor indicator of identification accuracy.

## 4. THE PERPETRATOR

Thus far, we have considered only one of the main actors in the drama that unfolds when an eyewitness observes an event, namely the eyewitness himself or herself. The other actor is, of course, the perpetrator (or perpetrators) of the crime, and several factors have been shown to influence memory for the perpetrator.

### 4.1. Disguise

Disguises are frequently used during the commission of a crime, and their presence can significantly impair encoding of the perpetrator's face. Disguises may include masks, sunglasses, or anything that obscures significant portions of the face. In general, the upper portions of the face (e.g., eyes and hair) provide the most important information for later identification, and disguises that hinder the encoding of these parts of the face are more likely to prevent identification of the perpetrator.

### 4.2. Distinctiveness and Typicality of the Perpetrator

In general, typical faces are more difficult to distinguish from other faces in memory, resulting in a higher likelihood of false identification. In contrast, if the offender is distinctive in some way, recognition may be enhanced because the presence of unusual attributes (such as Richard Nixon's chin or Mikhail Gorbachev's forehead birth mark) can make a face easier to remember and help to distinguish it from other faces in memory. However, if the lineup identification procedure is carried out properly (as discussed later) the distinctive suspect will be placed among others who share the same distinctive attribute, requiring the witness to identify him or her without relying solely on the memorable attribute. If the witness's attention was previously drawn to the distinctive attribute, to the neglect of other attributes or away from a holistic perception of the face, then the witness's identification accuracy may be decreased.

Other facial attributes, such as attractiveness or facial "typicality," have been shown to increase the sense that a face has been seen before, resulting in an increase in

the rate of false identifications. There also appear to be a number of widely held stereotypes about general facial appearance, which influence the encoding of information about faces and may therefore influence lineup identifications by eyewitnesses. For example, faces that are stereotyped as “criminal” in appearance are more likely to be identified from a lineup.

### **4.3. Familiarity Due to Repeated Viewing**

Not surprisingly, if the offender was previously unknown to the witness, recognition is less likely than if the offender was known to the witness. However, even when a perpetrator seems familiar, witnesses may not always be correct in their perceptions, especially if a sense of familiarity develops later in the investigative process. If the suspect was viewed committing the crime, the witness may have increased feelings of familiarity at the time of recognition and identification. However, if the suspect did not commit the crime, it is still possible that familiarity can develop after the witness views mug shots and photo spreads containing pictures of the innocent suspect. Following repeated viewings, the appearance of the suspect begins to seem familiar to the witness. Errors caused by repeated viewing have been attributed to errors in source monitoring, or source confusion. According to the source monitoring hypothesis, memory errors occur when a person attempts to identify where the memory (i.e., of the perpetrator) originated following the receipt of post-event information that leads the witness to erroneously attribute the new information (i.e., the innocent suspect) to the original witnessed event. For this reason, in-court identifications of a defendant provide little real evidence that the defendant actually committed the crime. Familiarity at this point has either been artificially strengthened (when the suspect is the offender) or created by previous viewing of the suspect’s person or photo (when the suspect is not the offender).

### **4.4. Perpetrator Race and Gender**

Although the race of the perpetrator alone generally has little influence on identification accuracy, when the race of the perpetrator and the witness are different, the interaction can impair identification accuracy. Studies of the cross-race effect or the own-race bias have shown that memory for same-race faces is generally superior to memory for faces of another, less familiar race. The effect

has been consistently demonstrated over a 35-year period and has been reliably observed with various ethnic groups (e.g., Europeans, Southern Africans, Americans, Asians, and Hispanics). The cross-race effect is of most significance to the criminal justice system when individuals mistakenly identify a suspect who is not the perpetrator.

Although female faces are generally better recognized than male faces, a similar interaction between the gender of the perpetrator and witness has also been noted. Studies of the own-gender effect have demonstrated that female participants tend to outperform male participants in remembering female faces. Curiously, though, male and female participants do not differ consistently in their ability to remember male faces.

## **5. OBTAINING INFORMATION FROM THE WITNESS**

Eyewitness accounts of criminal events play a vital part in solving crimes and prosecuting criminals, but research has shown that very little of what witnesses actually see ever gets reported to the authorities. One reason for this limited communication has to do with the manner in which information is elicited during police interviews of witnesses. Police interview techniques are often passed on from veteran officers to rookies or learned by trial-and-error on the job. Research has shown that even police officers without formal training tend to share similar interviewing styles, and police interviews of eyewitnesses are often very loosely structured. After requesting open-ended descriptions, officers frequently interrupt witnesses by asking specific, directive questions about the crime, typically pertaining to perpetrator height and weight. These questions usually elicit brief answers from witnesses and disrupt the flow of information. This directive interview style limits the information that witnesses convey and also appears to set up a dynamic between witnesses and officers in which witnesses wait passively for officers to direct the interview. As a result, information not specifically requested by the officer may never get mentioned during the interview, despite its importance to the case.

Interview methods can seriously influence the quality of information given by witnesses in other ways, producing information that is less accurate or even fabricated. A classic example of interviewer influence on witness reports comes from the highly publicized *McMartin* preschool case in which accusations of ritual child sexual abuse were made against seven preschool teachers.

Videotaped interviews revealed that the children were subjected to highly suggestive and leading questions. When these methods were applied in laboratory studies, children's false allegations greatly increased. Experiences such as the McMartin case and studies of actual police interviews reveal the importance of understanding the effects of interviewer influence on eyewitness reports.

### **5.1. The Cognitive Interview**

One interview technique developed by psychologists explicitly for the purpose of enhancing the retrieval of eyewitness memory (and limiting the detrimental effects of interviewer influence) is the cognitive interview. Developed by Geiselman and Fisher in the early 1980s, the cognitive interview consists of four main components: (i) context reinstatement, which includes mentally reinstating the environmental and personal context of the original event; (ii) instruction to "report all" information including partial information, even if it seems unimportant; (iii) recounting the event in a variety of temporal orders; and (iv) reporting the events from a variety of perspectives. A wealth of research has investigated the potential benefits of this technique, both in laboratory and field settings. A meta-analysis of these studies revealed a large increase in the number of correct details elicited by the cognitive interview and a smaller, yet significant, increase in the number of incorrect details elicited. However, the meta-analysis also indicated that accuracy rates (i.e., the proportion of correct details generated) elicited using the cognitive interview were about the same as accuracy rates using traditional interview methods (84% for the cognitive interview and 82% for standard interviews).

In subsequent research, Fisher, Geiselman, and colleagues devised the enhanced cognitive interview, a modified technique that attempted to incorporate strategies such as rapport building to manage the social dynamics of communicating effectively with a witness. These changes resulted in further increases in the amount of correct information obtained, although some studies also observed significant increases in the amount of incorrect details. Additional research is warranted. Training law enforcement officers and others to consistently employ all the aspects of the cognitive interview has proved difficult.

### **5.2. Guided Memory**

Research shows that when various aspects of the cognitive interview are broken down, context reinstatement

instructions appear to play a vital role in the effectiveness of a cognitive interview. Context reinstatement takes place when witnesses are provided with contextual cues about the to-be-remembered event. The basis for the effectiveness of context reinstatement comes from Tulving's encoding specificity principle, which maintains that retrieval cues will enhance memory when the information contained in the retrieval cues matches information contained in the original memory trace. A meta-analysis of facial recognition studies revealed that context reinstatement strategies produce some of the most substantial benefits to identification accuracy.

The guided memory technique was primarily designed to utilize context reinstatement strategies for enhancing memory. This technique involves having witnesses visualize aspects of the event, including physical features and traits of the perpetrator, along with various emotions and reactions elicited by the surrounding event. Although this method has been effective in enhancing correct recognition of perpetrators, little is known about the effects of guided memory on description quality or accuracy. Nevertheless, the guided memory technique, along with other context reinstatement strategies, has been shown to enhance the retrieval of descriptive information from witnesses and offers a promising alternative to current interview techniques.

### **5.3. Generating Descriptions**

Archival research has shown that although descriptions provided by witnesses are frequently vague and lacking in detail, they are generally quite accurate. The descriptors most often reported by witnesses relate to action events. Besides action events, witnesses are called on to provide person descriptions, crime scene descriptions (i.e., objects and environmental details), and accounts of what was said during the crime. With regard to the perpetrator, research indicates that witnesses are much more adept at describing character traits (e.g., he looked like an accountant) and psychological attributes (e.g., he was/looked crazy, kind) than physical aspects of a face, and that such global or holistic judgments improve subsequent attempts at recognition.

### **5.4. Verbal Overshadowing**

Although generating a description of the perpetrator is regarded as a benign activity, studies have indicated that verbally describing a target face may have a detrimental influence on subsequent accuracy at lineup identification. This phenomenon is referred to as verbal overshadowing.

Witnesses who provide a description of a target face perform worse on a subsequent identification task compared to witnesses who did not provide a description. The effect occurs regardless of whether the perpetrator is present or absent from the lineup or whether the lineup is administered sequentially or simultaneously. The negative effect is most pronounced when witnesses are encouraged to adopt a liberal response criterion and to report their memory in great detail.

### **5.5. Postevent (Mis)Information**

Many things can happen to eyewitnesses between the time they give their report to the police and the time they subsequently attempt a lineup or in-court identification. Some intervening events can adversely (or positively) affect their identification accuracy. Eyewitnesses may read newspaper reports about the event they witnessed, they may talk to other eyewitnesses about the same event, or they may be exposed to additional information during the course of the police investigation into the crime. This is known as postevent information. Its (potentially) strongly biasing effects on eyewitness testimony have been extensively investigated and documented by Elizabeth Loftus and colleagues.

### **5.6. Mug Shot Collections and Composite Reproductions**

Following the initial interview with the witness, police may seek to obtain a potential physical likeness of the perpetrator by having the witness attempt to recognize the perpetrator from among a library of criminal mug shots or by having the witness create a composite image based on his or her memory.

#### **5.6.1. Searching Mug Shot Collections**

Mug shot collections are maintained by many law enforcement agencies for the simple reason that criminal recidivism rates are high, and many perpetrators have been photographed by the police on a previous occasion. Researchers have investigated the influence of searching mug shot collections on later identification accuracy. This question has very important legal ramifications since eyewitnesses who have been exposed to a mug shot collection prior to making an identification of a suspect from a lineup may be more susceptible to a source monitoring error, mistaking the intervening mug shot for the perpetrator. Some studies demonstrated that mug shot

exposure can “contaminate” an eyewitness’s memory, but others did not. Legally, the fact that an eyewitness has been subjected to a potentially biasing and avoidable procedure may be enough to discredit that witness.

A second, often neglected, research question concerns the cumbersome manner in which mug shot searches are conducted. Typically, large collections of photographs are stored in albums (printed or electronic), leaving the eyewitness to page through several thousand photographs in a vain attempt to “spot” the perpetrator. Some researchers have tried to improve mug shot search procedures. For example, English researchers devised a description-driven system in the late 1980s called FRAME, which appeared to work quite well but was later criticized for depending on a limited database. For instance, there were very few mug shots of old men in the database; thus, it was comparatively easy to search the database for old men, inflating the accuracy rate. A similar, description-based system has been developed in Canada, and although it appears promising, there is not enough research to draw a definitive conclusion. Researchers in Hong Kong are exploring ways of structuring mug shot searches according to the physical similarity of faces, but initial results have been disappointing.

#### **5.6.2. Making and Utilizing Composite Portraits**

Law enforcement agencies in many parts of the world rely on practical technologies to help crime witnesses reconstruct likenesses of faces. These technologies range from sketch artistry to proprietary computerized composite systems such as Identikit, Photofit, E-Fit, Mac-a-Mug, Faces, and Comp-U-Sketch. Unfortunately, these technologies have not performed well under empirical examination. In most instances, they appear to produce poor-quality composites, which are difficult to match to target faces, even when the target is in full view of the witness. There was some hope that the move to computerized composite software would result in systems containing much larger libraries of features that would be easier to search and graphically “post process” (e.g., allowing the easy addition of specific facial features such as scars or alteration of the aspect ratio of a face). There is no evidence that computerized systems (e.g., Faces and E-Fit) lead to better reconstructions than “manual” systems (e.g., Identikit and Photofit), except when the target is in full view of the witness during the reconstruction, in which case there is a clear advantage for computerized systems.

Recent improvements in computing power have produced more versatile, mathematically sophisticated composite systems. Several groups of researchers are experimenting with “eigenface” systems that potentially allow witnesses to search limitless populations of faces with the assistance of genetic algorithms. Results are promising when witnesses are allowed to reconstruct faces while they are in full view but not from memory. There is evidence that verbal descriptions may be of more assistance to eyewitnesses than visual likenesses or reconstructions.

## 6. THE WITNESS IDENTIFICATION

### 6.1. Measuring Lineup Fairness

There are two key events for an eyewitness, as far as the law is concerned. The first is the event they witnessed, and the second is the identification of the person whom they saw commits the crime. When and how the second event unfolds is just as critical as the first because this identification will constitute the evidence against the defendant. Historically, eyewitnesses were primarily asked to identify perpetrators in court, but this practice is frowned upon nowadays since it is very suggestive (indeed, the witness would look foolish if he or she pointed anywhere but at the defendant). An alternative approach is for police to arrange an encounter between the eyewitness and the accused and to hope for a spontaneous identification (i.e., a “show-up” identification). This slightly less suggestible technique still jeopardizes the liberty of innocent people.

The most widely used alternative to in-court identifications in the United States, and in many other countries, is the “lineup,” in which the suspect is placed alongside a number of men (or women) of reasonably similar physical appearance and demeanor, and the witness is asked to choose the perpetrator if he or she is present in the array. Police lineups originated in English criminal law and procedure, and it is clear that the notion of “fairness” is their *raison d’être*. Lineups are intended to secure an identification that can potentially incriminate the perpetrator, but one that is fair to innocent people who might be suspects. Lineups are not invariably fair in everyday practice since innocent people have been convicted on the basis of eyewitness identification from a lineup. One of the strongest strands of eyewitness research is that investigating all aspects of the fairness of lineups, from their construction to their administration. An array of

methods and measures that can be applied to police lineups has been developed.

In a method known as mock witness evaluation, researchers ask people who did not see the crime event and are “blind” to the identity of the perpetrator (or innocent suspect) to try and identify the suspect in the lineup. This is usually achieved by giving mock witnesses a brief description of the perpetrator (preferably the very description the eyewitness gave to the police) and asking them to indicate the lineup member who best matches this description. If mock witnesses are able to identify the suspect at a rate greater than  $1/k$  (where  $k$  is the number of lineup members), the lineup is said to be biased. The proportion of mock witnesses choosing the suspect is a measure of “lineup bias.”

A second aspect of lineup fairness concerns the number of plausible foils it contains. There are records of police lineups in the United States and other countries where the suspect was the only representative of a particular race or ethnic group in the lineup. In such lineups, the suspect’s identity is immediately suggested to the identifying witness. The number of plausible lineup members is referred to as the effective size and is distinguished from the nominal (or actual) size of the lineup. There are currently two measures of effective size in use: (i) a descriptive measure formulated by Malpass in 1981 and (ii) a closely related inferential measure formulated by Tredoux in 1998. For a lineup to be considered fair, it should receive favorable scores on measures of both lineup bias and lineup size.

The validity of lineup fairness measures has been evaluated in a small number of studies, with positive results. Lineup fairness is frequently assessed by researchers examining police lineups in specific legal cases. A small literature exists regarding proactive use of researchers in constructing lineups that meet fairness standards on a priori grounds. A substantial discussion of lineup fairness appeared in a special edition of the journal *Applied Cognitive Psychology* in 1999.

### 6.2. Lineup Instruction Effects

Some legal and criminal jurisdictions prescribe the instructions, or admonitions, that officers are to provide to eyewitnesses when presenting them with a lineup. The key question is whether these warnings have an effect on the rate of incorrect identification of innocent suspects, which is the main scourge of police lineups. Empirical studies have explored the effect of biased instructions, and several staged-crime experiments have shown that instructions that presuppose the

presence of the criminal in the lineup (“Point out the person who committed the offence”) lead to much higher error rates than instructions that do not (“Note that the criminal may not be present. If the criminal is present, point him out”). The implication is that the instructions given to witnesses should be tempered quite carefully with a warning indicating the perpetrator’s possible absence. Research comparing a set of clearly biased instructions with instructions usually given by the Los Angeles Police Department and a set of “more balanced” instructions showed an increase in mistaken identifications for the biased instructions but no such increase with either of the other two sets of instructions.

### **6.3. Alternative Strategies: Blank Lineups, Sequential Lineups, and Relative Judgments**

That lineups produce high rates of false identification has been known for a long time, and researchers have puzzled over the causes. The cause was long thought to be that witnesses are led to frame their decision task as choosing the offender from among the people in the lineup, thereby selecting the best choice from among the alternatives presented. The witness enters the identification task presuming that the police would not conduct a lineup if they did not have a suspect firmly in mind and thus interprets the task as requiring him or her to identify the lineup member that the police suspect.

A number of additional strategies to reduce erroneous identifications, apart from the lineup instructions and admonitions discussed previously, have been proposed. One solution, suggested approximately 50 years ago by Glanville Williams, the famous English legal authority, is to present the witness with a blank lineup (a lineup known not to contain the offender/suspect) to “trap” people who feel they must make an identification into making a harmless one. Witnesses who choose someone from this blank lineup are clearly mistaken and are dismissed as too unreliable to complete the main identification task. Only witnesses who do not pick from the blank lineup are presented with a suspect-present lineup. Studies have employed blank lineups in staged crime experiments and discovered that witnesses who choose a member of the blank lineup are almost twice as likely to make an incorrect identification from a subsequent perpetrator-present lineup and likewise only about half as likely to correctly identify the perpetrator.

There are a number of problems introduced by such an approach. First, blank lineups are not feasible in practice since the public would quickly learn that a lineup consists of two parts, only the second of which involves the real suspect, rendering the blank lineup ineffectual. Second, it burdens law enforcement with constructing twice as many lineups as they currently construct. (The use of modern computing systems and large mug shot databases may solve this problem in the near future.) Third, law enforcement resists “throwing away” a witness who has not yet been given a chance to view the suspect on grounds that a witness may make a choice from the blank lineup and subsequently, when presented with the suspect-present lineup, enthusiastically identify the suspect, claiming that he or she was mistaken before but is now sure this is the offender. Expecting police officers to forego the potential for this “evidence” might be too much to ask.

Another structural manipulation or alteration that has attracted a great deal of attention from psychological researchers is known as the sequential lineup. Instead of presenting multiple lineups in sequence, or members of one lineup simultaneously (exposing the witness to the suspect and a number of foils at the same time), the individual lineup members are presented one at a time and the witness is required to identify the perpetrator from the members of this sequence. Witnesses are led to believe that there are more photos to be seen than they are actually shown and are instructed to decide if each photo is or is not the perpetrator. This procedure is based on an analysis of the lineup task articulated by Glanville Williams in 1963 (“The witness may . . . be inclined to pick out someone, and that someone will be the one member of the parade who comes closest to his own recollection of the criminal”) and psychologically interpreted by Rod Lindsay and Gary Wells in 1985. They suggested that the structure of the conventional police lineup (all members presented at once) invites witnesses to compare the members of the lineup to one another and identify the person that “best matches” the witness’s memory, leading to a decision based not on the virtual correspondence of the witness’s memory of the perpetrator to one of the lineup members but on the relative correspondence of the lineup members to the witness’s memory, a relative judgment strategy. Presenting lineup photographs sequentially is said to inhibit eyewitnesses from relying on relative judgments and, instead, fosters absolute judgments made by comparing their memory of the perpetrator with each lineup member. Empirical studies have demonstrated that the use of sequential lineups can reduce false identifications and increase correct rejections of lineups that do not

contain the perpetrator, although a significant reduction in the rate of correct identifications in lineups that do contain the perpetrator has also been noted.

The apparent advantages of sequential lineups has made them attractive to law enforcement authorities in some jurisdictions, and they have been promoted very actively to law enforcement by some members of the eyewitness research community. Research on sequential lineups has been very important, but there is also controversy over whether the research has been extensive and careful enough to warrant promotion of sequential lineups as a replacement for traditional police lineups.

#### **6.4. Lineups Involving Other Sensory Modes**

The lineup is usually a test of visual memory, but most legal systems also recognize that additional cues to identity may reside in the voice, mannerisms, and gait of the perpetrator.

##### **6.4.1. Earwitness Identification**

The case of voice identification has attracted much attention in the research literature. The most pertinent finding has been that familiarity of the voice plays a key role in identification, and that identification of unfamiliar voices is prone to yield alarmingly high false-alarm rates. Unfamiliar voices are very difficult to recognize and especially easy to confuse. Similar to the own-race bias in facial identification, several studies have observed that voices of another, less familiar race are also more difficult to recognize. In addition, a witness's memory for the voice of the perpetrator is equally susceptible to the many estimator variables discussed previously, including opportunity to hear the voice clearly and various factors that may draw the witness's focus of attention away from the auditory stimulus.

##### **6.4.2. Multimode Lineups**

Lineups typically used in Western countries restrict the presentation of the lineup to one mode, most frequently involving visual presentation. However, cues from a single (sensory) domain may be degraded with respect to the original context (e.g., physical disguise), and it is easier to identify the perpetrator with the assistance of cues from other domains. Researchers have begun to explore the utility of multimode lineups on witness identification performance. Although there is little

published research, the idea of allowing witnesses access to cues that are not solely visual in nature is already in practice in Sweden. The lineup is formed in a room in which lineup members are allowed to sit down, smoke, communicate, and behave in general as they ordinarily would over an extended period of time. These individuals are then viewed by the witness from behind a one-way mirror, and members of the lineup are not aware when the witness is present. Unfortunately, no systematic evaluation of this strategy has been performed.

##### **6.4.3. Live vs Photographic Lineups**

Lineups currently differ within single sensory modes, such as the practice of eliciting identifications from photographic lineups as an alternative to corporeal (live person) lineups. In many areas of the United States, for example, photographic lineups are favored over corporeal lineups. In many foreign jurisdictions, however, identifications from corporeal lineups are required if eyewitness identification evidence is to be presented in court. Intuitively, corporeal lineups would seem to have an advantage: The detail to be obtained from a live inspection of the suspect must surely exceed the detail present in a photograph. However, the consensus from several studies is that there is either very little difference between the capacity of photographic and corporeal lineups to elicit identifications or no difference at all. An extensive review of the literature reported no notable difference between live, video, or photographic lineups. There is no evidence to suggest that media that embellish cues will aid identification accuracy or reduce false alarms.

## **7. THE PROSECUTION**

After the police obtain a positive identification from the eyewitness that their suspect is indeed the perpetrator, they are ready to hand the case over for prosecution. The district attorney (or attorney general or director of public prosecutions) may decide not to prosecute, but in most jurisdictions eyewitness identification is considered strong evidence against the perpetrator and the case will proceed to court. A number of factors in the court process can influence an eyewitness's accuracy and can change the impact that an eyewitness's testimony will have at trial. For instance, the prosecution (or defense) can pose leading questions to the witness, eliciting false recollections, or an

expert can be called to testify on the vagaries of eyewitness identification. In the following sections, some of the research on two potentially important factors is summarized. First, we briefly discuss a debate that concerns the role and value of expert testimony by eyewitness researchers in legal trials. Second, we discuss research that has examined the effectiveness of jurors in assessing the reliability of eyewitness testimony at trial and whether juries benefit from expert assistance on eyewitness memory.

### **7.1. Expert Testimony**

Many researchers have offered expert testimony to the courts in cases in which eyewitness evidence has been at issue, although this undertaking has been somewhat controversial within the discipline of psychology. A variety of arguments have surfaced on this issue, including disputes over the consistency of findings and the adequacy of current knowledge for presentation to the courts. Several surveys of researchers and experts in the eyewitness literature by Saul Kassin and colleagues have attempted to address these concerns by identifying areas of agreement. In 1989, this survey revealed that the following areas were considered “reliable” by the majority of respondents: limited exposure time, lineup instruction effects, prior expectations on the part of the witness, postevent misinformation, and the weak relationship between confidence and accuracy. An update to this survey of experts in 2001 demonstrated some additions to this listing that included confidence malleability, mug shot-induced biases, child witness suggestibility, the effects of alcohol intoxication, the cross-race effect, the weapon focus effect, forgetting over time, and lineup presentation manipulations (simultaneous vs sequential lineups). Given the developing nature of the science of eyewitness identification, it is reasonable to assume some shifts in experts’ opinions regarding various phenomena. Nevertheless, Kassin and colleagues noted the remarkable consistency in ratings across the two surveys for the majority of topics covered.

### **7.2. Jurors’ “Common Knowledge” of Problems with Eyewitness Identification**

One basis for rejecting an eyewitness expert’s testimony in court is the belief that factors affecting the reliability of eyewitness identification are common knowledge to the

lay juror. Researchers have sought to determine exactly what laypersons know about factors affecting eyewitnesses. Three basic methodologies have been used to investigate this information: (i) surveying jury-eligible citizens as to their knowledge and beliefs, (ii) assessing jurors’ ability to predict the outcome in an eyewitness identification experiment, and (iii) using mock trials to assess the influence of trial techniques (such as cross-examination of the eyewitness or the presence/absence of expert evidence on eyewitness memory).

First, survey studies have been conducted by administering questionnaires, such as the Knowledge of Eyewitness Behavior Questionnaire, in order to assess beliefs about factors that affect the accuracy and reliability of eyewitness identification. The results of such studies have demonstrated that respondents appeared insensitive to such effects as the age of the witness (young or old) and the retention interval prior to identification, and that respondents tended to believe (contrary to research findings) that training could improve identification accuracy. Second, studies have asked participants to read written summaries of identification experiments and then to “postdict” the accuracy of participant-witnesses. Results in such studies have indicated that participants overestimate accuracy rates, suggesting that individuals often believe witnesses to be much more accurate in their judgments than they truly are. Finally, research employing the “mock trial” as a method for assessing jurors’ common-sense knowledge has manipulated different factors known to influence, or not to influence, identification accuracy and then assessed whether participant-jurors are sensitive to these factors in their verdicts. A common outcome is that witness confidence is given too much weight, with jurors believing confident witnesses to be more accurate, which may not necessarily be the case. In contrast, jurors failed to assign sufficient weight to other factors, such as opportunity to observe the perpetrator or the presence of a weapon, which should have been considered. Taken together, this research appears to demonstrate that lay jurors lack the requisite knowledge to appropriately evaluate eyewitness identification evidence, and that expert witnesses may be useful for providing such information at trial. Several studies have also attempted to assess the benefits of expert testimony within the mock trial paradigm. Their results have generally shown that exposure to an eyewitness expert leads jurors to more appropriately weight identification evidence based on factors known to influence the reliability of an identification.

## 8. CONCLUSION

In a 1978 paper that has been extraordinarily influential in the eyewitness literature, Gary Wells argued that the quintessence of eyewitness research is its application to a legal problem and advised that it is incumbent on eyewitness researchers to show the practical utility of their research. This approach has indeed yielded great rewards: One of the most important pieces in the eyewitness literature is the coconstruction of guidelines by psychologists, lawyers, and criminal justice professionals for collecting evidence from eyewitnesses, sponsored by the National Institute for Justice.

There is a cost, though, to the near-exclusive applied research focus taken by eyewitness researchers. An applied focus typically leads to research that is interested in improving practice (e.g., lowering the rate of false-positive identifications) but is less concerned with explaining how and why a particular intervention works. For example, we know that the sequential lineup returns a lower rate of false positives than the traditional police lineup, but we do not have a good answer for why this is the case. In general, eyewitness research has not appreciated the value of theoretical explanations and models. This is not surprising given the explicit applied orientation of researchers in the area, but it may now be time for researchers to redress the imbalance. A key reason for promoting such a redress is the belief that the applications that stem from applied eyewitness research could be more powerful if they were based on an explanatory model. If we understood the mechanism, for instance, that makes conservative instructions reduce the false-positive rate, then we might be able to design alternative forms of lineup that maximally reduce this rate.

We do not wish to propose that eyewitness research jettison its applied orientation: On the contrary, we wish to strengthen that orientation. The question is how this should be done. Our contention is that a return to the laboratory might, paradoxically, create a stronger applied foundation for the discipline.

### See Also the Following Articles

Forensic Mental Health Assessment ■ Interrogation and Interviewing

## Further Reading

- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing*. Springfield, IL: Thomas.
- Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist*, *56*, 405–416.
- Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law & Human Behavior*, *5*, 299–309.
- Malpass, R. S., & Devine, P. G. (1980). Realism and eyewitness identification research. *Law & Human Behavior*, *4*, 347–358.
- Malpass, R. S., & Lindsay, R. (1999). Measuring lineup fairness. *Applied Cognitive Psychology*, *13*, S1–S7.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, & Law*, *7*, 3–35.
- Ross, D., Read, J. D., & Toglia, M. (1994). *Adult eyewitness testimony: Current trends and developments*. London: Cambridge University Press.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*, 139–156.
- Sporer, S. L., Malpass, R. S., & Koehnken, G. (1996). *Psychological issues in eyewitness identification*. Hillsdale, NJ: Erlbaum.
- Stebly, N. M., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law & Human Behavior*, *25*, 459–473.
- Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement* [Booklet]. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.
- Tredoux, C. (1998). Statistical inference on measures of lineup fairness. *Law & Human Behavior*, *22*, 217–237.
- Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology*, *36*, 1546–1557.
- Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 277–295.
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law & Human Behavior*, *22*, 603–647.



# Eyewitness Testimony

Linda J. Levine and Elizabeth F. Loftus

University of California, Irvine, California, USA

1. Eyewitness Errors and Wrongful Convictions
  2. The Malleability of Memory
  3. The Misinformation Effect
  4. Creating New Memories
  5. Confidence and Memory Accuracy
  6. Emotion and Memory
  7. Facilitating Accurate Testimony
- Further Reading

## GLOSSARY

**amygdala** A brain structure, located in the anterior part of the temporal lobe, that modulates the strength of emotional memories.

**flashbulb memory** An unusually vivid and detailed memory of a surprising and consequential event.

**misinformation effect** The tendency for erroneous postevent information to lead to memory distortions.

**postevent information** Information introduced after the experience of an event.

**source amnesia** The inability to recall the origin of the information that comes to mind when attempting to recall an event.

**weapon focus** Witnesses' tendency to focus on and remember the weapons in crime scenes, often at the expense of memory for other information.

Law enforcement officials rely heavily on eyewitnesses to identify the perpetrators of crimes and to make decisions about exactly what happened at the times of

legally relevant events (e.g., what led to an accident). Eyewitness testimony also has a major impact on the verdicts returned by juries. Although eyewitnesses are often correct, and in some cases their testimony is the only evidence available, mistakes do occur and can lead to the wrong people being convicted and to the actual perpetrators of crimes remaining at large. How do these mistakes happen? Far from being a complete and objective record of past events, memory is malleable. Leading questions, differing recollections of others, and vividly imagining events at a later time can lead to memory errors and even to the creation of "memories" anew. Scientific research on the causes of eyewitness testimony errors has been used to develop recommendations for preventing errors and reducing the rates of wrongful convictions.

## 1. EYEWITNESS ERRORS AND WRONGFUL CONVICTIONS

It is estimated that as many as 7500 people in the United States are wrongfully convicted of serious crimes each year. One of the key factors leading to wrongful convictions is eyewitness error. Indeed, eyewitness misidentifications are responsible for more wrongful convictions than are all other causes combined. In 1996, Huff and colleagues analyzed factors that led to 205 cases of proven wrongful convictions of innocent people. Most cases resulted from a combination of errors, but slightly more than

half of all cases examined were convicted on the basis of eyewitness misidentification. In a report published the same year by the U.S. Department of Justice, Connors and colleagues analyzed 28 cases of wrongful conviction with DNA exoneration. They found that 80% of these innocent people had been convicted due to faulty eyewitness memory. These findings highlight the risks of uncritical reliance on eyewitness reports.

## **2. THE MALLEABILITY OF MEMORY**

Memories for past events can change over time. After people first witness an event such as a crime or an accident, new information can become incorporated into memory. People may mingle memories of different events and fill in gaps in their memories based on more recent information. "Source amnesia," or the inability to recall the origin of a memory of a given event, is an important contributor to memory malleability. Over time, people forget whether information that comes to mind when attempting to recall an event originated from actual past experience, from subsequent suggestions by others, or from vividly imagining an event. Once the source of the information has been forgotten, people can confuse suggested and imagined events with the true event. Because the malleability of memory has important implications for evaluating the reliability of eyewitness testimony in court proceedings, psychologists have examined the conditions under which erroneous information is likely to be mistaken for witnessed events.

## **3. THE MISINFORMATION EFFECT**

Research conducted by Loftus indicates that "postevent information," or information introduced after the experience of an event, can lead to errors in people's memories for witnessed events. Even subtle differences in the way in which a question is asked (e.g., changing from the indefinite article "a" to the definite article "the") can lead to distortions in memory. This was shown in a study where participants first viewed a brief film of a traffic accident. Afterward, they were asked a number of questions that were phrased in one of two forms: "Did you see the broken headlight?" or "Did you see a broken headlight?" Participants who were asked the question including the word "the" were more likely to report seeing a nonexistent broken headlight than were participants who were asked the question using the word "a".

In another study, people viewed slides of an automobile collision. Later, they were asked one of two questions: "How fast were the cars going when they hit each other?" or "How fast were the cars going when they smashed into each other?" The verbs, "collided," "bumped," and "contacted" were also used. Participants' speed estimates varied considerably depending on how the question was phrased. For example, estimates of the cars' speed averaged 40.8 miles per hour when the verb "smashed into" was used, whereas they averaged only 31.8 miles per hour when the verb "contacted" was used.

Memory errors of this type become more frequent as the gap of time increases between the occurrence of an event and the reporting of it. With increasing time, details of the actual event are forgotten or become less accessible, and the fragmentary memory tends to be filled in with information that is currently available. This information may be drawn from schematic knowledge, or knowledge of the ways in which events typically occur, from suggestions by others and even from the phrasing of a question (e.g., use of the verb "smashed" instead of "contacted"). Thus, the way in which a police officer or attorney questions a witness can influence what the witness recalls.

More broadly, scientific research shows that a person does not simply experience an event, file it away in memory, and then retrieve it whole at some later date and read off what was stored; instead, at the time of recall, past events are partially reconstructed using information from many sources. These sources include the person's original perception of the event and inferences drawn later after the fact. Over a period of time, bits of information from these sources become integrated, so that the witness becomes unable to say how he or she knows a specific detail. Only the unified reconstructed memory remains. Memories, then, are neither wholly accurate nor wholly distorted. They typically incorporate both memory traces from the experience of the actual event and postevent information.

## **4. CREATING NEW MEMORIES**

Not only can memories be altered, they can be created anew. A simple reliable method for creating false memories for words was developed by Deese, Roediger, and McDermott and is referred to as the DRM paradigm. After studying a list of related words (e.g., rest, bed, night, snooze), people are asked whether or not particular words were on the list. They often claim to

clearly remember seeing a word that was not presented but that was closely associated with other words on the list (e.g., sleep). Hancock and colleagues' 2003 research demonstrated that these phantom words can be even more accessible in memory than words that actually appeared on the list.

Memories of complete events that never occurred also can be created in the minds of normal adults by enlisting the help of their mothers, fathers, and other older relatives. In a 1995 study, Loftus and Pickrell told participants that their relatives had provided researchers with descriptions of events that had occurred during their childhoods. In fact, three of the events had actually occurred, but one was a fabricated account of being lost in a shopping mall at 5 or 6 years of age for an extended time and ultimately being rescued by an elderly person and reunited with the family. Approximately a quarter of the participants fell sway to the suggestions and described memories, either complete or partial, of having been lost in this specific way. Using the same procedure in 2001, Heaps and Nash found that 37% of the participants fell sway to the suggestion that they were in trouble in the water—nearly drowning—and had to be rescued by a lifeguard.

Suggesting to participants that relatives had witnessed an event is a strong form of suggestion, but distortions in memory can occur even with weaker suggestions such as guided imagination. This was demonstrated in Garry and colleagues' 1996 study in which participants first rated how confident they were that a number of childhood events had happened to them before 10 years of age (e.g., "broke a window with your hand"). Later, some participants were told the following: "Imagine that it's after school and you are playing in the house. You hear a strange noise outside, so you run to the window to see what made the noise. As you are running, your feet catch on something and you trip and fall." The script went on to guide participants through a detailed scenario in which they would break the window with their hand and get cut and bloody. During the final phase of the study, they once again answered questions about their childhood experiences. The results showed that the 1-minute act of imagination led a significant minority of participants to claim that an event was more likely to have happened (relative to a control group of participants who were not asked to imagine the event), even though they had previously said the event was unlikely to have occurred.

Imagination has also been shown to influence memory for complex events such as witnessing a crime. As

with the misinformation effect, the impact of imagining an event becomes more pronounced over time. Imagining an event increases its familiarity. As both the event and the act of imagination recede into the past, people become less able to distinguish whether the source of the event's familiarity is that they imagined it or that it actually occurred.

## 5. CONFIDENCE AND MEMORY ACCURACY

Jurors often understand that memory is fallible and look for cues as to whether an eyewitness's testimony can be believed. One commonly used indication of memory accuracy is confidence. Juries find confident witnesses to be far more believable than those who admit that they could be mistaken. Unfortunately, the assumption that confidence is a clear sign of accuracy is unfounded. Bothwell and colleagues' 1987 review of 35 studies involving staged incidents showed that the average correlation between witness confidence and accurate identification of the perpetrator was modest (the average correlation coefficient was  $r = .25$ ). Thus, with increasing confidence, only a slight increase in memory accuracy was found.

Further weakening its predictive power, witness confidence is highly malleable. Witnesses can become more or less confident about the accuracy of their memories for an event or individual as a result of events that happen afterward. Confidence increases when witnesses describe past events repeatedly or receive confirming feedback such as being told that another witness identified the same individual as the culprit. Confidence is more strongly predictive of recall accuracy when viewing and reporting conditions are optimal, for example, when a witness had a long time to observe a perpetrator's face and was not exposed to repeated or suggestive questioning.

## 6. EMOTION AND MEMORY

Witnessing a crime, and certainly being the victim of a crime, is a highly emotional event. How does intense emotion affect the accuracy and completeness of eyewitness testimony? In comparison with mundane events, memories for emotional events tend to be long-lasting, vivid, and detailed. When asked to describe how they first learned about highly emotional events (e.g., the terrorist attacks on the World Trade

Center and U.S. Pentagon on September 11, 2001), people can typically report, in considerable detail, where they were, what was happening at the time, who told them the news, and how they felt. In their classic 1977 article on flashbulb memories, Brown and Kulik argued that because of the obvious survival value, there may be some mechanism in the brain that leads to remembering biologically crucial but unexpected events with close to photographic accuracy.

More recently, animal and brain imaging studies have been conducted to investigate the biological underpinnings of the effects of emotion on memory. The results indicate that the amygdala, an almond-shaped structure located in the anterior part of the temporal lobe, modulates the strength of emotional memories stored elsewhere in the brain. The greater the level of amygdala activation, the more enduring the emotional memory. For example, in Canli and colleagues' 2000 study, participants viewed negative and neutral pictures. Functional magnetic resonance imaging (fMRI) was used to assess the level of activation in the amygdala as participants rated each picture for emotional intensity. The more emotionally intense participants found the pictures to be, the more activation was found in the amygdala. Three weeks later, participants were given a surprise recognition test with previously seen and new pictures. They were asked to indicate whether they had seen each picture before and, if so, whether it seemed familiar or whether they clearly remembered seeing it. For those pictures rated as emotionally intense, the degree of amygdala activation predicted whether the picture would be forgotten, would appear to be familiar, or would be clearly remembered.

Emotional memories, then, are more lasting and vivid than nonemotional memories, but this does not mean that they are error free. Indeed, Neisser and Harsch's 1992 research showed that inaccuracies were the norm when college students recounted how they found out about the explosion of the *Challenger* space shuttle after a delay of 2½ years. Furthermore, emotional arousal might not enhance memory for all aspects of emotional events. Several laboratory studies have shown that, in comparison with their memory for neutral events, people showed enhanced memory for information that is central to emotional events but showed poorer memory for the peripheral details. In real-world settings, a similar phenomenon has been termed "weapon focus." This refers to witnesses' tendency to focus on and remember the weapon in crime scenes where a knife or gun is present, often at the expense of memory for other information such as the culprit's face.

In summary, emotional arousal improves memory for central or gist information but can disrupt memory for details that are peripheral to a witness's current concerns (e.g., survival). Investigators are currently working to define more precisely the types of information that are "central" to people experiencing intense emotion and whether this information varies depending on the particular emotional state being experienced (e.g., fear vs anger). This issue is important because, in the case of eyewitness testimony, memory for details (e.g., a license plate number, the color of a jacket) can be telling.

What about memory for emotions themselves? Victims of crimes or accidents due to negligence may seek damages for the suffering they endured, but how accurately can people remember their own past emotional reactions? Levine and Safer's 2002 review of recent research showed that memories for past emotions, like memories for past events, can be biased by postevent information. In one study, Levine and colleagues asked people to recall how they felt when they first learned that O. J. Simpson had been acquitted of murder. Participants rated the intensity of their emotions shortly after the verdict was announced. Two months later, and then again after more than a year, they recalled how they had felt. At each time point, they also described how certain they were of Simpson's innocence or guilt. The results showed that the more people's beliefs about Simpson's guilt changed over time, the less stable their memories were for how happy or angry they felt when Simpson was acquitted. For example, those who came to believe that Simpson was guilty overestimated their feelings of anger, and underestimated their feelings of happiness, relative to their initial reports.

In another study, Safer and Levine had undergraduates recall how anxious they had felt before a major exam. Students who learned that they had done well on the exam underestimated how anxious they had felt before the exam (in comparison with a control group of students who had not yet learned their exam grades), whereas students who learned that they had done poorly overestimated how anxious they had felt before the exam. Thus, postevent information about their grades led to distortions in students' memories for their past feelings of anxiety.

Like memories of events, then, people's memories of their emotions are partially reconstructed; that is, they change subtly over time and become more consistent with current beliefs about the emotion-eliciting event. Put differently, people are constantly learning. Memories of past events are not sealed off from this subsequent learning; instead, they are informed by it.

The malleability of memory may be adaptive in some circumstances, but it makes it necessary to use caution when evaluating eyewitness testimony, whether that testimony concerns a neutral event (e.g., a traffic violation), an emotional event (e.g., a crime), or even a person's own past emotional reaction.

## 7. FACILITATING ACCURATE TESTIMONY

Scientists investigating the factors that lead to inaccurate memories have developed recommendations to ensure that witnesses provide as complete and accurate testimony as possible. Careful questioning of witnesses is essential for minimizing errors in eyewitness testimony. Using verbal or nonverbal cues, an interviewer can unwittingly influence witnesses toward identifying the person who the interviewer believes is the culprit and toward remembering events in the way in which the interviewer believes they occurred. Therefore, experts recommend that witnesses be interviewed by individuals who have no preconceptions concerning the facts surrounding the case.

Witnesses can be biased simply by limiting the answer options with which they are provided. Thus, interviewers are advised to phrase questions in a clear and open-ended manner that avoids suggesting that a particular response is the correct one (e.g., "They drove away in the white van, didn't they?"). Attention also should be paid to the wording of questions because the use of one term can produce a different response from the use of another, even when the two terms have similar meanings (e.g., "smashed" vs "hit").

Based on a detailed review of scientific evidence concerning eyewitness identification errors, Wells and colleagues in 1998 developed recommendations regarding the best procedures for constructing suspect lineups and photo spreads. When eyewitnesses are asked to identify a culprit from a lineup or photo spread, they tend to identify the person or photo most resembling the culprit they remember relative to the other members of the lineup. If the culprit is not present in the lineup or photo spread, this relative judgment process often leads to a false identification. Wells and colleagues recommended that eyewitnesses be informed that there is a good possibility that the perpetrator is not present in the lineup or photo spread. This makes it less likely that an innocent individual will be selected simply because he or she resembles the criminal. They also recommended that suspect lineups and photo spreads be

composed of individuals who fit the basic recollection of the witness. For instance, if the perpetrator was of a certain race, having only one person of that race present in the lineup or photo spread makes that person more likely to be selected regardless of whether he or she committed the crime. Because people recall the same events differently, it may be necessary to have individual witnesses view different lineups composed of suspects who resemble each witness's description.

Witnesses' confidence in their recollections can be influenced by subsequent events and suggestions. Therefore, experts recommend that confidence be assessed and recorded at the time a perpetrator is first identified or an event is first described. Interviewers should avoid making comments to witnesses regarding their accuracy afterward because this may inflate confidence in witnesses' memories and affect the quality of testimony at later stages.

Inaccurate eyewitness testimony can lead to a double tragedy: an innocent person being punished and the actual perpetrator remaining free to victimize others. Although memory errors can never be completely eliminated, adopting identification and interview procedures that are informed by scientific evidence will minimize inaccuracies and help to safeguard the justness of the criminal justice system.

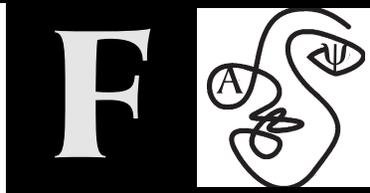
### See Also the Following Articles

Child Custody ■ Child Testimony ■ Eyewitness Identification ■ Psychology and the Law, Overview

### Further Reading

- Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*, 691–695.
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73–99.
- Canli, T., Zhao, Z., Brewer, J., Gabrieli, J., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience, 20*(RC99), 1–5.
- Connors, E., Lundregan, T., Miller, N., & McEwan, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial*. Alexandria, VA: National Institute of Justice.
- Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*, 208–214.

- Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation of critical lures in the Deese–Roediger–McDermott paradigm. *American Journal of Psychology, 116*, 1–14.
- Heaps, C. M., & Nash, M. (2001). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 920–930.
- Huff, C. R., Rattner, A., & Sagarin, E. (1996). *Convicted but innocent: Wrongful conviction and public policy*. Thousand Oaks, CA: Sage.
- Levine, L. J., & Burgess, S. L. (1997). Beyond general arousal: Effects of specific emotions on memory. *Social Cognition, 15*, 157–181.
- Levine, L. J., & Safer, M. A. (2002). Sources of bias in memory for emotions. *Current Directions in Psychological Science, 11*, 169–173.
- Loftus, E. F. (1996). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- Loftus, E. F. (2001). Imagining the past. *The Psychologist, 14*, 584–587.
- Loftus, E. F., & Pickrell, J. (1995). The formation of false memories. *Psychiatric Annals, 25*, 720–725.
- Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. In E. Winograd, & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of “flashbulb” memories*, Vol. 4: *Emory Symposia in Cognition* (pp. 9–31). Cambridge, UK: Cambridge University Press.
- Reisberg, D., & Heuer, F. (2004). Remembering emotional events. In D. Reisberg, & P. Hertel (Eds.), *Memory and emotion* (pp. 3–41). New York: Oxford University Press.
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 23*, 603–647.



# Fair Treatment and Discrimination in Sport

Vikki Krane

Bowling Green State University, Bowling Green, Ohio, United States

1. Introduction
2. Social Prejudice in Sport
3. Celebrating Difference through Sport
4. Social Equity in Sport
5. Applied Sport Psychology and Social Justice  
Further Reading

## GLOSSARY

**classism** Differential treatment based on social class or economic status; the privileging of the wealthy social group so that that group has greater power and status than other social class groups.

**equality** Identical treatment across social groups.

**equity** Fair treatment, but not identical treatment, across social groups; equal access to necessary opportunities and resources.

**ethnicity** A social category characterized by cultural heritage, including norms, values, and customs.

**ethnocentrism** Differential treatment based on ethnicity; the privileging of one ethnic group so that the group has greater power and status than other ethnic groups.

**heterosexism** Differential treatment based on sexual orientation that privileges heterosexuals over other sexual orientations; stigmatizes behaviors, identities, and relationships that are not heterosexual.

**nationalism** Differential treatment based on national origin or country of residence; the privileging of people of one

nation, or group of nations, so that the group has greater power and status than people of other nations.

**race** A social category characterized by biological or genetically transmitted traits.

**racism** Differential treatment based on race; the privileging of one racial group so that the group has greater power and status than other racial groups.

**sexism** Differential treatment based on gender; the privileging of one gender so that the group has greater power and status than other ethnic groups.

**social class** A social category based on economic position in society, including income, wealth, education, occupation, and social connections.

**social justice** Social tolerance ensures that people of all social groups have the necessary resources and opportunities.

**social prejudice** Systematic mistreatment, bias, or bigotry toward people based on social group membership.

**stereotypes** Characteristics perceived to be shared by members of a social group, without consideration for individual differences.

Fair treatment in sport occurs when there are equitable resources and opportunities for all participants, regardless of social group membership (e.g., gender, race, ethnicity, nationality, sexual orientation, and social class). Sport offers the opportunity to traverse social identities, but it also can be divisive. In some cases, differential treatment occurs in which minority social group members are treated unfairly. In other

situations, sport is an avenue to educate people and increase awareness about a wide range of social issues, fight social injustice, and provide humanitarian assistance. Applied sport psychologists may employ a variety of strategies to promote fair and equitable sport.

## **1. INTRODUCTION**

Sport often is considered a microcosm of society: Common practices in everyday life are replicated in sport environments. Issues such as equitable or non-equitable conduct found in society generally are reproduced in sport settings. Sometimes sport seems to traverse racial, ethnic, social, national, and gender boundaries; sport has the power to bring diverse people together while deemphasizing social or cultural differences. International sport is predicated on the ideal of achieving peace and friendship among world peoples. It provides a common language, connects people worldwide, and offers an avenue for cultural exchange and enhanced understanding across national boundaries. This is seen in events such as the Olympics, Football World Cup, Tour de France, and the Commonwealth Games. Athletes who may be from divergent cultural groups will cooperatively engage in sport. This has occurred with such disparate groups as Iranians and Iraqis or Arabs and Israelis. In fact, competitions have been created specifically to mend political strife. The Goodwill Games, for instance, were constructed to bring together the United States and former Soviet Block nations after the boycotts of the 1980 and 1984 Olympic Games.

Although sport offers the opportunity to traverse cultural identities and unite different peoples, it also can have the opposite effect. Cultural differences may be exacerbated through sporting events, which in turn become propaganda for the victorious side. Sport reproduces dominant ideologies and is accessible disproportionately to members of privileged social groups. Obviously, these matters negatively affect sport experiences.

## **2. SOCIAL PREJUDICE IN SPORT**

People naturally group like individuals into social groups; this categorization is a perceptual process that allows individuals to make sense of the vast social world. Stereotypes emerge as people identify characteristics perceived as common among the members of a social group. Sometimes these common characteristics

are benign (e.g., basketball players are tall). However, some stereotypes may be incorrect (e.g., black people cannot swim) or even malicious (e.g., Islamic men are violent or terrorists). Prejudice occurs when stereotypes lead people to have negative attitudes toward all members of certain social groups, whereas discrimination occurs when people behave negatively toward members of a social group.

Unfair treatment in sport, as in society at large, often arises from prejudices such as racism, ethnocentrism, sexism, heterosexism, classism, and nationalism. These terms reflect various forms of social exclusion based on socially constructed categories or identities (e.g., gender, race, nationality, sexual orientation, and ethnicity). Identities are socially constructed as social norms and expectations create and reinforce attitudes toward individuals based on social categories. For example, beliefs about how males and females should behave and look are based on social customs. These customs create expectations for males and females that affect actual behavior, values, and beliefs. Social beliefs and social norms often lead to stereotypes about certain groups of people, which reflect societal values and convey acceptable and unacceptable behaviors and values. They also deprive people of their individuality and limit opportunities.

Stereotypes and negative attitudes are the major roadblocks to sporting achievement, not lack of ability. Innately, there is no reason that people in various social groups should have any less athletic talent than individuals in majority social groups, but examples of disproportionate accomplishments abound. For example, far fewer African, Latino/a, or Asian athletes are represented on Canadian, UK, or Australian professional sport teams compared to white athletes. Stereotypes such as “black players are not as smart as white players” and “Asians lack athletic ability” impede their progress in sport. Beliefs that “women are not interested in sport” or that “Aboriginal people are not as ‘civilized’ or cognitively developed as white people” reduce their opportunities in mainstream (i.e., majority-ruled) sport.

Because these athletes are not expected to succeed, it is likely that coaches will treat them differently than athletes expected to be more skilled. The athletes for whom coaches have low expectations likely receive less constructive feedback, less praise, fewer practice opportunities, fewer chances to correct mistakes, and more criticism. This feedback, in turn, may decrease confidence and increase anxiety or may not provide learning opportunities for certain groups of athletes. Thus, these athletes do not attain the same skill level as

athletes from majority social groups, reinforcing the correctness of the stereotype. The concomitant lack of successful role models and inability or unwillingness of sport clubs to support minority athletes further minimize their chances of excelling in sport. Additionally, individuals with minority identities are often socially or economically disadvantaged and therefore do not have the resources to develop their athletic abilities.

Cultural and prejudicial attitudes evolve and change over time; therefore, they are socially and historically bound. For example, although today black players appear to be represented proportionately in British football, Asian players are not. However, not long ago, black players were struggling for representation. As social values change, so do attitudes toward various social groups in sport. Judgments about specific social groups also may differ across cultures. French fans, for instance, backed openly lesbian, professional tennis player Amelie Mauresmo, whereas in other countries she would not be supported. There is not a single way to explain prejudice or inequity in sport because it is played out in a myriad of ways given the dominant culture where it occurs. The underlying commonality, though, is that opportunity, status, and power are disproportionately afforded consistent with the social hierarchy supporting the dominant social group.

### 2.1. Racism

Inequitable treatment due to racial or ethnic identities occurs across sporting environments. Racism affects university, professional, and elite sports, from youth to masters' levels. It has been documented in sport worldwide. Incidents of racist fan behavior at World Cup or international rugby events, for example, are not uncommon. Racist actions in sport result in fewer racial minority athletes engaging in particular sports (e.g., swimming and golf), minimal numbers of athletes of color competing at elite levels of sport, few individuals of color as coaches or managers of athletic programs, overt prejudicial behaviors toward racial minority athletes, lack of fan support for minority athletes, and limited opportunities to learn sports. Conflated with these indignities, female athletes of color face additional barriers, as do disabled, indigenous, or lesbian, gay male, bisexual, or transgendered (LGBT) athletes of color.

A common racist action in sport is stacking, which occurs when a certain social group of athletes is funneled into peripheral positions on athletic teams. For example, in Western team sports, white athletes historically held the leadership and central positions.

(White) coaches believed that black athletes were not smart enough to take on the responsibility of this position. Rather, black athletes were guided into positions requiring speed and jumping ability (e.g., pass receivers). This practice led coaches to encourage young black players to gain skills in peripheral positions if they wanted an opportunity to participate in higher levels of the sport. It also encouraged self-selection into positions where there were black role models. Stacking ultimately limited the number of athletes of color on a team because these players competed for only a few positions. This practice occurs across a wide array of sports.

Although the overt practice of stacking has declined in many sports, lack of minority representation in sport administration and coaching is widespread. Because these are the powerful, decision-making positions, minority athletes have little representation in how sport is conducted or who is selected for teams or management positions. Additionally, sport has become a big business. The owners, managers, and organizers in the business of sport have become quite wealthy and powerful, allowing them to wield considerable power concerning how sport is run and who can be involved. Few men of color rise to these positions and almost no minority females are sport team owners or managers.

### 2.2. Sexism

Research on gender in sport consistently documents fewer opportunities and less support for female sport participants compared to males. International competitions, such as the Olympics, have more events for male athletes; governing bodies are male dominated so women have little voice in how sport is managed; the media reinforces gendered ideologies and provides limited coverage of female athletes; sexual harassment toward female athletes is widespread; and females in sport must contend with stereotyping and "accusations" of being lesbians. These actions have led to sport being considered a patriarchal institution; that is, it is dominated by males who hold social, political, and economic power.

Sport also reinforces and supports hegemonic masculinity and femininity. These ideals reinforce socially sanctioned mannerisms specific for males or females. Most societies encourage males to be masculine and females to be feminine. Characteristics associated with hegemonic masculinity include strength, competitiveness, assertiveness, confidence, and independence. Being gentle, compassionate, emotional, and dependent, for example, are

characteristics of hegemonic femininity. These characteristics are considered hegemonic because they are accepted and reinforced by social norms, customs, and behaviors. Hegemonic customs and behaviors are so widely visible that they are rarely questioned; rather, they are accepted as “natural.” Males who act contrary to hegemonic masculinity or females who counter hegemonic femininity face discrimination. Because sport is considered a masculine domain, especially contact sports, there is more support (financially and socially) for boys and men in sport than for girls and women.

Fundamental, non-Western societies are even more restrictive toward women’s sport participation. Within traditional Islamic cultures, women face rigid barriers for participation in sport. This culture also emphasizes modesty, and for women this encompasses completely covering their bodies in clothes. This expected attire may impede participation in sport altogether or limit women’s ability to compete in elite events in which competitive attire is inconsistent with Islamic culture. Only in all-female environments may Islamic women wear shorts and t-shirts; otherwise, they are expected to be covered properly. For Islamic women to compete at elite events, they must forsake traditional attire and modesty, participate in mixed-gender venues, and resist the strongly patriarchal social structure. They also risk severe reprisals for doing so. Female Islamic athletes have received public disavowals and death threats for participating in settings not consistent with Islamic law. Although facing different ideologies of intolerance, Aboriginal women have similar struggles. Whereas women in general are underrepresented in sport, women who are non-white, economically disadvantaged, older, physically challenged, or lesbian, bisexual, and transgendered are even less visible.

### **2.3. Heterosexism**

LGBT individuals in sport face paradoxical circumstances. On the one hand, they may be able to “pass” and be undetected in sporting environments. In which case, they may avoid overt discriminatory actions based on their sexual orientation. On the other hand, that they must conceal their LGBT identities suggests the depth of prejudice toward this social group. LGBT athletes often are perceived as countering hegemonic femininity or masculinity, which appears to underscore prejudice against this social group.

Bias against LGBT athletes is well documented in Western societies. They face verbal harassment, lack

of media attention and endorsements, negative bias by officials or judges, and prejudicial treatment by administrators and coaches. Incidents of physical and verbal abuse by teammates are also documented. Some sport organizations have rules specifically aimed toward prohibiting LGBT participants. For example, the U.S. Ladies Professional Golf Association only allows women who were born female to compete, effectively barring transsexuals from participation. An athlete marching in the opening ceremonies at the 1998 Gay Games exemplified the heightened fear of detection among LGBT athletes from fundamental Eastern countries. Holding a placard for Lebanon, he wore a mask over his face. In Lebanon and other countries with fundamental religious-based governments, LGBT individuals may be imprisoned because of their sexual orientation.

### **2.4. Nationalism**

As international competitions bring together individuals from around the world, they also encourage nationalism. International sport and politics seem inextricably linked. Often, governmental involvement in sport reflects a desire for global recognition or prestige. In the Olympics, for example, if the ideal of promoting world togetherness was truly embraced, news agencies would not report the numbers of medals won by participating countries. However, these medal counts are used by various countries to confirm their sovereignty. This was most evident during the Cold War era when the campaign for democracy versus socialism was partially played out in the international sport scene.

International sport is important economically in terms of world recognition and global esteem. Today, rising industrial nations vie to host the Olympics, and other prominent international sport events, to demonstrate their new strength in the international market. Nationalism also is seen as the sport media constructs national identities through coverage of major sporting events. Positive national identities are reinforced, whereas stereotypes about, and difference between, athletes from other countries are accentuated. Nationalism has led to situations of poor sporting behaviors, such as fan hooliganism, fights during competitions, and unfair judging or refereeing. Additionally, some national sport organizations, such as in Japan, only support people of local heritage. Foreigners and immigrants are excluded due to institutionalized nationalism.

### 2.5. Religious Intolerance

As previously discussed, religious beliefs may influence sport experiences. One social group particularly affected by religious influence on sport is Islamic women. Discrimination based on religion may include situations in which important sporting events are scheduled on non-Christian religious holidays, religious expectations are not accommodated, religious customs are belittled, prejudicial jokes are shared or prejudicial comics are displayed, or players are overtly discriminated against because of their religious beliefs.

Perhaps the most evident occurrence of religious intolerance manifest in sport was at the 1972 Olympics when 11 Jewish athletes and coaches were assassinated by terrorists. In the current global climate, sportsmen of Arab descent are viewed cautiously and considered potential terrorists due to their heritage. Also, the revealing uniforms common in Western women's sport are not consistent with the modesty espoused in many Asian and Islamic cultures. Still, some international sport organizations require athletes to wear these styles. Similarly, many sport settings do not accommodate the religious practices of Islamic athletes who engage in obligatory prayer times.

### 2.6. Interactions among Social Prejudices

The various forms of social exclusion interact to create a hierarchy of status and power in sport. For example, when examining racism in sport, the focus is often on men's sport. The omission of considering race in women's sport is a form of sexism and neglects consideration of the interaction between racism and sexism. Female athletes of color face challenges related to their gender as well as their race. Across sporting cultures, males seem to reap the most privilege, particularly white, heterosexual males. Male athletes of color are afforded the privilege of being male, although they are marginalized due to their race or ethnicity. In countries where people of color are the high-status majority, heterosexual males still garner the most privilege and power. Additionally, it is important to consider the influence of social class. When examining sport across diverse countries, people of color or ethnic minorities disproportionately are the most socially disadvantaged people. Females are also represented disproportionately in socially disadvantaged classes; in this case, racial and ethnic minority females face the largest social and economic challenges.

Throughout the world, LGBT individuals face much discrimination, regardless of, or in addition to, their race, ethnicity, social class, or gender. However, white, gay males still have more social benefits than females or people of color in these communities. All of these patterns of gender, racial, ethnic, and sexual orientation superiority are consistently evident across Western societies. In non-Western societies, there may be different cultural forces affecting different patterns of privilege and discrimination, but discernable patterns are evident. As global sport becomes more commodified, minority participants continue to become even more distanced.

## 3. CELEBRATING DIFFERENCE THROUGH SPORT

A number of sporting events have been developed to provide safe and comfortable environments for individuals who commonly experience discrimination in traditional sporting venues. These events cater to specific social groups and include competitive, social, and cultural activities. Examples include competitions for Jewish athletes, LGBT athletes, Arab athletes, and Islamic athletes. Such events promote cultural pride and self-esteem. Participants engage in competitive events, but they also are surrounded by individuals with similar social identities. They are able to work toward sporting goals and immerse themselves in a setting that celebrates their unique culture. Athletes are freed from discrimination and inequity, creating a personally and athletically empowering environment.

The Maccabiah Games are an Olympic-style event held in Israel every 4 years. The best Jewish athletes in the world are invited to participate in athletic, cultural, and educational events. The games offer the opportunity to share participants' customs and heritage in a competitive setting. The Gay Games offer a similar setting for LGBT athletes. Every 4 years the Gay Games celebrate the sporting and cultural practices of the participants. These games began as a venue for competition in an environment free from discrimination. They also counter stereotypes about LGBTs as well as celebrate the distinctiveness of LGBT life.

The Arab Games were created as a competitive event for male athletes from the nations of the Arab League. These games cater only to Arab males; participation by women is contrary to dominant religious beliefs in these countries. Recently developed, the Islamic Countries Women's Sport Solidarity Games provide a

competitive environment for women consistent with Islamic law. Competitive events are held in an all-female environment. Males may be spectators only at the opening ceremonies, where the women wear traditional Islamic attire.

To promote national pride and enthusiasm in European small nations, the Games of the Small States of Europe bring together athletes from Andorra, Cyprus, Iceland, Liechtenstein, Luxembourg, Monaco, San Marino, and Malta. Because athletes from these nations generally cannot keep up competitively with athletes from larger countries, these games provide the opportunity to compete successfully at an elite level. The North American Indigenous Games celebrate native people in the United States and Canada. A variety of competitions also exist for Australian indigenous peoples. These events often include traditional Aboriginal games as well as contemporary sport competitions.

The personal and cultural influence of these types of competitions is remarkable. For many athletes, these are the only places where they can compete devoid of prejudice and discrimination. Personal, cultural, and athletic empowerment often result. Athletes celebrate their shared history and culture while also engaging in international competition. Typically, there is extraordinary energy and excitement in such passionate environments. However, the necessity for these types of competition may be questioned because they may be perceived as separatist. In the current social climate, worldwide, these competitions seem to provide an important experience for the athletes. When athletes with devalued social identities can participate comfortably and safely in integrated competitions, perhaps independent competitions no longer will be needed.

#### **4. SOCIAL EQUITY IN SPORT**

Although examples of inequitable treatment in sport are evident, in many ways, sport brings together people from diverse social groups. It has the potential to counter stereotypes and inequitable treatment of minority athletes. For example, Amir Hadad, an Israeli Jew, and Aisam-Ul-Haq Qureshi, a Pakistani Islamic, competed as doubles partners in the 2003 Wimbledon tennis tournament. During international events such as the Olympics, diverse athletes are seen interacting on the playing fields as well as socializing in the Olympic Village. After the September 11, 2001, terrorist attack on the United States, American athletes were met with great encouragement and support throughout the world.

Sport has become a platform to increase awareness of health concerns such as breast cancer and AIDS. Races and bicycle rides are used to raise money for research and assistance programs. Examples of sport as social agent are widespread: In Kenya, sport helped fight AIDS; through sporting events money was raised to fight cancer in the United Kingdom; sport programs are used to reduce crime in Germany and the United States; and sports assist in overcoming religious intolerance in Ireland and the Middle East. Many sport stars have created programs to reach out to disadvantaged children to provide them the benefits of sport. For example, Tiger Woods created the Tiger Woods Foundation, which introduces golf to minority and at-risk, inner-city youth. Sport organizations, such as the Professional Golf Association and the National Hockey League, are also reaching out to minority youth. The potential of sport to educate people and increase awareness about a wide range of issues, fight social injustice, and provide humanitarian assistance is immeasurable.

##### **4.1. Opposing Social Prejudice through Sport**

Because of the wealth, status, power, and popularity associated with sport, it can be a strong catalyst for social change. Many national and international sport organizations are working toward more equitable sport environments. A wide array of international sport organizations and conferences have addressed issues of inequity in sport. Sporting organizations are making policy to minimize sexism, racism, and heterosexism. Attempts to provide sport opportunities for economically disadvantaged people are also incorporated into these policies.

##### **4.2. Reducing Racism in Sport**

Perhaps the most poignant example of sport as an agent of social change was during the anti-apartheid movement. Most international sport events banned South Africa from participating until apartheid was abolished. Along with other political pressures, South Africa eventually succumbed and changed its social system. This international campaign in sport helped bring down the horrific racism in South Africa and permanently changed international sport. This effort toward reducing racism continues today. For example, participation in the Commonwealth Games is contingent upon commitment to countering racism in disadvantaged countries.

Sport, the Olympic Games in particular, led to serious reconsideration of the inequitable treatment of Aboriginal people in Australia. In response to national protests after being awarded the 2000 Olympic Games, the International Conference on Human Rights and Sports was held in Sydney in 1999. Here, the plight of the Aboriginal people was highlighted and propositions to reduce prejudice and discrimination were developed. The Indigenous Sport Program, developed to increase sport opportunities at all levels as well as provide economic assistance for rising elite Aboriginal athletes, was one outcome of this conference.

Many professional sport organizations have enacted policies specifically to counter racism. FIFA, soccer's world governing body, held an antiracism conference in Buenos Aires in 2001. Delegates from 200 football (soccer) national associations engaged in brainstorming to develop strategies to minimize racism in football. Ultimately, FIFA passed a resolution expecting that all clubs, associations, and officials actively take steps to eradicate racism in football. In 1999, Football against Racism in Europe (FARE) was developed, involving organizations from 13 countries. FARE assists and links various local and national endeavors to counter racism in football (soccer). By 2003, all 92 European professional soccer clubs supported the "Kick Racism out of Football" campaign. "Show Racism the Red Card" is another antiracist education campaign in the United Kingdom. This program engages professional footballers as antiracist role models. Sportscotland developed a comprehensive equity strategy for sport. It encourages all sport governing bodies to address inequities related to race, ethnicity, gender, sexuality, ageism, religion, and disability. Similar programs have emerged throughout the world.

### 4.3. Reducing Sexism in Sport

A variety of international organizations to counter sexism in sport have emerged. The International Association of Physical Education and Sport for Girls and Women was founded in 1949. This group represents and lobbies for women and girls in sport and physical education at the international level. It is a member of the International Council for Sport Sciences and Physical Education and therefore is involved in developing international policy regarding women's sport.

In 1994, the International Working Group on Women and Sport was established during the first World Conference on Women and Sport. This group drafted the Brighton Declaration, which implored governments

and sport organizations to "comply with the equality provisions of the Charter of the UN, the Universal Declaration of Human Rights, and the UN Convention on the Elimination of all Forms of Discrimination Against Women." The Second World Conference on Women and Sport was held in Windhoek, Namibia, in 1998. It resulted in the Windhoek Call for Action, which reaffirmed these principles. In response to these declarations supporting women's sport, many conferences and resolutions worldwide followed. For example, in 2001, the First Asian Conference on Women and Sport was held in Osaka, Japan. At this conference, delegates reaffirmed their commitment to the Brighton Declaration and the Windhoek Call for Action and dedicated themselves to the advancement of women in sport.

In addition, the Women's Sport Foundation in the United States, United Kingdom, and Australia lobbies for increased sport opportunities for women and girls in sport. In Canada, the Canadian Association for the Advancement of Women in Sport works with Sport Canada to increase opportunities for girls and women in sport and to ensure healthy environments for female participants. Similarly, the Australian Sports Commission expressly supports women's sport. The European Women and Sport organization, representing 41 nations, promotes gender equality in European sport organizations.

National policies to reduce sexism and increase opportunities for women have been applied to sport. In the United States, Title IX makes it illegal to discriminate against individuals based on sex in any public educational setting, including sport. This law greatly affected women's sport by decreeing that girls and women receive equitable opportunities, funding, coaching, and facilities. Although still not equal to boys and men, since the enactment of Title IX, the number of girls and women participating in sport, and social acceptance of female athletics, has greatly increased. In Canada, the Charter for Rights and Freedoms protects the sport opportunities for girls and women.

### 4.4. Securing Child Welfare in Sport

Concern for the well-being of child participants is another area in which sport organizations encourage fair treatment. An international organization, Defense for Children International, based in Switzerland, raised concerns about the exploitation of children in elite sport. It advocated that children should have the following rights in sport: "(i) to do as one chooses with one's body, (ii) to be consulted, (iii) to physical and

psychological integrity, (iv) to associate freely with whom one chooses, and (v) to education.” In the United Kingdom, the British Association for Sport and Exercise ratified a welfare policy mandating that its members safeguard and protect children and vulnerable adults from sexual abuse.

#### **4.5. Economic Reforms through Sport**

Through sport, international awareness has increased concerning the horrendous working conditions of sweatshops in developing countries. United Students against Sweatshops (USAS), an international student movement among university students, has led the campaign fighting for better working conditions and rights for workers. It has lobbied universities and colleges to be proactive and discerning when making contracts with various sporting goods companies that sell items with university logos. Groups such as USAS have led boycotts against sporting goods companies (e.g., Adidas, Nike, and Umbro), imploring the elimination of exploitive child labor, procurement of fair wages, and safe work environments.

Although all the organizations and sport group actions described previously focus on specific social justice issues (e.g., racism or sexism), they also recognize the interactions among these issues. For example, groups fighting sexism also forward policies against heterosexism and racism. While focusing on the abysmal working conditions in sweatshops, it is apparent that women and children of color are most likely to be engaged in these positions. Sport organizations such as national Olympic committees, professional sport organizations, and national governing bodies have programs encouraging diverse athletes to participate in sport as well as policies countering sexism, racism, and other social injustices.

When considering fair play in sport, there are two sides to the issue. Certainly, inequitable treatment of diverse sports people occurs. Individuals with minority social identities may be marginalized and devalued in sport; they may face stereotypes, prejudice, and barriers. However, sport also brings together diverse people. In the name of fair play, sport celebrates achievements of all people and provides an environment in which divergent social groups can leave behind political agendas and simply engage in sport. Sport also has become a tool for social justice activities. Through sport, intolerance is challenged and social change continues.

### **5. APPLIED SPORT PSYCHOLOGY AND SOCIAL JUSTICE**

Given the potential of sport to encourage social justice and to challenge unfair treatment, sport psychologists are in an ideal position to effect such action. Applied sport psychologists work with athletes, coaches, parents, and administrators to develop productive sport behaviors leading to improved or consistent performance and enjoyment. Inequitable sport settings limit both of these possible outcomes. Thus, efforts to resist inequitable treatment logically complement more customary sport psychology goals.

Challenges to inequitable treatment may be focused on a societal level, as previously discussed, or it may be approached on a local level by creating positive and fair team environments. Teaching young athletes to value fairness and rebuke prejudice and discrimination will eventually affect all of sport. One team or league can make a major difference. For example, when teenage members of a golf team refuse to compete because a teammate is not allowed to play due to his or her skin color, their action speaks volumes to the larger society. Members of sport teams that engage in community service learn about people living in different situations than their own. College athletes in the United States, for instance, help build shelters for homeless people and tutor economically challenged children. Sport psychologists can engage the team in discussions highlighting the social justice issues inherent in such settings.

Many sport psychologists already endorse the importance of developing constructive team climates through creating mastery or task-oriented sport environments. In such settings, coaches emphasize skill development, cooperation, improvement, and development of peer relationships. In contrast, performance- or ego-oriented climates emerge when coaches encourage interpersonal competition and promote a “win at all costs” attitude. When athletes are rewarded for effort, improvement, and teamwork, they are likely to have a long-term investment in sport, use healthy training techniques, and adopt positive sporting behaviors. These athletes also tend to believe that success in sport is due to hard work and working cooperatively with others. Athletes who are rewarded primarily based on social comparison (e.g., winning) are likely to be willing to do whatever it takes to win, such as using aggressive or harmful behaviors including trash talking in a prejudicial manner or other overtly discriminatory behaviors.

A logical extension of traditional means of creating a task-oriented team setting is to address issues of social justice and encourage fair treatment of all athletes. It seems reasonable that athletes who participate in a task-oriented climate will be likely to accept the importance of being sensitive to individual differences. If, for example, athletes of different races or ethnicities are dedicated to team improvement and are invested in being successful, they will also be willing to work together in a productive manner. Creation of a task-focused environment may be the first step toward promoting fair treatment within sport teams or sport leagues.

Sport psychologists and coaches wanting to challenge unfair treatment due to prejudice may employ a variety of strategies. An initial, but important, step in this direction is to consider the language used by leaders and team members. Coaches and sport psychologists should model appropriate language. Racist, sexist, homophobic, or any other prejudicial comments should not be tolerated. As individuals point out the inappropriateness of such language and reinforce its unacceptability, athletes will become more aware of how their words may be hurtful. Situations in which prejudice or discrimination occur provide opportunities to open dialogue about sensitive issues. Furthermore, most strategies to encourage team cohesion will also promote acceptance of diverse people within a team. Creating team goals that include cooperative efforts, providing athletes opportunities to learn about athletes with different social identities, and emphasizing team diversity as a strength will lead to enhanced cohesion while also minimizing prejudice. Athletes should also be encouraged to “do the right thing.” Empowering athletes to acknowledge and challenge prejudice and discrimination will further increase the likelihood that sport settings will promote fair treatment of all participants.

Ensuring equitable treatment in sport will require the involvement of athletes, coaches, and administrators.

Sport psychologists are positioned to interact and educate each of these factions. As sport psychologists make social justice in sport a priority, it is more likely that sport will achieve its potential to bring together diverse people for fair competition, create avenues for cultural exchange, and ensure positive experiences for majority and minority athletes alike.

### See Also the Following Articles

Diverse Cultures, Dealing with Children and Families from  
 ■ Gender and Culture ■ Homophobia ■ Homosexuality  
 ■ Prejudice and Discrimination ■ Sexual Behavior ■ Sexual Behavior and Culture ■ Stereotypes

### Further Reading

- Andersen, M. (Ed.) (in press). *Understanding applied sport psychology*. Champaign, IL: Human Kinetics.
- Carrington, B., & McDonald, I. (Eds.) (2002). *“Race,” sport and British society*. London: Routledge.
- Coakley, J., & Dunning, E. (Eds.) (2000). *Handbook of sports studies*. London: Sage.
- Collins, M. F. (2003). *Sport and social exclusion*. London: Routledge.
- Griffin, P. (1998). *Strong women, deep closets: Lesbians and homophobia in women’s sport*. Champaign, IL: Human Kinetics.
- Hargreaves, J. (2000). *Heroines of sport: The politics of difference and identity*. New York: Routledge.
- Kidd, B., & Donnelly, P. (2000). Human rights in sport. *International Review for the Sociology of Sport*, 35, 131–148.
- Maguire, J. (1999). *Global sport: Identities, societies, civilizations*. Malden, MA: Blackwell.
- Messner, M. A. (2002). *Taking the field: Women, men, and sports*. Minneapolis: University of Minnesota Press.
- Schaffer, K., & Smith, S. (Eds.) (2000). *The Olympics at the millennium: Power, politics, and the Olympic Games*. Piscataway, NJ: Rutgers University Press.



# Family and Culture

James Georgas

*University of Athens, Athens, Greece*

1. Culture
  2. Family Theories
  3. Family
  4. Kinship Patterns and Family Matters
  5. Family Roles and Power
  6. Psychological Dimensions
  7. Conclusion
- Further Reading

There are two general approaches to the study of family. One approach has been the focus on the family, its structure, interactions between family members, child development, and other issues. The second approach has been the study of the family within its cultural context. This approach assumes that the family is one system within other social systems of a culture, and that the study of the family necessarily entails studying its relationships with the institutions and other aspects of the culture. In addition, the usual approach in cultural psychology studies the family in one culture or society, whereas the cross-cultural approach is comparative, studying universals and differences in families across cultures or countries. Cultural anthropology, sociology, psychology, cross-cultural psychology, psychiatry, economics, political science, and geography have each contributed to the study of family from the perspective of its own discipline. However, there is considerable overlap in theories and methods so that one can also talk about a multidisciplinary approach employed by many in each field. A number of issues are related to

the study of family and culture. There are different family types: nuclear families, extended families, and one-parent families. Family types differ according to subsistence patterns, social systems, religion, and other cultural features. Marriage takes place in different ways in different cultures. Kinship ties, postmarital residence, and descent are important determinants of family types. The social power of the mother and the father is related to the institutional structure and values of the culture. Socialization of children by the family is related to ecocultural variables and family types and results in different psychological characteristics across cultures.

## 1. CULTURE

Culture is a very broad term referring to all human-made patterns of activity, including societal institutions, practices, norms, rules, means of subsistence, symbols, communication, settlements, transportation, art, religious beliefs, ideas about nature and humans, interactions among humans, child-rearing practices, values, beliefs, cognitive processes, personality, and social behavior. There are many definitions of culture, each stressing different elements, including the patterns of human activity, the unique symbols in each culture, the symbols transmitted through language, the achievements of the group, the artifacts, the transmission of the traditions from generation to generation, and the mutual process of humans creating cultural elements and culture-shaping human and social patterns.

The term culture, however, is also used to designate a group with a common heritage as described in the previous definition of culture. That is, culture is used as a synonym for “ethnic group,” “nation,” “society,” or “country.” Anthropologists initially employed the term to refer to small exotic groups of people, often referred to in the past as primitive peoples, savages, or indigenous groups. Employed in this sense, larger geopolitical groups are referred to as nations or countries. However, culture is also employed in taxonomies of nations, such as Western culture or Islamic culture. If culture is used to define an ethnic group with a specific “culture,” then the number of cultures has been estimated by anthropologists to be in the thousands, whereas there are approximately 200 nations or countries throughout the world.

Thus, culture is employed in both ways in family research: It refers to a specific country or to those elements that characterize the society.

## 2. FAMILY THEORIES

Family theories have been developed in sociology, cultural anthropology, psychology, psychiatry, and cross-cultural psychology. The first family theories were developed in the 19th century. A form of social Darwinism theorized that the structure and function of family adapted, as a social organism, to the environment. Marxist theory employed the concept of economic determinism to explain how economic resources determined social power, which in turn determined class struggle. The concept of social power was extended to explain authority and power structures of the father and mother. Durkheim’s theory explained the existence and the changes in family structure and function in terms of the family’s functional role in the preservation of society. Functionalism analyzed the role of family as part of a greater whole, in which an equilibrium with other social institutions was established. This was essentially a systems theory, in contrast to the biological or economic determinism of social Darwinism and Marxist theory.

One of the most influential theories of family in the second half of the 20th century was Parsons’ structural–functional perspective, in which society was viewed as an organism striving to resist change and to maintain itself in a state of equilibrium. Structure refers to the members of the family (parents, child, and kin), and function refers to how families satisfy physical and psychological needs for survival and maintenance. According to Parsons, the adaptation of the family

unit to the industrial revolution required a nuclear family structure that could carry out societal functions and satisfy the physical and psychological needs of family members. However, in its urban setting, the nuclear family is fragmented from its kinship network, leading to psychological isolation. Distancing itself from the extended family results in loss of its productive, political, and religious functions. Social mobility, particularly in the highly mobile U.S. culture, was made possible by the breaking of family ties.

However, most sociologists, anthropologists, and psychologists today disagree with Parsons’ depiction of the nuclear family as isolated from its kin. Evidence from studies of family networks indicates that ties are maintained between members of the nuclear family and kin at different generational levels in the United States and many European countries. Many ethnic groups in the United States maintain fairly close family and kinship ties, whereas others do not. Cross-cultural studies have indicated that there are not polar opposites of no family ties versus close family ties but, rather, different degrees of family ties. In countries such as the United States, Sweden, and Britain there are looser ties between the nuclear family and kin, whereas the ties are closer in countries such as China, India, and Greece.

Family theories based on systemic concepts have also been developed (Parson’s theory is also a systems theory). Family system theory has been developed within general systems theory, and its basic concepts apply to physics, biology, and social sciences, among other sciences. The family is a system within larger suprasystems of society and attempts to maintain equilibrium by adapting to demands and to changes in the larger system. The focus is on communication processes between family members and on recurring family transactional patterns. Family system theory was developed in the fields of clinical psychology and psychiatry as a psychotherapeutic method. This has resulted in the construction of models of family functioning and techniques of psychotherapy directed toward the family. Ecological theories of family trace the relationships between environment, social institutions, the family, and psychological variables. The human ecology theory of Bronfenbrenner is basically a family systems model. One of the earliest ecological models is that of cultural anthropologists John and Beatrice Whiting and Irvin Child and their many collaborators in their six-culture study. Their basic hypothesis was that child training throughout the world is in many important aspects identical and that there are universal

aspects of behavior but also differences from culture to culture. Six cultures were studied in detail: those of Raira, Okinawa; Tarong, the Philippines; Khalapur, India; Nyansongo, Kenya; Juztlahuraca, Mexico; and Orchard Town, United States. The elements of the model were the environment (climate, flora, fauna, and terrain), history of the settlements, maintenance systems (subsistence patterns, means of production, settlement patterns, social structure, etc.), the child's learning environment (caretakers and teachers, mothers' workload, etc.), psychological variables (behavioral styles, skills and abilities, values, conflicts, and defenses), innate needs, and projective expressive systems or elements of the culture (religion, magic beliefs, ritual and ceremony, etc.) (Fig. 1). Cultural features such as the economy, social structure, settlement pattern, and household and family organization were found to be related to behavior. For example, the complexity of the socioeconomic system and the composition of the household were predictive of social behavior of the system. This is a classic cultural anthropological model for cross-cultural studies of family.

Other family theories employed include feminist theory, symbolic interactionism, family development, phenomenology, family power, and exchange theory.

### 3. FAMILY

#### 3.1. Definition of Family

Currently, there are many definitions of family based on different theoretical perspectives. An early definition of the family was that of the anthropologist Murdock: "The family is a social group characterized by common residence, economic cooperation, and reproduction. It includes adults of both sexes, at least two of whom maintain a socially approved sexual relationship, and one or more children, own or adopted, of the sexually cohabiting adults." The functions of family were considered to be sexual, reproduction, socialization, and economic. The sociologist Popenoe defined family in terms of recent changes in the social and economic position of the family, the increase in one-parent families [divorced, adoptive, widows(ers), or unmarried mothers], and homosexual families. In this definition of the family, the minimum number is one adult and one dependent person, the parents do not have to be of both sexes, and the couple does not have to be married. The functions of the family are procreation and socialization of children, sexual regulation, economic cooperation, and provision of care, affection, and companionship.

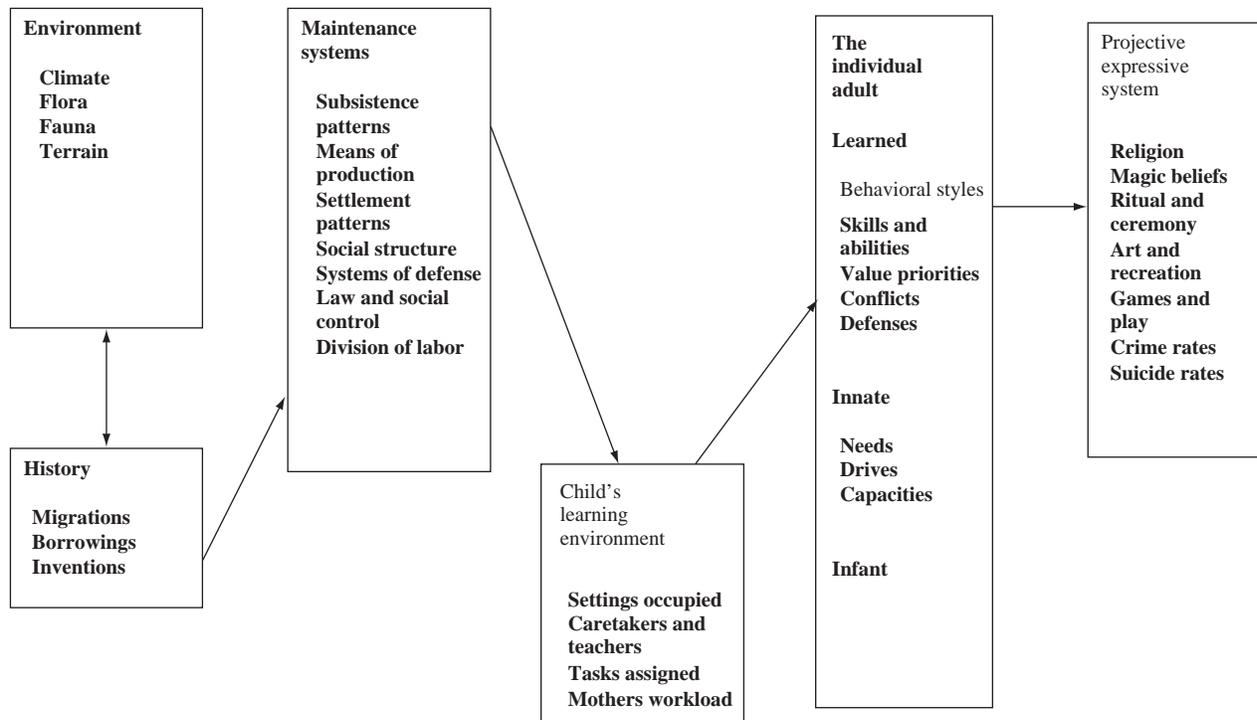


FIGURE 1 A model for psychocultural research. From Whiting and Whiting (1975).

### 3.2. Family Typology

As inferred in the previous definitions, there are different types of families. The structure refers to the positions of the members of the family (e.g., mother, father, daughter, grandmother, etc.) and the roles assigned to the family members by the culture. For example, traditional roles of the nuclear family in North America and northern Europe in the mid-20th century were the wage-earning father and the housewife and child-raising mother. Cultures have social constructs and norms related to the proper roles of family members—that is, what the role of the mother, father, etc. should be.

Family types or structures have been delineated primarily by cultural anthropological studies of small cultures throughout the world. However, family sociologists have also contributed to the literature on family typology, although sociology has been more interested in the European and American family and less interested in small societies throughout the world.

There are a number of typologies of family types, but a simple typology would be the nuclear and the extended family systems. To these can be added the one-parent family.

The nuclear family consists of two generations: the wife/mother, husband/father, and their children. The one-parent family is also a variant of the nuclear family. Most one-parent families are divorced-parent families; unmarried-parent families comprise a small percentage of one-parent families, although they have increased in North America and northern Europe. The majority of one-parent families are those with mothers.

The extended family consists of at least three generations: the grandparents on both sides, the wife/mother and the husband/father, and their children, together with parallel streams of the kin of the wife and husband. There are different types of extended families in cultures throughout the world. The following is one taxonomy:

- The polygynous family consists of one husband/father and two or more wives/mothers, together with their children and kin. Polygynous families are found in many cultures. For example, four wives are permitted according to Islam. However, the actual number of polygamous families in Islamic nations is very small (e.g., approximately 90% of fathers in Qatar, Kuwait, United Arab Emirates, Oman, Bahrain, and Saudi Arabia have only one wife). In Pakistan, a man seeking a second wife must obtain permission from an arbitration council, which requires a statement of consent from the first wife before granting permission.

- In a few societies in Central Asia there are polyandrous families, in which one woman is married to several brothers and thus land is not divided. However, this is a rare phenomenon in cultures throughout the world.

- The stem family consists of the grandparents and the eldest married son and heir and their children, who live together under the authority of the grandfather/household head. The eldest son inherits the family plot and the stem continues through the first son. The other sons and daughters leave the household upon marriage. The stem family was characteristic of central European countries, such as Austria and southern Germany. The lineal or patriarchal family consists of the grandparents and the married sons. This is perhaps the most common form of family and is also found in southern Europe and Japan.

- The joint family is a continuation of the lineal family after the death of the grandfather, in which the married sons share the inheritance and work together. Joint families were found south of the Loire in France, as were patriarchal families, whereas the nuclear family was predominant north of the Loire. Joint families are also found in India and Pakistan.

- The fully extended family, or the *zadruga* in the Balkans countries of Croatia, Bosnia, Serbia, Montenegro, Albania, Macedonia, Bulgaria, had a structure similar to that of the joint family but with the inclusion of cousins and other kin. The number of kin living and working together as a family numbered in the dozens.

A point needs to be made regarding the different types of extended families. Historical analyses of the family by anthropologists and sociologists indicated that people considered to be members of a family or a household were not necessarily kin. For example, in central European countries until the 18th century, servants (who were often relatives), semipermanent residents, visitors, workers, and boarders were considered to be members of the household. The term *familia* was used to denote large households rather than “family” in the modern sense. Until the 18th century, no word for nuclear family was employed in Germany but the term “with wife and children.” Frédéric Le Play, considered to be the father of empirical family sociology, discussed the emergence of the nuclear family as a product of the industrial revolution. He also characterized the nuclear family, the *famille*, as unstable in comparison with the stem family.

One theory regarding the change from feudal *familia* to the *famille* of Western Europe is based on

the following analysis. After the reformation, vassals left the feudal towns to seek work in the cities. This led to the separation of the dwelling place and place of work and resulted in privacy and the sentimentality of the nuclear family. This pattern, however, was not found among the peasants in the agricultural areas. The strengthening of the relationship between parents and children was also a result of the religious influence of the Age of Enlightenment. These changes led to the releasing of servants from the close community of the household. Servants and workers became less personal and part of the household and more contractual. This led to the emergence of many new nuclear families (e.g., those of early factory workers and clerks). A new word in German, *Haus*, referred only to those living within it.

Historical analyses of the family during this period in Western Europe also emphasize that not all families were large extended families because establishing this type of household was dependent on land ownership. Most families worked for large feudal types of households and were essentially nuclear in structure. In England during this period, where land ownership was restricted to the nobility, the vast majority of families, which either worked for the landowners or rented small plots, were necessarily nuclear families.

### **3.2.1. The Nuclear Family: Separate or Part of the Extended Family?**

The key element in studying different types of family structure and its relationships with psychological development of the children, its economic base, and its culture is the nuclear family. In 1949, Murdock made an important distinction regarding the relationship of the nuclear family to the extended family: "The nuclear family is a universal human social grouping. Either as the sole prevailing form of the family or as the basic unit from which more complex familial forms are compounded, it exists as a distinct and strongly functional group in every known society."

Murdock made an important point: The nuclear family is prevalent in all societies, not necessarily as an autonomous unit but because the extended family is essentially a constellation of nuclear families across at least three generations. Parsons' theory that the adaptation of the family unit to the industrial revolution required a nuclear family structure resulting in its isolation from its traditional extended family and kinship network, leading to psychological isolation and anomie, has had a strong influence on psychological and

sociological theorizing about the nuclear family. However, studies of social networks in North America and northern Europe have shown that the hypothesized isolation of the nuclear family is a myth. Nuclear families, even in these industrial countries, have networks with grandparents, brothers and sisters, and other kin. The question is the degree of contact and communication with these kin, even in nations of northern and southern Europe.

A second issue relates to the different cycles of family, from the moment of marriage to the death of the parents or grandparents. The classic three-generation extended family has a lifetime of perhaps 20–30 years. The death of the grandparent, the patriarch of an extended family, results in one cycle closing and the beginning of a new cycle with two or three nuclear families, the married and unmarried sons and daughters. These are nuclear families in transition. Some will form new extended families, others may not have children, some will not marry, and others (e.g., the second son in the stem family) will not have the economic base to form a new stem family. That is, even in cultures with a dominant extended family system, there are always nuclear families.

A third issue is the determination of a nuclear family. This is related to place of common residence or the "household" of the nuclear family. Demographic studies of the family usually employ the term household in determining the number of people residing in the residence and their roles. However, there is a paradox between the concepts household and family as employed in demographic studies. Household refers to counting the number of persons in a house. If there are two generations, parents and the children, they are identified as a nuclear family. However, this may lead to erroneous conclusions about the percentage of nuclear families in a country. For example, in a European demographic study, Germany and Austria had lower percentages of nuclear families than Greece. This appears to be strange because Greece is known to be a country with a strong extended family system. However, demographic statistics provide only "surface" information, which is difficult to interpret without data about attitudes, values, and interactions between family members. Nuclear households in Greece, as in many other countries throughout the world, are very near to the grandparents—in the apartment next door, on the next floor, or in the neighborhood—and the visits and telephone calls between kin are very frequent. Thus, although nuclear in terms of common residence, the families are in fact extended in terms of their relationships and interactions.

In addition, there is the psychological component of those who one considers to be family. Social representation of his or her family may consist of a mosaic of parents, brothers and sisters, grandparents, uncles, and aunts and cousins on both sides, together with different degrees of emotional attachments to each one, different types of interactions, bonds, memories, etc. Each person has a genealogical tree consisting of a constellation of overlapping kinship groups—through the mother, father, mother-in-law, father-in-law, but also through the sister-in-law, brother-in-law, cousin-in-law, etc. The overlapping circles of nuclear families in this constellation of kin relationships are almost endless. Both the psychological dimension of family—one’s social representation—and the culturally specified definition of which kin relationships are important determine which kin affiliations are important to the individual (“my favorite aunt”) or the family (“our older brother’s” family) and which are important in the clan (the “Zaman” extended family) or community (the “Johnsons” nuclear family). Thus, it is not so important “who lives in the box” but, rather, the types of affiliations and psychological ties with the constellation of different family members or kin in the person’s conception of his or her family, whether it is an “independent” nuclear family in Germany or an “extended family” in Nigeria.

### 3.3. Subsistence Patterns and Family Types

Ecological theories specify relationships between the family structure and function, the ecological context (physical environment, climate, flora, fauna, terrain, and settlement patterns), the sociopolitical context (social structure, subsistence patterns, traditions, religion and beliefs, language and means of communication, etc.), and psychological variables (cognition, personality, values, social interactions, etc.). The ecological context and the sociopolitical context can shape family types, but the model is not strictly deterministic: Individuals or the community can also shape family types. The number of wives, the dowry system, and relationships with kin in a family type can be shaped by traditions and religion. However, perhaps the most important determinant of family type is the subsistence system, which is also shaped by the ecological context, particularly in small societies. In hunting and gathering societies, in which mobility of the family is necessary in order to obtain food, the nuclear family type is

prevalent. In agricultural and pastoral societies, in which there is a permanent base and many hands are necessary for cultivation of the land, the extended family types are prevalent. In modern industrial society, in which people are hired to provide services or to work in industries, money is a means of exchange, people live in urban areas, and apartments or houses are expensive, the nuclear family is prevalent.

## 4. KINSHIP PATTERNS AND FAMILY MATTERS

### 4.1. Kinship Relationships

Cultural anthropology has categorized the different types of kinship relationships in extended families in a variety of cultures. These are called lineal, collateral, and affinal relationships. Lineal relationships are those between the grandparents and the grandchildren. Collateral relationships are those with uncles and aunts, cousins, and nephews and nieces. Affinal relationships are those between parents-in-law, children-in-law, and siblings-in-law as well as with matrilineal and patrilineal kin. However familiar the terms for these kin are in Europe and North America, they are critical concepts in different cultures because they are also related to the types of relationships and obligations toward affinal, collateral, and affinal kin, to lines of descent, to residence, to inheritance of property, and to roles.

The terms for kin are more differentiated in various cultures. Murdock employed a system with six categories based on categorization of cousins: Hawaiian, Eskimo, Sudanese, Iroquois, Crow, and Omaha. Other categorizations are based on classification of paternal and maternal relatives. For example, in Iran the term in Farsi for mother’s sisters is *khaleh* and it is *ammeh* for father’s sisters, with analogous terms for father’s brothers. In Pakistan, several cousins may be raised together in the same household. They are referred to in Urdu as brothers or sisters, with the prefix added, “through-my-maternal-aunt” or “through-my-paternal-uncle.” Generic words, such as aunt, uncle, or even grandparents, have no equivalent in Urdu, but specific terms related to matrilineal kin, as in Iran, are employed; for example, *khala* is the mother’s sister and *phuphee* the father’s sister.

### 4.2. Residence

Residence refers to where the couple resides after marriage. One theory is that subsistence patterns

determine rules of residence, which in turn affect the rules of descent. The theory is based primarily on small, noncomplex societies.

- Patrilocality refers to residence with or near the husband's patrilineal kinsmen. This is the most common form of postmarital residence, with approximately 50% of small societies sampled.
- Virilocality refers to residence near the husband's father and is similar to patrilocal residence. The difference is the absence of patrilineal kin groups. This represents the second largest percentage (20%) of postmarital residence after patrilocal. The remainder of residence patterns sum to the remaining 30% of societies.
- Matrilocality is similar to patrilocal but refers to residence with or near the wife's matrilineal kinsmen. This is the case, for example, with the Chiapas of Mexico.
- Avunculocal refers to residence with or near the maternal uncle or other male matrilineal kinsmen of the husband.
- Bilocal is residence with or near the parents of the husband or the wife.
- Matrilocality–avunculocal is similar to bilocal but can be either matrilocality or avunculocal residence.
- Avunculocal–virilocality is also similar to bilocal but can be either avunculocal or virilocality residence.
- Neolocal means residence apart from the relatives of both spouses. This is characteristic of nuclear family residence in northern Europe and North America. In complex industrial societies in which there are few small independent farms, the economy is based on wages rather than land ownership. Thus, land ownership as a basis of production is not a factor, and small nuclear families with neolocal residence are more functional.
- Duolocal means separate residence of each spouse in the paternal and maternal household.

### 4.3. Descent

Descent refers to the relationships with paternal and maternal kin. The relationships concern many aspects of life, such as economic activities, religious activities, child-rearing practices, political behavior, inheritance, and residence. These aspects of living are organized around strict sets of rules with kin, many of them as ritualized behaviors. Societies with specific rules for different kin are usually hierarchical in structure and in the distribution of power. The following are rules of descent in different cultures.

*Bilateral*: Affiliation is with both mother's and father's relatives.

*Patrilineal*: Affiliation is with kin of both sexes through the maternal and paternal fathers only, but not through maternal and paternal mothers.

*Matrilineal*: Affiliation is with kin of both sexes through the maternal and paternal mothers only, but not through maternal and paternal fathers.

*Ambilineal (cognatic)*: Affiliation is with kin through either the maternal parents or the paternal parents. Some may affiliate with kin through their mothers and others through their fathers.

*Double*: Affiliation is with both father's patrilineal kin and mother's matrilineal kin.

### 4.4. Marriage

Societies have norms regarding whom one is permitted to marry (endogamy) and restrictions regarding whom one cannot marry (exogamy). In some societies, such as in India or Pakistan, endogamy means that marriage is restricted to the same caste, the same village, the same religion, and the same race. These social norms are not as restrictive in North America and Europe, although marriage to someone of another racial or ethnic group or to someone of a different level of education or social status may not be accepted in some groups. Exogamy means that, e.g., marriage to a member of the nuclear family, or incest, is a universal taboo in almost all societies. In China, marriage to someone with the same family name was disapproved. In many societies, marriage is not permitted with first cousins or with the son or daughter of a godparent. In other societies, such as Saudi Arabia, marriage with first cousins is desired, specifically to the sons and daughters of the paternal brother (uncle) or to members of the clan. In other societies, if a husband dies, the wife must marry the brother.

In the past, in most societies throughout the world, marriages were arranged between two families, and a verbal or written contract was agreed upon regarding the dowry or the bride wealth; however, this is changing gradually in many societies in Africa and Asia. In patrilineal societies, the fathers usually made the decision. In the few matrilineal societies in the world, the mother's brothers made the decision. Decisions to unite their sons and daughters in marriage are based on certain criteria. Because endogamy is an important restriction, families look for the best possible alliances between them, which may be based on social and economic factors and, in many cases (such as in Saudi

Arabia, where marriage between paternal first cousins is highly desirable), strengthening the wealth, prestige, and social power of the extended family within the clan or tribe. The decision of a young couple to marry on the basis of love with or without the consent of the parents occurs in a minority of societies, including North America and northern Europe, although this is changing rapidly in many societies. One theory is that in societies with neolocal postmarital residence, romantic love as a basis of marriage is more common. It is suggested that marriage based on love and companionship provides a stronger union when the couple must depend on their own economic assets and not on their parents' aid for survival.

The dowry is an integral part of the marriage contract between the two families of the couple. There are three general types of dowries. The dowry system refers to the transfer of property, money, or animals by the bride's parents to the husband after marriage. The bride wealth system, the most common dowry system, is the opposite of the dowry system: The transfer is from the husband's family to the bride's family. A third type is bride theft, in which the man abducts the woman without the consent of the respective families. In the past, the dowry system was in effect in European countries, in addition to other societies, whereas the bride wealth system was in effect in Islamic nations as well as other societies.

Inheritance of property is an integral part of marriage and lineal descent. Japan and China adapted differently to industrialization in the 19th century. Both had an agrarian economy, rapid population growth, and extensive but corrupt and inefficient bureaucracy. In China, inheritance was egalitarian, but in Japan a single child inherited the property, which made it possible to accumulate capital. In China, loyalty was to the family and nepotism a duty, so every family member could benefit from upward social mobility. In Japan, which had a more feudal system, a father could disinherit his son and adopt a young man who seemed more worthy.

#### 4.5. Divorce

There is considerable cultural diversity regarding divorce. Divorce is socially disapproved in all societies, and families usually make great efforts to keep the couple together. The reasons for this are many: the social stigma of divorce, the consequences for the children, the financial provisions, and the issues of property and inheritance.

Societies differ in the degree to which divorce is controlled and by which institutions. In some societies,

divorce is controlled directly by the family, whereas in others there is indirect control by social institutions and by the dominant religion. For example, Catholicism does not permit divorce except under highly unusual situations requiring a special dispensation. The Orthodox Church permits three marriages and three divorces. Islamic law, the *sharia*, permits polygamy up to four wives and also divorces. However, in Islamic nations, the intent of a husband to take a second wife may lead to the first wife seeking a divorce. Despite the conventional notion in the West that in Islam husbands can divorce a wife by stating three times the phrase, "I divorce you," divorce is not an easy matter because of its legal and social consequences. According to Islamic law, the daughter inherits property from the father—half of the sum of the son's. The wife retains property in her name after marriage, even gifts from her family, and the husband has no legal claim to her property. After divorce, the woman retains her property and custody of the children until they reach 7 years of age. However, the father remains the legal guardian and is responsible for the financial care of the children. Remarriage of a divorced woman is permitted, but the children of the first marriage are subsequently raised by the maternal grandmother.

### 5. FAMILY ROLES AND POWER

Role theory refers to the cultural norms regarding psychological and interactional aspects of members of society, such as mothers, fathers, sons, daughters, and grandparents. The originators of role theory are Ralph Linton in sociology and George Herbert Mead in social psychology. Role refers to the social expectations and the social scripts of family roles—how roles have been shaped by cultural conventions and by the collective ideologies of a society. One aspect of role theory studies how roles are learned during the process of social interaction. That is, people interact with others, they see themselves and others as occupants of particular statuses, and they learn guides for action. In other words, there are certain social scripts or expectations associated with certain roles.

Structural role theory as developed by Linton and Parsons refers to the structural and functional aspects of status and role. By structure is meant the positions recognized by the specific social system and the content of the role. That is, what are the social scripts associated with the roles of mother and father in the society? For example, social scripts of gender roles in a society might

be that the place of mother is in the home, she should raise the children, cook, etc., whereas in other societies mothers should work, share with the husbands the daily work of the home and care of children, etc. The functional analysis is concerned with the function of role. That is, what are the consequences for the rest of the social system of the specific role? How does the role contribute to the maintenance of the system? How does a particular role help in the achievement of the goals of the system? How does the role help solve the process of adaptation to the social system and contribute to the maintenance of patterns, such as the values system? Structural role theory also has a comparative element, attempting to determine which structure features of social systems are universal through cross-cultural analysis.

Power, the degree of gender equality, and the division of resources within the family are another aspect of family dynamics in the family. Anthropology has established that power is related to the control of the economic activity or land ownership and institutionalized through the norms of the society. Power can be defined as the influence of the spouse, children, and other relatives. However, the influence is not a personality trait, although there may be individual differences in the ability to influence others. The basis of the influence is the authority bestowed by social norms to different family roles and internalized in terms of values. Power may be expressed in the control of the finances, in the ability to command respect and obedience, in feelings of autonomy and control over one's life, in the capacity to regulate one's dependency on others, in the power to decide where the family lives and whose career is given priority, and in the freedom to be able to leave difficult family situations.

Most agricultural societies are patrilineal, but some are matrilineal. In some agricultural societies, the male does most of the plowing or animal tending and the woman's role is to care for the house and the children. However, in many patrilineal agricultural societies, the woman does much of the work, such as harvesting and tending the crops, in addition to the household tasks and caring for the children. Boys learn the roles of the males and the fathers, and girls learn the roles of the females and the mothers in small societies. The roles in small societies are hierarchical, in which the patriarch of the family has the ultimate power and males have more power than females. In matrilineal societies, the matriarch has more power than the father. The brothers of the matriarch have more power than the father, who is absent from the family and may visit the matriarch secretly at night. However, although the father

may have the institutionalized social power within the family, the influence of the mother in extended families regarding many family matters may be greater than that of the father.

It is important to note that role theory refers to social values and social scripts—what the social scripts should be in the society. However, there is also a psychological dimension; there are individual differences in each society regarding agreement with these roles. Some members of society insist on the strict application of these roles, whereas some members rebel against these roles. In many societies throughout the world undergoing transition, women are rejecting the traditional roles of mother, housewife, and caretaker and are entering the workforce, as in Western societies. However, the process is more painstaking in these societies, and working mothers often experience strong family pressures, which may come from the grandparents or the in-laws, to be both a working mother and fulfill the traditional roles of housewife and caretaker. The structure of the traditional family in many countries provides an alternative caretaker of the child while the mother works—the grandmother.

## 6. PSYCHOLOGICAL DIMENSIONS

So far, the discussion has covered features of culture and family. Societies differ in terms of degree of complexity, from small, simple societies of a few hundred members who subsist through fishing or hunting to large, complex societies of western Europe and North America in which the economic basis is industry and services. Societies also differ in terms of family structure and function. Family type is related to the subsistence system, property and residence, and descent. This section discusses whether there are different psychological characteristics related to different family types.

As indicated previously, there are relationships between means of subsistence and family types. Agricultural and pastoral societies tend to have extended families, whereas hunting, fishing, and gathering societies tend to have nuclear societies. Agricultural societies tend to have a base, land and permanent houses, and large families and kin nearby. A large number of kin is necessary for crop cultivation. Cultivation is a process that requires patience and the cooperation of many, and it is affected by droughts, floods, and other natural forces. Studies have found that children in agricultural and pastoral societies are taught to be responsible, compliant, obedient, and to respect their elders. Hunting and

gathering societies move about in order to obtain food, hunting requires particular skills, and good hunters are respected no matter what their age. Children in these societies tend to be self-reliant, independent, and achievement oriented. It is likely that these personality traits and the social values that promote them are a function of the adaptation of the individual and the family to the ecological constraints of their environment through their means of subsistence.

In the six-cultures studies of Whiting *et al.*, relationships between the economy, social structure, settlement pattern, household and family organization, and the behavior of children were studied. Minturn and Lambert studied the effects of differences in living patterns and economic activities on child-rearing practices and the behavior of children (e.g., succorance, achievement, self-reliance, obedience, nurturance, responsibility, sociability, dominance, and aggression). Mothers were least warm in Mexico and India, where the living patterns were characterized by courtyard or semicourtyard living. In African and Philippine societies, where mothers lived in separate houses surrounding a partially enclosed yard, mothers were intermediate in warmth. Okinawan mothers, who had more privacy, were the warmest. New England mothers, who had the most privacy and almost exclusive responsibility for child care, ranked lowest in warmth. In general, children in nuclear, joint, and lineal family households were treated more warmly and were indulged when alternate caretakers, such as grandparents, sisters, and aunts, were available.

Kagitcibasi's model of family change based on socioeconomic development posits three family interaction patterns: the traditional family in developing countries characterized by total interdependence between generations in material and emotional realms, the individualistic nuclear family model of Western society based on independence, and a dialectical synthesis of these two involving material independence but psychological interdependence between generations. The parenting orientations differ significantly among these patterns, as do the distinctive characteristics of the emergent selves. With increasingly urban lifestyles and increasing affluence in collectivistic cultures, material interdependence between generations decreases because elderly parents no longer depend on the economic support of their adult offspring, but emotional interdependence continues. This leads to increased autonomy in child rearing, but together with psychological interdependence.

These changes are theorized to lead to the development of three types of the self in the three family models: the related self, high in relatedness but low in autonomy

in the family model of total interdependence; the separated self, high in autonomy but low in relatedness in the family model of independence; and the autonomous-related self, high in both relatedness and autonomy in the family model of psychological interdependence, revealed in the changing family in collectivistic societies.

The previously discussed studies refer primarily to personality and emotional variables. Family type is also related to cognitive variables. An example is Berry's ecocultural framework, which traces the relationships between ecological, cultural, and acculturational variables and cognitive styles. Initially, links were established between hunting-based and agriculture-based peoples regarding ecological demands and cultural practices and the required cognitive performances (i.e., visual disembedding and analytic and spatial abilities). These links were analyzed in detail in terms of the variables subsistence patterns, settlement patterns, population density, family type, social/political stratification, socialization practices, education, and wage employment. It was found that nomadic hunters and gatherers, with a loose social structure, nuclear family systems, and an emphasis on assertion in socialization, were relatively field independent (i.e., were able to focus on an object without being influenced by its context), whereas sedentary agriculturalists, with tight social structures, extended family systems, and an emphasis on compliance in socialization, were relatively field dependent (i.e., strongly influenced by the context of stimuli). In addition, people in societies undergoing acculturation and developing Western types of education became more field independent (Fig. 2).

## 7. CONCLUSION

This article presented a picture of the family primarily from the perspective of cultural anthropology, sociology, and cross-cultural psychology. In this brief synopsis, it was not possible to present the contributions of other social sciences, such as economics, geography, psychiatry, and political science. The primary goal was to focus on the study of family from the standpoint of cross-cultural psychology and psychological anthropology. Thus, emphasis was placed on an ecological or ecocultural perspective: the ecological context, social structure, maintenance systems, means of subsistence, the family, child development, roles, and psychological consequences. The picture may appear to be like a kaleidoscope in that there is wide cultural variety and there are many variations of families and social structure.

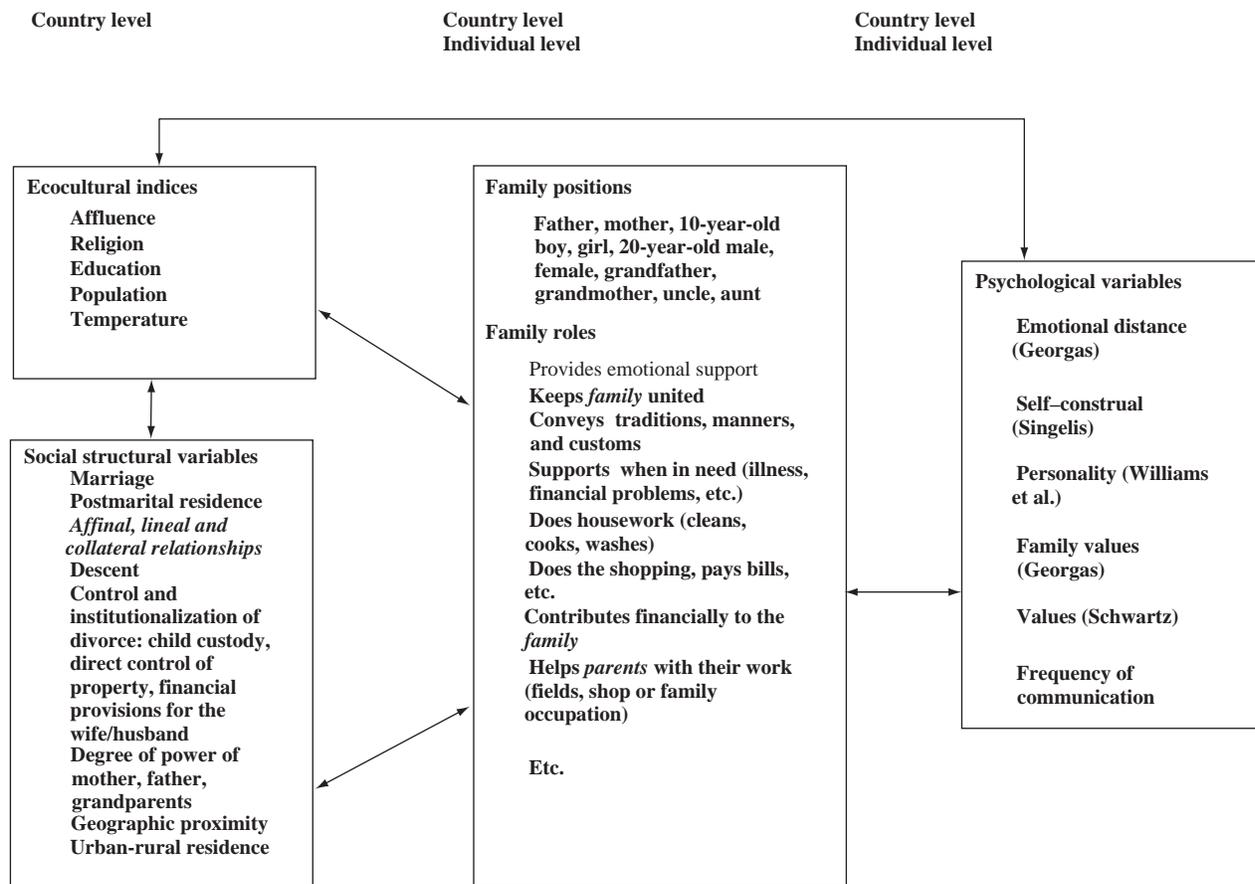


FIGURE 2 An ecocultural framework for the study of family.

However, there have been many efforts to determine the structure underneath this seemingly chaotic plethora of findings. This entails a comparative approach, consistent with the approach of psychological anthropology and cross-cultural psychology in seeking universals in family structure and function as well as variations due to cultural differences. A criticism of sociology, which has made significant contributions to the study of the family, is that it has been concerned primarily with the study of the family, particularly the nuclear family, only in the countries of northern Europe and North America.

An important issue in making sense of results from thousands of societies is the chronological dimension. That is, anthropologists have been studying small societies for more than a century. Thus, some findings refer to the study of a culture 100, 80, 40, or 30 years ago and some to recent studies. However, in a changing world in which small societies have been exposed to economic changes, technology, television, tourism, and trade with economically developed nations of the West and Asia,

these societies have also been adapting. Acculturation and enculturation in response to these pressures for change have also affected the links between ecology, social structure, family types, and psychological variables. There is a trend toward more families becoming structurally nuclear, even in small societies. However, one must be careful in employing the phrase "becoming nuclear." There is increasing evidence from studies of small societies or developing nations that the predictions of modernization theory that the driving force of modernization, the economic engine, will eventually lead to the Western type of nuclear family system may be incorrect. The number of nuclear families is increasing in many developing societies; the traditional family systems are no longer totally dependent on subsistence systems, such as hunting, gathering, or even agriculture; young couples are increasingly choosing their spouses rather than entering arranged marriages; women are increasingly entering the workforce; traditional family roles have changed; and the male is no longer the

absolute patriarch in the family. However, even though the number of nuclear families is increasing in most societies, these families still maintain very close relationships with their kin. It appears that although the structure of the family is changing from extended to nuclear, the functions of the family and the kin relations have not changed as much. Interactions and relations with kin are the psychological components of the family, which are of particular concern to psychologists.

The study of culture, family, and the psychological consequences is thus in a very dynamic phase and of considerable interest to social scientists and students.

### See Also the Following Articles

Marital Therapy

### Further Reading

- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications*. New York: Cambridge University Press.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Georgas, J., Berry, J. W., van de Vijver, F. J. R., Kagitcibasi, C., & Poortinga, Y. H. (Eds.). (2005). *Cultures, family and psychological functioning*. Cambridge, UK: Cambridge University Press.
- Ingoldsby, B. B., & Smith, S. (Eds.). (1995). *Families in multicultural perspective*. New York: Guilford.
- Levinson, D., & Malone, M. J. (1980). *Toward explaining human culture*. New Haven, CT: HRAF Press.
- Levi-Strauss, C. (1969). *The elementary structures of kinship*. London: Eyre & Spottiswoode.
- Malinowski, B. (1962). *Sex, culture, and myth*. New York: Harcourt, Brace, & World.
- Minturn, L., & Lambert, W. W. (1964). *Mothers of six cultures*. New York: Wiley.
- Munroe, R. H., Munroe, R. L., & Whiting, B. B. (1981). *Handbook of cross-cultural development*. New York: Garland.
- Murdock, P. M. (1949). *Social structure*. New York: Free Press.
- Parsons, T. (1943). The kinship system of the contemporary United States. *American Anthropologist*, 45, 22–38.
- Parsons, T. (1949). The social structure of the family. In R. N. Anshen (Ed.), *The family: Its functions and destiny*. New York: Harper.
- Popenoe, D. (1988). *Disturbing the nest—Family changes and decline in modern societies*. New York: Aldine.
- Segalen, M. (1986). *Historical anthropology of the family*. Cambridge, UK: Cambridge University Press.
- Whiting, B., & Whiting, J. W. M. (1975). *Children of six cultures: A psycho-cultural analysis*. Cambridge, MA: Harvard University Press.



# Feminist Psychology

María Inés Winkler Müller

Universidad de Santiago (USACH), Santiago, Chile

1. Historical Background
  2. Psychoanalysis and Feminist Psychology
  3. Feminist Psychology and Moral Reasoning
  4. A Feminist Ethic
  5. The Feminist Perspective in the History of Psychology
  6. Clinical Psychology and the Feminist Perspective
  7. Community Psychology and the Feminist Perspective
  8. Feminist Psychological Methodology
  9. Conclusion
- Further Reading

## GLOSSARY

*feminist ethics* A philosophical stance carrying the assumption that women and their experience have moral meaning, focusing attention on feminine ethical values such as the ethic of care, implying a critique of all discriminatory distortions, and promoting the necessary social action to get rid of all kinds of oppression.

*moral reasoning* The ways in which people analyze and resolve ethical questions.

*psychoanalytic critical posture* Theoretical stance based on women's libidinal and genital differences from men.

Feminism is mainly a political, social, and cultural perspective that is pro-women. That is, it is an ideology or an intellectual posture that attempts to explain why things are the way they are while advocating for political, social, and cultural equality between women and men. It also tries to eradicate gender discrimination in society.

Feminist psychology refers both to a collection of efforts to understand the different ways in which women and women's issues have been considered (or not considered) in psychology and to a body of knowledge about women gathered since the 1960s. Feminist psychology, also known also as psychology of women, analyzes the process through which gender is constructed, emphasizing the worth in women's perspectives and experiences. One can recognize different perspectives in feminist psychology, but all of them poignantly signal and challenge the invisibility of women in psychology. Furthermore, it is a double exclusion; women are not considered, and one does not even recognize that there is an omission. Feminist psychology challenges the validity of gender role stereotypes and expectations that have traditionally maintained patterns of male dominance and female subordination in our society. More specifically, feminist psychology disputes unquestioned premises, some research traditions, and theoretical positions that maintain a patriarchal order in which women remain secondary. At the same time, it questions those professional practices and institutional and societal structures that limit the understanding, treatment, professional achievement, and responsible self-determination of women and men.

## 1. HISTORICAL BACKGROUND

Historically, feminist psychology has emerged as a grassroots movement, but it is possible to trace its origins and

identify some landmarks in the United States, especially in the works of Weisstein and Gilligan, where they denounced prevailing sexist biases and women's non-existence in psychology. These pioneers emphasized that there were no female authors in any of the major theories of human development and psychological functioning and that women have been missing and ignored, even as research participants, at the formative stages of psychological theories. As Weisstein pointed out in 1971, "psychology is men's fantasy." That is, scientists draw their conclusions from investigating men and blindly generalize men's experiences to the experiences and lives of women but not vice versa. Weisstein's article, "Psychology Constructs the Female," challenged the predominant sexism in psychology while pointing out a double criticism of mainstream psychology: that it lacks scientific rigor and that it ignores the role played by social contexts. The lack of scientific rigor refers to the omission of the experiences of half the population (i.e., of women) in research that supports psychological theorizing. This lack of awareness about social context refers to the individualist emphasis that has characterized psychology in methodology as well as in theory construction, for example, ignoring the importance of power and gender in behavior. These criticisms are still valid, and many feminist psychologists reiterate this assertion. Today, the goal is still to do "improved science," taking into account all relevant variables and not just men's experiences. Cole's 1979 work was a pioneering study, titled *Fair Science*, that focused on sex-based discrimination and the stratification of science. Its main contribution was an increased understanding of women's challenges, opportunities, and obstacles when attempting a career in science.

Some interest and emphasis in the matters of sex differences started to appear from the beginning of psychology in the United States at the end of the 19th century and the beginning of the 20th century. Functionalism then grew as central theory strongly associated with prevailing Darwinian postures, especially with the variability hypothesis (men would have more variations among them than would women) and the functional variability hypothesis (women's intellectual performance would be affected by the menstrual cycle). These beliefs about a supposed female inferiority (despite the early disagreement and opposition found in the research done by Bradford Thompson, Puffer, Whiton Calkins, and other psychology female pioneers) prevailed for many decades as a subtext in psychological theorizing. This supposed inferiority showed itself in different social and professional fields,

and a succession of studies developed during the 1970s and 1980s revealed that sexism was clearly present at several levels. Discrimination against women hindered women's access to, and success in, certain professions. Women encountered discriminatory practices when applying for graduate training, in hiring, and in publishing their research. In addition, research methodology and practice were characterized by male bias that influenced the selection of research design, assessment measures, and sex of participants studied as well as the analysis and interpretation of data.

Based on this new awareness, and because they were interested in achieving real changes in the theory, research, and practice of psychology, Weisstein, Chesler, Evans Gardner, and others in 1970 founded American Women in Psychology, now Division 35 of the American Psychological Association, created in 1973. In Canada, some of the most important events that marked the beginning of feminist psychology were the Underground Symposium, the Canadian Psychological Association (CPA) Task Force on the Status of Women in Canadian Psychology, the founding of the CPA Interest Group on Women and Psychology, and the establishment of the CPA Status Women Committee. These and other initiatives, mainly in the academic domain, contributed to highlight the presence and experiences of women in various subfields of psychology (e.g., moral reasoning, feministic ethics, history of psychology, clinical psychology, community psychology).

## 2. PSYCHOANALYSIS AND FEMINIST PSYCHOLOGY

Outside mainstream psychology, the feminist perspectives most known are related to psychoanalytic perspectives. Using her own mothering experience, Deutsch, one of the first woman analysts in the Vienna Psychoanalytic Society, underlined the consequences of the Freudian undervaluing of women in women's developmental psychological well-being in her pioneering text, *The Psychology of Women*. Horney, a psychoanalytic feminist, criticized Deutsch for equating women with masochism, challenging the role of gender in personality development while stressing the importance of society and culture as determining factors for normal and pathological development. Contemporary stances have also embraced a critical position, particularly object relations theory as developed by Chodorow and Benjamin as well as by French psychoanalysts such as Kristeva, Cixous,

and Irigaray. The French feminist project of “writing the body” is an especially revolutionary effort to provide women’s writings with an authority based in women’s libidinal and genital differences from men. Theories about the body are particularly important for feminists because historically the body has been considered equivalent to the feminine, the female, or the woman—and, as such, has been denigrated as weak, immoral, or unclean.

### 3. FEMINIST PSYCHOLOGY AND MORAL REASONING

In regard to moral reasoning, Kohlberg, in his well-known developmental research, studied the ways in which people resolve ethical questions and proposed the postconventional stage as the highest level of moral reasoning. This level is characterized by the adult’s ability to base morality on the logic of principled decision making, which in turn is based on principles thought to be universal and not dependent on culture—the ethic of justice, in Kohlberg’s words. But given that all of his research participants were male, Kohlberg could not have taken into account the different socialization of little girls and little boys in our culture. Males are traditionally socialized to be autonomous and independent, whereas females are supposed to be passive but loving caretakers of others. Gilligan argued that these differences in socialization practices lead to different values. For men, these differences lead to a morality based on equal rights and acceptance of abstract principles, even at the sacrifice of people’s well-being. For women, these differences lead to a morality based on caring in which increasing maturity broadened the scope of their sense of responsibility and compassion. So, women have a different way—not an inferior one—of solving ethical dilemmas—the ethic of care. Olsen, in her 1978 book *Silences*, documented the history of women writers and pointed out the relative silence of women as literary voices. Gilligan, in her 1982 book *A Different Voice*, delineated one of the most well-known psychological theories about women’s development. By listening to girls and women resolve serious moral dilemmas in their lives, Gilligan formulated the development of a moral reasoning organized around concepts of responsibility and care. She argued that the experiences of inequality and subordination that outline the lives of women and girls creates a moral sense of self that is grounded in human attachments and characterized by a strong connection with relationships. Gilligan and other representatives of the Harvard Group of

Women Studies expressed it using the metaphor of “voice” and “silence.” There were two voices, or two ways of experiencing oneself in relation to others, but only one voice—the male’s—was acknowledged and delineated in the field of psychology and education. To give voice to women and listen to the unheard voice was the new motto. Gilligan recognized a deep affinity with the work of Baker Miller, and she was inspired by the radical consciousness regarding the situation of women as key for understanding the psychological order and voice central to women’s experience and women’s exercise of power. These American feminists saw women’s strength emerging from the experience of recovering an authentic voice and the capacity to express it. Conversely, women’s subordination was rooted in the silencing of their voice. But Gilligan’s work has not been free of controversies. The question of sex differences is not always considered a feminist question, and her proposal of an alternative moral system reflecting women’s experience can also be understood as a consequence of a male-biased culture.

### 4. A FEMINIST ETHIC

New developments in feminist psychology and in the domain of ethics suggest a feministic ethic that draws from women’s experience and feminist theory to describe its principles. Feminist ethics embraces much of traditional ethics but goes beyond and provides unique lenses for determining what one ought to do and how to decide on courses of action in situations involving competing ethical principles. Because moral philosophy has been largely a male enterprise, feminists seek to expose the individual and institutional practices that have negated and devalued feminine thinking. Feminist ethics assumes that women and their experience have moral meaning, and it focuses attention on feminine ethical values such as the ethic of care. This perspective also implies that a feminist critique must be accompanied by a critique of all discriminatory distortions and by the necessary social action to get rid of all kinds of oppression and not just women’s oppression.

### 5. THE FEMINIST PERSPECTIVE IN THE HISTORY OF PSYCHOLOGY

In the history of psychology, mainly U.S. and Canadian psychologists denounced the androcentric bias in the

chronicling of the discipline. As with history, sociology, literature, and other disciplines, evidence shows that in psychology women have been systematically excluded as protagonists when historicizing the discipline. The growing awareness of this absence gave rise to a body of research that enriched the field and introduced the names of several female pioneers in the historic chronicle, as shown today in registers of the American Psychological Association and Canadian Psychological Association. In the United States, Bohan, Scarborough, Felipe Russo, and Furumoto have assumed the task of repositioning women in the history of psychology. In Canada, Pike and Wright, among others, have assumed this task. Bosch and Ferrer, from Spain, have developed multidisciplinary research that undertakes to reflect about psychological thought and to identify androcentrism and misogyny in the theories of Freud, Jung, and others. They also have formulated the idea of a nonandrocentric science. The reconstruction of female participation and contributions in various countries of Latin America is an unfinished task.

## **6. CLINICAL PSYCHOLOGY AND THE FEMINIST PERSPECTIVE**

In clinical psychology, feminism has a strong hold in psychotherapy. Feminist psychotherapy emerges as a response to male dominance of mental health professions and institutions, and it objects to the role performed by the mental health establishment in maintaining social and power inequities between men and women as well as between members of the dominant culture and those of ethnic minority groups. Although feminist therapists use diverse therapeutic techniques, they have a common goal of raising awareness of the harmful effects of discrimination on clients, for example, the effects of job discrimination based on gender, race, or sexual orientation. Probably more important still is that feminist therapists dispute the tendency of traditional (male-centered) psychotherapies to catalog as normal those attributes identified with the male-dominated mainstream culture and to label as abnormal those characteristics associated with femininity and with minority cultures. Feminist therapists advocate that women's expressions of anger and distrust toward dominant group members can be seen as successful adaptive responses to prejudice and discrimination. In denouncing traditional psychotherapies

for failing to recognize the harmful effects of sexism, racism, and heterosexism, feminist therapists assert that traditional therapies categorize any behaviors that do not fit the interests of the dominant group as problematic. In so doing, traditional therapies silence women. Feminist psychology also mandates that therapists explore, recognize, and understand their own social biases; that is, they should cultivate awareness of the influence and importance of those biases in their work as therapists. Feminist theorists assume that social inequality, rather than individual psychopathology, often plays a main role in creating and maintaining many of the difficulties presented by clients in psychotherapy, especially when those clients belong to the oppressed groups. Furthermore, the main process in feminist therapy is the search for clients' sources of power rather than the search for pathologies in people. Supervision, advocacy for victims of sexual harassment, and treatment of eating disorders also constitute important topics in feminist psychotherapy where gender is of utmost importance.

## **7. COMMUNITY PSYCHOLOGY AND THE FEMINIST PERSPECTIVE**

Between community psychology and feminist psychology, we find notable convergence, not only in similar values but also in a common historical background, with both having emerged during the 1960s as part of broad movements for social justice and equality. Both perspectives share in criticizing victim-blaming ideologies and reject simplistic or restrictive standards of health, normality, and behavior. Respect for differences, distributive justice, and equality are some of the distinctive values that community psychology and feminist psychology endorse. Both perspectives dispute individual and adjustment-oriented analyses, explanations, and answers. Instead, both elaborate models and strategies emphasizing social policy, advocacy, prevention, and empowerment. In particular, the social community psychology represented by Latin American psychologists Montero and Serrano-García shares the commitment to the empowerment of the powerless, including social action and promotion of social change. Nevertheless, community psychology has not fully embraced feminist perspectives. It has been criticized for not adequately integrating women into the profession, not addressing women's issues in graduate training and

professional practice, and not incorporating feminist perspectives and processes into research, theory, and practice.

## 8. FEMINIST PSYCHOLOGICAL METHODOLOGY

In the methodological arena, feminist psychology states that more socially situated methods are needed because individualistic research methods have dominated feminist psychology and feminist research despite the theoretical command to study and understand people in context. It is suggested that the inclusion of more socially oriented methods, such as group interviews and focus groups, may help to overcome this deficiency. In addition, from a feminist ethics perspective, new questions arise about the role of power inequalities between the researcher and the researched and about the risk of the imposition of meaning. The task is to generate high-quality interactive data offering the possibility of theoretical advances through of the co-construction of meaning in which both the researcher and the researched participate.

Not only gender inspires different types of discrimination. In 1998, Burman denounced feminist psychology's exclusion of the experiences and voices of women whose cultural identities differ from those of White Western women. She included studies and reflections from contributors who lived and worked in diverse countries such as South Africa, Yugoslavia, and Puerto Rico. In these other cultural contexts, feminist psychology itself has only incipient development and has yet to influence mainstream psychology in any meaningful way.

From its beginnings, feminist psychology has continually expanded its scope of inquiry and its modes of studying gender. Over time, the frontiers that demarcate different approaches have become more visible. Postmodern feminist psychologists reject the possibility of a whole identity that includes a single true voice, and they sustain that diverse contradictory identities can have a liberatory potential. On the other hand, in 2001, Marecek delineated what she termed "feminist skepticism," an approach that draws on postmodern thought and critical theory. The principal question is whose interest and ideas are served by psychology as science, assuming that psychology is not engaged in a harmless and disinterested search for the truth. Contemporary feminist theory calls into question the

possibility of finding fixed, reliable, objective meanings that are applicable to everyone. New categories within current feminist analysis emerge, and differences among women are now in the foreground.

## 9. CONCLUSION

In synthesis, feminism has forced psychology to take a stand that includes women in theory as well as in research and practice. The introduction of women's psychology and girls' psychology as legitimate topics of study is no less important than the naming and exploring of main issues in women's lives and experiences. By now, feminist psychology has shown the power of women's voice in broadening our understanding of human development, and it is slowly succeeding in putting gender issues on mainstream psychology's agenda. Other dimensions of feminist psychology's contributions involve reconstructing research methods and priorities to study women in the context of their experiences, integrating multiple diversities into all subfields of the discipline, developing innovative approaches in psychotherapy, and transforming institutions toward being more inclusive and collaborative. Last, but not least, this perspective dictates advocacy for social action and public policies that benefit the health and well-being of both women and men.

Feminist's insights are transforming our understanding and the study of psychology and human development. Women's psychology faces future challenges—the call to address diversity among women, to promote and achieve developments in different countries and places, and to become more interdisciplinary as the discipline becomes more inclusive, especially in the task of integrating new ethnic and gender identities.

### See Also the Following Articles

Gender and Culture ■ Gender and Education ■ Gender Role Development ■ Occupation and Gender

### Further Reading

Baker Miller, J. (1987). *Toward a new psychology of women*. Boston: Beacon.

Belenky, M., Clinchy, B., Goldberger, N., & Tarule, J. (1997). *Women's ways of knowing: The development of self, voice, and mind*. New York: Basic Books.

- Bohan, J. (1995). *Re-placing women in psychology*. Dubuque, IA: Kendall/Hunt.
- Bond, M., & Mulvey, A. (2000). A history of women and feminist perspectives in community psychology. *American Journal of Community Psychology*, 28, 599–630.
- Bosch, E., Ferrer, V., & Gili, M. (1999). *Historia de la Misoginia* (History of misogyny). Barcelona, Spain: Anthropos.
- Bosch, E., Ferrer, V., & Riera, T. (2000). *Una Ciencia no Androcentrica: Reflexions Multidisciplinars* (To a non-androcentric science: Multidisciplinary reflections). Palma, Spain: Universitat de les Illes Balears.
- Brabeck, M., & Ting, K. (2000). Feminist ethics: Lenses for examining ethical psychological practice. In M. Brabeck (Ed.), *Practicing feminist ethics in psychology* (pp. 17–35). Washington, DC: American Psychological Association.
- Brown, L. (2000). Feminist therapy. In C. R. Synder, & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes and practices for the 21st century* (pp. 358–380). Toronto: Wiley.
- Burman, E. (1998). *Deconstructing feminist psychology*. London: Sage.
- Chrisler, C. J., Golden, C., & Rozee, D. P. (1995). *Lectures on the psychology of women*. New York: McGraw-Hill.
- Cole, J. (1979). *Fair science: Women in the scientific community*. New York: Free Press.
- Deutsch, H. (1944). *The psychology of women. A psychoanalytic interpretation*. New York: Grune & Stratton.
- Gannon, L., Luchetta, T., Rohes, K., Pardie, L., & Segrist, D. (1992). Sex bias in psychological research: Progress or complacency? *American Psychologist*, 47, 389–396.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Henwood, K., Griffin, C., & Phoenix, A. (1998). *Standpoints and differences: Essays in the practice of feminist psychology*. London: Sage.
- Horney, K. (1967). *Feminine psychology*. New York: Norton.
- Kristeva, J. (1986). *The Kristeva reader*. New York: Columbia University Press.
- Mahoney, M. (1996). The problem of silence in feminist psychology. *Feminist Studies*, 22, 603–625.
- Marecek, J. (2001). After the facts: Psychology and the study of gender. *Canadian Psychology*, 42, 254–267.
- Matlin, W. M. (1987). *The psychology of women*. New York: Holt, Rinehart & Winston.
- Olsen, T. (1978). *Silences*. New York: Delacourte Press.
- Pyke, S. (2001). Feminist psychology in Canada: Early days. *Canadian Psychology*, 42, 4.
- Unger, R. (1998). *Resisting gender: Twenty-five years of feminist psychology*. Thousand Oaks, CA: Sage.
- Unger, R., & Crawford, M. (2000). *Women and gender: A feminist psychology*. New York: McGraw-Hill.
- Weisstein, N. (1971). Psychology constructs the female. In V. Gornick, & B. K. Moran (Eds.), *Woman in sexist society* (pp. 207–224). New York: Basic Books.
- Worell, J. (2000). Feminism in psychology: Revolution or evolution? *Annals of the American Academy of Political and Social Science*, 571, 183–196.
- Worell, J., & Johnson, G. N. (1997). *Shaping the future of feminist psychology: Education, research, and practice*. Washington, DC: American Psychological Association.



# Forensic Mental Health Assessment

Kirk Heilbrun and Tammy Lander

Drexel University, Philadelphia, Pennsylvania, USA

1. Introduction
  2. Context and Principles of FMHA
  3. Clinical–Legal Issues
  4. Important Developments
  5. Future Directions
  6. Conclusions
- Further Reading

## GLOSSARY

*forensic mental health assessment* The evaluation of defendants or litigants, usually conducted by mental health professionals such as psychologists, psychiatrists, or social workers, that is performed to assist the court in making better informed decisions or to help the attorney present evidence relevant to his or her case.

John Muhammad and Lee Boyd Malvo were charged in the offenses that became known as the “Washington, DC, sniper shootings.” Both were charged with capital (death penalty) offenses and tried separately. What forensic mental health assessment might be important in such cases? Possibilities include juvenile transfer to criminal (adult) court, competence to stand trial, competence to proceed as one’s own attorney (Muhammad briefly attempted to act as his own attorney), sanity at the time of the offense (Malvo entered a plea of not guilty by reason of insanity), and aggravating and mitigating factors in capital sentencing (both Muhammad and

Malvo were convicted of capital murder). This article will help you understand the nature of forensic mental health assessment, the kinds of questions it tries to answer, as well as the limits of its applicability and usefulness.

## 1. INTRODUCTION

The field of forensic psychology has developed almost entirely during the past 50 years. The broad meaning of “forensic psychology” might include the research or professional activities of any psychologist—developmental, social, cognitive, human experimental, or clinical—when performed in anticipation that the work might be used by a judge or attorney as evidence to promote better informed legal decision making. However, this article focuses more specifically on one aspect of forensic psychology: assessment. This aspect, forensic mental health assessment (FMHA), refers to the evaluation of defendants or litigants by psychologists (or other mental health professionals such as psychiatrists or social workers) that is performed to assist the court in making better informed decisions or to help the attorney present evidence relevant to his or her case.

Described in this way, forensic psychology is a fairly new specialty within the broader field of psychology. Psychologists have not always been welcomed by the courts. In an important 1962 case, *Jenkins v. United States*, a federal appellate court ruled that psychologists with appropriate training and expertise were qualified

to offer expert testimony on questions regarding mental disorder. Prior to *Jenkins*, courts typically required mental health experts to have a medical degree. Since the *Jenkins* decision, however, there has been a steady increase in psychologists' involvement in legal questions requiring mental health expertise.

Just how frequently do forensic psychologists provide FMHA? We do not have good studies that would answer this question, but the estimated frequency for certain kinds of evaluations, such as criminal defendants' competence to stand trial, is more than 25,000 annually, according to Otto and Heilbrun. Other common legal questions that may be informed by FMHA include competence to waive Miranda rights, sanity at the time of the offense, sentencing (capital and noncapital), juvenile commitment, juvenile waiver to criminal court and reverse waiver back to juvenile court, child custody, and personal injury. Psychologists are involved in conducting many of the evaluations performed in these areas.

## 2. CONTEXT AND PRINCIPLES OF FMHA

### 2.1. Sources of Authority in Forensic Assessment

#### 2.1.1. Courts, Statutes, and Case Law

The law is the first place forensic psychologists look in structuring FMHA. Many legal questions are defined, to some extent, by existing statutes. When the criteria for a given legal decision are described in a statute or related evidence code or administrative code, these criteria can be used in the report and possible testimony. This use indicates that the psychologist is both familiar with the relevant law and has structured the evaluation around it, thus making the evaluation more relevant to the proceedings. For any given issue, it is up to the court to answer the ultimate legal question, so it is not the most desirable focus for the report. However, it is helpful to show that the evaluator is aware of the legal standard and can use the defining criteria to structure the report.

Other sources of legal authority may also be helpful, although they may be less directly relevant to the legal criteria used in the report. Appellate case law (particularly cases from the U.S. Supreme Court) and "model" law (e.g., the *Criminal Justice Mental Health Standards*, published in 1989 by the American Bar Association) can provide useful background on the sources of and

reasoning behind the current legal standard in the jurisdiction in which the evaluation is being conducted.

#### 2.1.2. Ethical Standards

Forensic psychologists rely on two major sources of ethical authority. First, the *Ethical Principles of Psychologists and Code of Conduct* (Ethics Code), published by the American Psychological Association (APA) in 2002, serves as the strongest authority for ethics in psychology. Those conducting FMHA are ethically obligated to do so in a way consistent with the ethics code and can be sanctioned by the APA for violating it. A second important source of guidance is the *Specialty Guidelines for Forensic Psychologists*, published by the Committee on Ethical Guidelines for Forensic Psychologists in 1991. This document was written for those who regularly conduct FMHA, and it provides more specific guidance for those in legal settings. Finally, and most specifically, there are documents such as the "Guidelines for Child Custody Evaluations in Divorce Proceedings," published by the APA in 1994, that offer direction for particular kinds of FMHA such as child custody evaluation.

#### 2.1.3. Scientific Literature

FMHA is based on and informed by science. Studies in areas such as mental health disorders, behavioral functioning, violence and antisocial behavior, response to treatment, and the assessment of specific functional legal capacities are all important to the development of FMHA procedures and the practice of FMHA in particular cases. Under the 1993 holding of the U.S. Supreme Court in *Daubert vs. Merrell Dow Pharmaceuticals, Inc.*, and the subsequent 1999 Supreme Court decision in *Kumho Tire Company, Ltd. vs. Carmichael*, FMHA evidence is weighed by the trial court judge for admissibility based in part on whether it is based on procedures consistent with science. (Not all states have the *Daubert* standard for admissibility of expert evidence; some continue to use the standard described in the 1923 *Frye vs. United States* case, in which a technique must be sufficiently established to have gained general acceptance in the field to which it belongs to yield admissible expert evidence.)

#### 2.1.4. Professional Literature

Finally, there is a literature that might best be described as "professional," describing recommended approaches to FMHA based on experience and emerging or

established standards in the field. Often such literature appears in practice-oriented journals or books, particularly those that are interdisciplinary. Leading interdisciplinary journals in forensic psychology include *Law and Human Behavior*, *Behavioral Sciences & the Law*, *Criminal Justice and Behavior*, *International Journal of Forensic Mental Health Services*, and others focusing both on empirical studies and on more practice- or theory-oriented approaches to FMHA. This literature forms another important source of authority and guidance for the forensic psychologist.

## 2.2. Principles of FMHA

Evaluations conducted by psychologists are typically performed to help reach a diagnosis or problem formulation and to plan appropriate treatment (we might call these evaluations “therapeutic assessment”). However, FMHA differs from this usual process and goals in some important ways. Heilbrun pointed out the following 10 differences (Table I) in his 2001 book, *Principles of Forensic Mental Health Assessment*:

TABLE I  
Principles of Forensic Mental Health Assessment<sup>a</sup>

---

### Preparation

1. Identify relevant forensic issues.
2. Accept referrals only within area of expertise.
3. Decline the referral when evaluator impartiality is unlikely.
4. Clarify the evaluator’s role with the attorney.
5. Clarify financial arrangements.
6. Obtain appropriate authorization.
7. Avoid playing the dual role of therapist and forensic evaluator.
8. Determine the particular role to be played within the forensic assessment if the referral is accepted.
9. Select the most appropriate model to guide data gathering, interpretation, and communication.

### Data collection

10. Use multiple sources of information for each area being assessed.
11. Use relevance and reliability (validity) as guides for seeking information and selecting data sources.
12. Obtain relevant historical information.
13. Assess clinical characteristics in relevant, reliable, and valid ways.
14. Assess legally relevant behavior.
15. Ensure that conditions for evaluation are quiet, private, and distraction free.
16. Provide appropriate notification of purpose and/or obtain appropriate authorization before beginning.
17. Determine whether the individual understands the purpose of the evaluation and the associated limits on confidentiality.

### Data interpretation

18. Use third-party information in assessing response style.
19. Use testing when indicated in assessing response style.
20. Use case-specific (idiographic) evidence in assessing clinical condition, functional abilities, and causal connection.
21. Use nomothetic evidence in assessing clinical condition, functional abilities, and causal connection.
22. Use scientific reasoning in assessing causal connection between clinical condition and functional abilities.
23. Do not answer the ultimate legal question.
24. Describe findings and limits so that they need change little under cross-examination.

### Communication

25. Attribute information to sources.
  26. Use plain language; avoid technical jargon.
  27. Write report in sections, according to model and procedures.
  28. Base testimony on the results of the properly performed FMHA.
  29. Testify effectively.
- 

<sup>a</sup>Adapted from Heilbrun (2001).

1. The purpose for which the evaluation is conducted
2. The nature of the relationship between the evaluator and the individual being evaluated
3. The need for a formal notification of purpose
4. Who is being served
5. The nature of the standard(s) applied in the evaluation
6. The sources of data considered
7. The response style of the individual being evaluated
8. The need to clarify the reasoning and limits of knowledge
9. The nature of the written report
10. The expectation of expert testimony

When these differences are considered, it becomes very clear that FMHA is not simply psychological assessment that is conducted in a legal setting. Instead, it is a specialized form of psychological assessment that takes into account that there may be very different questions asked, standards applied, motivations observed, and approaches to communicating results.

General principles that describe therapeutic assessment do not necessarily apply to FMHA, so it is helpful to review the FMHA principles described by Heilbrun in 2001. These principles, discussed in the following section, are summarized and elaborated in how they might apply in FMHA in [Table II](#).

These 29 FMHA principles each apply, to some extent, to the many forms of FMHA. Heilbrun *et al.* provided examples of case reports on different legal questions to which these principles can be applied, so it seems that both theoretically and practically these principles are potentially useful. They are reviewed in the remainder of this section, with a comment about each principle and the associated general guideline:

1. *Identify relevant forensic issues.* This principle distinguishes between the legal question, to be decided by the judge or the jury in the case, and the forensic issues, which are the capacities and abilities included within the legal question. These capacities and abilities can be described and measured more readily than the broader legal question, which typically contains questions of both science (e.g., Did the defendant have a mental illness?) and values (e.g., Was the individual sufficiently disturbed to be unable to know the wrongfulness of his behavior?). The associated guideline suggests that both the legal question and the forensic issues be described in the first section of the report.

2. *Accept referrals only within area of expertise.* Expertise in FMHA has two components: (i) clinical

and didactic training and experience with individuals similar to the one being evaluated and (ii) forensic training and experience in evaluating similar individuals and questions. The forensic psychologist can demonstrate both kinds of expertise by submitting a professional resume (a *curriculum vitae*) when requested by the lawyers or the judge and by displaying professional credentials, such as degree, licensure, and board certification, on the letterhead used for the report.

3. *Decline the referral when evaluator impartiality is unlikely.* The evaluator's impartiality is important in FMHA. If the psychologist has personal beliefs or experiences circumstances that would interfere with his or her ability to be fair and even-handed in collecting data and drawing conclusions, then this principle and guideline suggest that the forensic clinician should decide not to conduct the FMHA in this particular case. For example, psychologists with strong feelings about the death penalty that they cannot set aside in a particular case should decline to conduct a capital sentencing evaluation.

4. *Clarify the evaluator's role with the attorney.* When psychologists conduct an FMHA, they may do so upon order of the judge or by request of one of the attorneys involved in the litigation. A psychologist may also serve as a consultant to one of the attorneys, without any expectation that a report or expert testimony will result. These roles have differing expectations for who will receive the results of the evaluation or consultation and in what circumstances. These expectations stem from relevant evidence law, applicable ethical guidelines, and professional practice standards. It is important, though, that the psychologist make it clear which role is being played so the appropriate expectations will guide his or her professional conduct. When roles are unclear or merged, there can be substantial problems with the evaluation or consultation.

5. *Clarify financial arrangements.* Psychologists are paid for conducting FMHA as they would be for any professional service. The details of such payment, however—the rate or total amount, who will be responsible for payment, and whether there are special arrangements to be made (such as a retainer or an escrow account)—should be agreed upon by all parties before the FMHA is begun. If the payment process is established under law or policy, then arrangements are not needed; the evaluator simply bills the court, the government, or whoever is designated under policy for the purpose. In other instances, however, the financial arrangements must be discussed and clarified in the beginning.

**TABLE II**  
**Principles and Guidelines for Application of FMHA Principles in Forensic Contexts<sup>a</sup>**

<i>Principle</i>	<i>General guidelines for application in FMHA</i>
Identify relevant forensic issues.	Cite legal question and include forensic issues in first section of report.
Accept referrals only within area of expertise.	Give degree and licensure, board certification status. Provide curriculum vitae and/or summary of qualifications if requested.
Decline the referral when evaluator impartiality is unlikely.	Avoid involvement in cases in which there is substantial incentive for the forensic clinician (personal, professional, or monetary) to have case decided in a particular direction.
Clarify the evaluator's role with the attorney.	Ensure that both the forensic clinician and the referral source are clear on whether the clinician will serve as court-appointed evaluator, attorney-requested evaluator, or consultant.
Clarify financial arrangements.	Ensure that terms of payment for evaluation are understood by both the forensic clinician and the party responsible for payment.
Obtain appropriate authorization.	Cite basis for evaluation request (e.g., court ordered or attorney requested). Describe whether informed consent was obtained, if evaluation was not court ordered.
Avoid playing the dual roles of therapist and forensic evaluator.	Minimize the frequency of this combination. If such roles are combined, it should be with explicit justification, advance planning, and clear notification to the individual involved.
Determine the particular role to be played within forensic assessment if the referral is accepted.	If report will be submitted into evidence, evaluator should be <i>impartial</i> —tone of report should reflect this.
Select the most appropriate model to guide data gathering, interpretation, and communication.	Use the Morse model (mental disorder, functional abilities, and causal connection) or the Grisso model (functional, contextual, causal, interactive, judgmental, and dispositional characteristics).
Use multiple sources of information for each area being assessed.	Obtain self-report, psychological testing data, third-party interviews, and collateral records data.
Use relevance and reliability (validity) as guides for seeking information and selecting data sources.	Use data sources with demonstrated reliability and validity (when this has been researched) and that will provide information relevant to the area being assessed.
Obtain relevant historical information.	In a separate section, document the individual's history and previous functioning in areas relevant to current clinical condition and functional legal capacities.
Assess clinical characteristics in relevant, reliable, and valid ways.	Describe clinical characteristics using measures that are reliable, valid for the purpose used, and/or weighed against information from collateral sources.
Assess legally relevant behavior.	Document information collected from multiple sources regarding the individual's functional legal capacities.
Ensure that conditions for evaluation are quiet, private, and distraction free.	Note any deviation from reasonably quiet, private, and distraction-free conditions. Describe impact on data collected.
Provide appropriate notification of purpose and/or obtain appropriate authorization before beginning.	Describe elements of notification of purpose or informed consent given to individual being evaluated and to third parties who are interviewed.

*Continues*

Continued

Principle	General guidelines for application in FMHA
Determine whether the individual understands the purpose of the evaluation and the associated limits on confidentiality.	Document how the individual's understanding was assessed and to what extent he or she understood the relevant information.
Use third-party information in assessing response style.	Describe the consistency of third-party information with self-reported information, and be particularly cautious about self-report when it is significantly different from third-party accounts.
Use testing when indicated in assessing response style.	Administer a test(s) sensitive to response style, particularly when there is concern about the accuracy of self-report.
Use case-specific (idiographic) evidence in assessing clinical condition, functional abilities, and causal connection.	Describe the individual's clinical condition and functional legal abilities in the context of his or her history of symptoms and demonstrated capacities.
Use nomothetic evidence in assessing clinical condition, functional abilities, and causal connection.	Describe the results of psychological tests, structured instruments, and specialized tools validated for assessing (i) clinical condition or (ii) functional legal capacities.
Use scientific reasoning in assessing causal connection between clinical condition and functional abilities.	Describe explanations for clinical condition and functional abilities that have the most supporting evidence and least disconfirming evidence.
Do not answer the ultimate legal question.	When evidence is mixed, or competing explanations seem comparably well-supported, say so.
Describe findings and limits so that they need change little under cross-examination.	Present conclusions about forensic capacities but not the larger legal question(s).
Attribute information to sources.	Be careful, impartial, and thorough in presenting data and reasoning.
Use plain language; avoid technical jargon.	Consider alternative explanations.
Write report in sections, according to model and procedures.	Describe data so that the source(s) of any specific finding is clear.
Base testimony on the results of the properly performed FMHA.	Make minimal use of technical language, and define technical terms when they must be used.
Testify effectively.	Include sections on referral information, sources of information, relevant history, clinical functioning, relevant functional legal capacities, and conclusions.
	Describe causal relationship between clinical symptoms and functional legal capacities.
	Master the contents of the report, which contains thorough documentation of evaluation, and use report contents to guide testimony.
	Use effective style in presenting substantive FMHA findings in testimony.

<sup>a</sup>Adapted from Heilbrun (2001).

6. *Obtain appropriate authorization.* The nature of the authorization needed to conduct the evaluation varies, depending on who requested it and the role being played by the evaluator. The forensic clinician needs a signed order from the court for a court-ordered evaluation. Consent from both the referring attorney and the individual being evaluated is necessary when

the forensic clinician is in the role of expert for the defense/prosecution/plaintiff.

7. *Avoid playing the dual roles of therapist and forensic evaluator.* Serving as both therapist and forensic evaluator for the same individual should be avoided because it can create significant problems for the psychologist, the person being evaluated and treated, the

lawyers, and the judge. Occasionally, such a dual role cannot be avoided. When it cannot, the guideline emphasizes the importance of making the reason clear, doing the planning in advance, and informing the affected individual(s) in plain language.

8. *Determine the particular role to be played within forensic assessment if the referral is accepted.* If the forensic psychologist respects the previous principle, then most FMHA cases will involve only a single role. The psychologist must identify this role first. When it involves writing a report and possibly testifying, the psychologist must strive to be impartial, focusing only on the questions that are requested. In addition, this role should be maintained throughout the entire case.

9. *Select the most appropriate model to guide data gathering, interpretation, and communication.* Stephen Morse, in 1978, and Thomas Grisso, in 1986, each described a model to guide the ideas, procedures, interpretation of results, and reasoning that are important components of FMHA. Each model helps the psychologist describe the relationship between clinical symptoms and functional legal capacities, making it easier for the reader to understand how the individual being evaluated might meet the applicable legal test. They also help describe the deficits that keep him or her from doing so and the source of these deficits when the individual does not.

10. *Use multiple sources of information for each area being assessed.* A single source of information is usually less accurate than multiple sources considered together. This is particularly applicable in FMHA because some of the information sources can be particularly distorted. A defendant might be inclined to exaggerate or minimize symptoms or capacities, depending on the possible implications of each in litigation. A psychological test might not be sensitive to such deliberate exaggeration or minimization. An individual who was interviewed to provide more information about the defendant might have some bias against or for the defendant. When multiple sources are considered, however, the psychologist can be more sensitive to particular distortions from any single source and better gauge the overall accuracy through considering consistency across sources.

11. *Use relevance and reliability (validity) as guides for seeking information and selecting data sources.* Whether evidence will be admitted into a legal proceeding depends on its relevance to the matter before the court and its reliability (accuracy). Therefore, this principle emphasizes both relevance and reliability in sources of information that are part of FMHA. Some

sources of information, such as psychological tests and structured interviews, have measured levels of reliability and validity as part of their development. It is fairly straightforward to select such tests and interviews with higher levels of reliability and validity. Other sources of information, such as the interviews with the individual being evaluated and with collateral observers, may be more difficult to gauge in their validity—but may be selected because they are so clearly relevant to the issues being assessed.

12. *Obtain relevant historical information.* Relevant historical information about the individual being evaluated is almost always important in FMHA. The level of detail needed varies, however, according to the type of FMHA. Some evaluations, such as those of competence to stand trial, are present-state oriented and need a more focused history in areas such as criminal and mental health history. Other evaluations, such as capital sentencing, are much broader and demand a history that is much wider in scope.

13. *Assess clinical characteristics in relevant, reliable, and valid ways.* The decision about which symptoms and problems should be assessed, and how to assess them, is informed by using relevance and reliability, according to this principle. This guideline calls for using appropriate measures, such as psychological tests and structured interviews, and considering these results in conjunction with information from collateral sources in order to assess the relevant clinical aspects under consideration.

14. *Assess legally relevant behavior.* There are particular capacities and behaviors that are part of the specific legal question(s) being considered by the court. These capacities and behaviors (sometimes called functional legal criteria) are a very important part of FMHA. For some kinds of evaluations, there are validated measures that can be used to help the evaluator learn how the individual being evaluated compares with other such individuals on certain relevant legal capacities. Such measures have been developed for competence to stand trial, competence to consent to treatment, and competence to waive Miranda rights. However, for many legal questions, there has not yet been a validated measure developed. Therefore, evaluators must identify the functional legal demands and obtain information about the individual's thinking, feeling, and behavior at the appropriate time, when they do not have a validated measure available.

15. *Ensure that conditions for evaluation are quiet, private, and distraction free.* FMHA is often conducted in detention centers, jails, prisons, or secure

psychiatric hospitals. Security is important in these settings; staff are responsible for preventing escape, harm to others, or harm to self. The need for evaluation conditions that are quiet and distraction free (without significant noise or visual distraction) as well as private (in which communication cannot be overheard by a third party) must therefore compete with security, sometimes in the context of a physical plant that was not designed to facilitate privacy or quiet. This principle emphasizes that the FMHA cannot be performed meaningfully unless there is a certain degree of privacy and limit on distraction afforded during the evaluation period.

16. *Provide appropriate notification of purpose and/or obtain appropriate authorization before beginning.* The forensic psychologist provides a notification of purpose or obtains informed consent before starting the evaluation. Informed consent is needed when an attorney retains a forensic psychologist to evaluate the attorney's client because the evaluation is legally voluntary and the client is under no legal obligation to participate. However, a client undergoing a court-ordered evaluation does not have a legal right to refuse to participate. An evaluatee in these circumstances should receive the same information that would be provided as part of obtaining informed consent—the nature and purpose of the evaluation, its authorization by the court, and the associated limits on confidentiality, including how it might be used. However, since there is no legal right to refuse, the necessity for consent is eliminated when a court-ordered evaluation is conducted.

17. *Determine whether the individual understands the purpose of the evaluation and associated limits on confidentiality.* If the information provided in the notification of purpose or request for informed consent is not understood, or is only partly understood, then the process is less meaningful. The forensic psychologist must gauge how well such information is understood by the individual being evaluated. To the extent it is not understood, the psychologist should describe the limits on understanding and what deficits interfere with such understanding.

18. *Use third-party information in assessing response style.* Some individuals evaluated through FMHA respond to questions openly and honestly. Others exaggerate or fabricate symptoms and problems. Yet others minimize or deny symptoms that are genuinely experienced. Response style is a term used to describe how an individual reports his or her own symptoms, behaviors, and problems, and it is used to classify an

individual's responding as (i) reliable, (ii) exaggerated/malingering, (iii) defensive, or (iv) uncooperative. Self-reported information from an individual who is not reliable will have significant inaccuracies. This principle emphasizes the value of third-party information (records and interviews of collateral observers) in both detecting and correcting these inaccuracies.

19. *Use testing when indicated in assessing response style.* In addition to third-party information, there are also tests that have been developed and validated to assess response style. These are reviewed by Rogers in his 1997 book titled *Clinical Assessment of Malingering and Dissimulation*. This principle advocates using tests and specialized measures whenever available and appropriate for assessing malingering or defensiveness.

20. *Use case-specific (idiographic) evidence in assessing clinical condition, functional abilities, and causal connection.* Science is an important part of FMHA. The first way it can contribute is through promoting the gathering of individualized information, specific to the case and present functioning of the individual, and comparing it to that individual's capacities and functioning at other times. This approach is consistent with the law's goal of individualized justice, so it may appeal particularly to judges and attorneys.

21. *Use nomothetic evidence in assessing clinical condition, functional abilities, and causal connection.* Science can also contribute to FMHA through empirical data collected with groups of individuals similar to the person being evaluated and by using tests developed and validated on similar groups. Such tests facilitate the reliable measurement of relevant abilities and symptoms and allow their comparison with those in "known groups." For example, a validated test to measure capacities relevant to competence to stand trial would allow the psychologist to compare an individual's performance on a given capacity (e.g., understanding the adversarial nature of the legal system) to that of other known groups, such as (i) defendants who had been adjudicated incompetent to stand trial, (ii) defendants for whom the issue had been raised but were not adjudicated incompetent, and (iii) defendants for whom the issue had never been raised.

22. *Use scientific reasoning in assessing causal connection between clinical condition and functional abilities.* Scientific reasoning, as contrasted with scientific data, can also contribute to FMHA. The results obtained from one source of information (such as the interview) can be treated as "hypotheses" to be supported or

rejected through further information obtained from other sources of information. Hypotheses that account for the most information with the simplest explanation are likely to be the most accurate.

23. *Do not answer the ultimate legal question directly.* Many in the field of forensic psychology have debated whether experts should directly answer the “ultimate legal question” before the court. Some have argued that this is permitted under the law of evidence in most jurisdictions, that many judges seem to want it, and that there is little harm in doing so. Others have suggested that ultimate legal questions (such as “Is the defendant competent to stand trial?”) have embedded moral and community values to a greater extent than do functional legal capacities (such as whether the defendant understands his charges and can assist in his defense). Thus, answering ultimate legal questions directly intertwines clinical and scientific conclusions with moral values, possibly to the confusion of the judge or jury. This principle suggests that functional capacities rather than ultimate legal questions should be the focus of FMHA.

24. *Describe findings and limits so that they need change little under cross-examination.* FMHA should be a careful and thorough process, incorporating multiple sources of information and accompanied by acknowledgment of the limits on data accuracy and consistency. Using this approach, the forensic clinician can expect that the description of findings and conclusions will not change significantly during cross-examination.

25. *Attribute information to sources.* All information in the FMHA report should be attributed by a specific source(s). This allows the evaluator to describe consistency across sources. It also means that the judge and opposing counsel can directly assess the credibility of specific information by source.

26. *Use plain language; avoid technical jargon.* The “consumers” of FMHA, such as judges, attorneys, and jurors, usually do not have advanced training in psychology or other behavioral science. Thus, it is useful for FMHA results to be communicated without the use of technical jargon. When important points or concepts are conveyed in nontechnical language, it is usually possible to capture their essential meaning. Occasionally, a technical term may be needed, but a parenthetical definition can help ensure that it is understood by the reader or listener in the way its meaning is intended.

27. *Write reports in sections, according to model and procedures.* There is a great deal of information in an FMHA report. Without an organizing structure for the report, the reader can become lost or fail to understand some of the important aspects of the report. One

organizing structure for reports was suggested by Heilbrun in his 2001 book *Principles of Forensic Mental Health Assessment*:

1. Referral: with identifying information concerning the individual, his or her characteristics, the nature of the evaluation, and by whom it was requested or ordered

2. Procedures: times and dates of the evaluation, tests or procedures used, different records reviewed, and third-party interviews conducted, as well as documentation of the notification of purpose or informed consent and the degree to which the information was apparently understood

3. Relevant History: containing information from multiple sources describing areas important to the evaluation

4. Current Clinical Condition: appearance, mood, behavior, intellectual functioning, thought, and personality

5. Forensic Capacities: varying according to the nature of the legal questions

6. Conclusions and Recommendations: addressed toward the relevant capacities rather than the ultimate legal questions

28. *Base testimony on the results of the properly performed FMHA.* Contrary to what is often perceived, most of FMHA does not result in expert testimony. Often, reports are accepted and presented in evidence without such testimony. Sometimes, reports may not be used, depending on the circumstances involved in requesting the report (e.g., court ordered vs attorney requested) and whether the attorney has that choice under the evidence law in the jurisdiction. However, when expert testimony is provided, it should be firmly based on the results of a thorough evaluation documented in the FMHA report. This allows the referring attorney to present the expert’s findings effectively, the opposing attorney to challenge them, the judge or jury to understand them, and the expert to communicate them.

29. *Testify effectively.* Effective expert testimony combines strong substance with useful style. The substance of FMHA has been discussed in the previous 28 principles. The style of expert testimony involves appearing knowledgeable, credible, likeable, interesting, and professional. Without both substance and style, the expert testimony provided by a forensic psychologist cannot be maximally effective. Strategies and suggestions for improving style while testifying have been documented by Brodsky in his 1991 book, *Testifying in Court*:

*Guidelines and Maxims for the Expert Witness*, and his 1999 book, *The Expert Expert Witness: More Maxims and Guidelines for Testifying in Court*.

### 3. CLINICAL–LEGAL ISSUES

#### 3.1. Criminal

Forensic psychologists may provide FMHA that is focused on a variety of legal questions in criminal, civil, and juvenile/family areas. These are described in detail in the 1997 book by Melton *et al.* titled *Psychological Evaluations for the Courts*. In some cases, only one of these questions will be addressed in a single FMHA; in other instances, the forensic psychologist may be asked to address two or three questions in a single evaluation. Some of the most commonly evaluated legal questions evaluated through FMHA for adult criminal defendants include the following:

1. Competence to waive Miranda rights—whether the defendant waived his or her Fifth Amendment right to remain silent and Sixth Amendment right to counsel in a knowing, voluntary, and intelligent fashion when questioned by the police while in custody

2. Competence to stand trial—whether the defendant understands his or her legal situation and can assist counsel in his or her own defense

3. Sanity at the time of the offense—the particular test varies across jurisdictions; in many states, it involves whether the defendant did not know the nature or wrongfulness of criminal behavior as a result of mental disease or defect; in others, there is the additional question of whether the defendant was unable to conform his or her conduct to the requirements of the law as a result of mental disease or defect

4. Non-capital sentencing—often involving the broad questions of the defendant’s risk for recidivism and rehabilitation needs and amenability; in specific kinds of cases, such as those involving sexual offenders, the questions are focused particularly on the need for specialized treatment

5. Capital sentencing—whether the defendant has characteristics, symptoms, or history specified in the aggravating and mitigating factors outlined in the law of the jurisdiction

#### 3.2. Civil

In addition to FMHA in criminal contexts, forensic psychologists conduct assessments for judges and

attorneys on questions within civil law. Two of the most frequent legal questions for FMHA in the civil area concern (i) civil commitment (whether an individual, because of mental illness, meets the criteria for involuntary hospitalization as presenting a danger to self or others) and (ii) personal injury [often whether an individual, allegedly the victim of a tort (or noncriminal wrong), suffered psychological injury as a consequence, and what might be needed to treat this injury].

#### 3.3. Juvenile/Family

There are several legal questions within the juvenile and family area that are often addressed through FMHA. These include (i) juvenile commitment (whether a youth adjudicated as delinquent should be committed for residential placement or can be served in a less restrictive intermediate placement such as boot camp or in a community-based placement; often involving the criteria of reoffense risk and treatment/rehabilitation needs and amenability), (ii) juvenile transfer and reverse transfer (whether an adolescent charged with a crime should be handled in juvenile court or adult criminal court, with the criteria for assessment often similar to those for juvenile commitment), and (iii) child custody (in divorce cases, the question of whether custody should be awarded solely to one of the divorcing parents or jointly can be informed by FMHA addressing the needs of the children and the parenting capacities of each parent).

### 4. IMPORTANT DEVELOPMENTS

As forensic psychology has matured, there have been important developments in the areas of training, research, and practice that may enhance the quality of FMHA that is conducted by forensic psychologists.

#### 4.1. Training

Specialized training is crucial to the goal of high-quality forensic practice. One important development in this area during the past decade has been the approval of forensic psychology as a formal applied specialization within the APA. Forensic psychology joins the areas of clinical, counseling, school, behavioral, clinical health, clinical neuropsychology, clinical child, family psychology, and psychoanalysis in psychology as approved APA specializations. This will promote the development and structure of formal postdoctoral training in forensic

psychology because postdoctoral programs will be able to apply for APA accreditation with a specialization in forensic psychology.

There are a number of such postdoctoral fellowship programs in forensic psychology today, which is another important development in the training area. With a total of 16 fellowships available nationally, those training to become forensic psychologists have the opportunity to receive broad-based training in clinical or counseling psychology prior to the doctoral degree followed by specialized forensic training during the fellowship year. Accreditation of such fellowships should further enhance the quality and uniformity of training in FMHA.

However, not all psychologists wishing to provide FMHA have had the opportunity to receive such specialized training. For those who received the doctoral degree before 1980, postdoctoral fellowship training was not available, nor was formal training during internship or doctoral training available except in rare instances. The role of continuing education in the form of workshops presented to practicing professionals has therefore been quite important. Some continuing education is provided by university–state partnerships, as in the state of Virginia. Other continuing education programs are offered by specialty organizations, such as the American Academy of Forensic Psychology. Such continuing education efforts have filled a gap between formal education and experience that was important to more senior psychologists conducting FMHA but have also been helpful in providing relevant research and new tools and approaches to even those who are already largely specialized in forensic.

#### 4.2. Research

The development of specialty tools has been very important in the maturation of the field. Such tools, developed and validated through scientific research and designed to measure functional legal capacities described in the law, have allowed forensic psychologists to consider individualized impressions and compare the individual being evaluated to others in similar circumstances. Tools have been developed during the past decade to measure constructs such as those relevant to competence to waive Miranda rights, competence to stand trial, competence to consent to treatment, violence risk assessment, and response style. As a consequence, FMHA on these issues can be conducted with more sophistication and scientific support than was possible before the development of these specialized, validated FMHA tools.

Foundation support of research networks has been a second significant development in research relevant to FMHA. In particular, the John D. and Catherine T. MacArthur Foundation sponsored the MacArthur Research Network on Mental Health and Law, which developed and carried out important programmatic research in the areas of violence risk, competencies, and coercion in the decade between the late 1980s and the late 1990s. Currently, the MacArthur Foundation sponsors two additional research networks—one on mandated community treatment and the second on juveniles—that have also designed and conducted important research that will enhance the quality and scientific foundation of FMHA conducted in these areas in the future.

One of the areas that have witnessed the most active research and development during the past decade is risk assessment of violent behavior. The field has moved from the relatively primitive approach to violence risk seen in the 1970s and early 1980s to a far more sophisticated, theoretically coherent, and empirically supported process. The research conducted by the MacArthur Research Network on Mental Health and Law is described in the 2001 book by Monahan and colleagues titled *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence*. In addition to this research, other works have yielded a number of tools assessing both risk factors for violence and needs for risk reduction (sometimes called risk-needs tools). As a consequence of this research, practitioners of FMHA can now address violence risk and risk reduction needs in those they evaluate in a way that is far more coherent and empirically supported than was ever possible before.

#### 4.3. Practice

Finally, there have also been important developments in the practice of forensic psychology, particularly FMHA. Some states (e.g., Massachusetts) have developed training curricula for practicing psychologists in the forensic area and developed a formal certification process for those wishing to become certified examiners in FMHA. More advanced individual certification has been available since the late 1970s from the American Board of Forensic Psychology, which has continued to offer a rigorous board certification examination to those who meet the criteria for training and experience and wish to be certified as forensic practitioners at an advanced level. The development and publication of specialty guidelines for the practice of FMHA was an additional important step. The Committee on Ethical Guidelines for Forensic Psychologists, which published the *Specialty Guidelines*

for *Forensic Psychologists* in 1991, effectively provided the field with standards that were consistent with the APA ethics code but more specific to forensic contexts, allowing both psychologists and legal professionals a view of aspirational standards of practice in forensic psychology.

## 5. FUTURE DIRECTIONS

The field of forensic psychology has undergone tremendous change during the past two decades. The practice of FMHA, as an integral part of forensic psychology, has become more legally relevant, better supported by empirical research, and more theoretically sophisticated. There are three developments that are likely to continue this growth and improvement during the next decade. The first involves the development of more formal practice standards and updating the *Forensic Specialty Guidelines* published in 1991 to reflect the changes in the field since that time. The second involves additional research, particularly in the functional legal capacities of legal questions that have not been researched in a way that promotes specialized tool development. With the development of an increased number of empirically supported, legally relevant specialized tools, the overall quality of FMHA will certainly be enhanced. Finally, it will be important to continue to expand formal training in forensic psychology. In part, this will involve the accreditation of postdoctoral fellowships specializing in forensic psychology and the development of additional fellowships. As more psychologists receive forensic specialty training, however, it will become feasible to offer less specialized but still relevant forensic training at the predoctoral and internship levels to psychologists who would provide FMHA less frequently or in a more targeted fashion but who still wish to provide courts and attorneys with a product that promotes better informed legal decision making in areas served by the field of forensic psychology.

## 6. CONCLUSIONS

How might the questions raised in the Washington, DC, sniper case be addressed through FMHA? One of the defendants, John Muhammad, briefly attempted to act as his own attorney. Was he competent to do so? Although this question is evaluated less frequently than a defendant's competence to stand trial, it is one that could be addressed through FMHA that described his mental state

and his abilities under relevant law to represent himself. Since Mr. Muhammad was tried and convicted of capital murder, the jury had to consider whether he should receive a sentence of life in prison or death. FMHA describing Mr. Muhammad in terms of the aggravating and mitigating factors under the law of Virginia could have helped the jury consider this question carefully. Lee Boyd Malvo, Mr. Muhammad's accomplice, was tried separately and also charged with capital murder. Malvo was 17 years old at the time of the offense, so the defense might have argued that he should be tried in juvenile court. (Many states do not permit adolescent defendants charged with murder to be treated as juveniles, but this would have been a possibility if the charge had been less serious.) The defense next argued that Mr. Malvo was insane at the time of the alleged offense. Both defense and prosecution introduced evidence based on FMHA conducted by various experts describing Malvo's mental state at the time of the offense. Finally, FMHA on capital sentencing, describing the existing aggravating and mitigating factors for this defendant, could (and did) help inform the jury considering the death sentence. The Washington sniper case illustrates the different ways that FMHA could be used to assist the court in making various kinds of decisions. The continued growth and expansion of the field of forensic psychology promises to bring more relevant and reliable information into the courtroom to continue to assist legal decision makers.

### See Also the Following Articles

Legal Competency ■ Psychology and the Law, Overview

### Further Reading

- Brodsky, S. (1991). *Testifying in court: Guidelines and maxims for the expert witness*. Washington, DC: American Psychological Association.
- Brodsky, S. (1999). *The expert expert witness: More maxims and guidelines for testifying in court*. Washington, DC: American Psychological Association.
- Committee on Ethical Guidelines for Forensic Psychologists. (1991). Specialty guidelines for forensic psychologists. *Law and Human Behavior*, 15, 655–665.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 113 S.Ct. 2786 (1993).
- Frye v. United States*, 293 F. 1013 (D.C. Cir 1923).
- Goldstein, A. (Ed.). (2003). Forensic psychology. In I. Weiner (Editor-in-Chief), *Comprehensive handbook of psychology* (Vol. 11). New York: Wiley.
- Grisso, T. (1998). *Forensic evaluation of juveniles*. Sarasota, FL: Professional Resource Press.

- Grisso, T. (2002). *Evaluating competencies* (2nd ed.). New York: Kluwer/Plenum.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer/Plenum.
- Heilbrun, K., Marczyk, G., & DeMatteo, D. (2002). *Forensic mental health assessment: A casebook*. New York: Oxford University Press.
- Hess, A., & Weiner, I. (Eds.). (1999). *The handbook of forensic psychology* (2nd ed.). New York: Wiley.
- Kumho Tire Company, Ltd. v. Carmichael*, 526 U.S. 137 (1999).
- Melton, G., Petrila, J., Poythress, N., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and attorneys* (2nd ed.) New York: Guilford.
- Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., Roth, L., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York: Oxford University Press.
- Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford.



# Full-Service Schools

Adena B. Meyers and Mark E. Swerdlik

*Illinois State University, Normal, Illinois, USA*

1. Introduction
  2. Rationale for Full-Service Schools
  3. Generalist Approach to Service Delivery
  4. The Obstacle of Categorical Funding
  5. Variations in Organizational Structure and Function
  6. Psychologists in Full-Service Schools
  7. Implications for Continuing Professional Development of Psychologists
  8. What Is the Future of Full-Service Schools?
- Further Reading

In addition to the traditional academic instruction typically provided in public schools, full-service schools offer an array of integrated health and social services to children and their families. This model of service delivery is based on the assumption that children will be successful learners if their social, emotional, and physical needs are met, and that many of these needs can be addressed effectively, conveniently, and efficiently through school-based service delivery.

## GLOSSARY

**categorical funding** A system of funding that allocates resources to separate programs with specialized professionals who address narrow domains of functioning.

**developmental psychopathology** An area of study focused on understanding how emotional, behavioral, and learning problems develop over time. The influence of interacting risk and protective factors is emphasized.

**ecological theory** A theoretical perspective that views human behavior and human development in the context of interconnected social systems.

**intervention** Services delivered to ameliorate identified problems in learning, behavior, health, or mental health. Examples include medication, psychotherapy, and behavior modification.

**prevention** Services delivered to promote healthy functioning thereby reducing the chances that problems will develop.

## 1. INTRODUCTION

Although schools exist primarily to teach children academic skills, links between educational systems and other services have existed for more than a century in the United States (Fig. 1). Public health, in particular, enjoys a natural partnership with public education. Requiring all children attending school to be immunized, for example, benefits schools by protecting students and staff from many communicable diseases and advances the public health goal of universal immunization. As Joy Dryfoos noted in her 1994 book, *Full Service Schools: A Revolution in Health and Social Services for Children and Adolescents*, this collaboration dates back to 1870 when the New York City Board of Health (founded 4 years earlier in response to a cholera outbreak) began working with the New York City Board of Education to vaccinate school children. At the turn of the century, several cities initiated school nursing

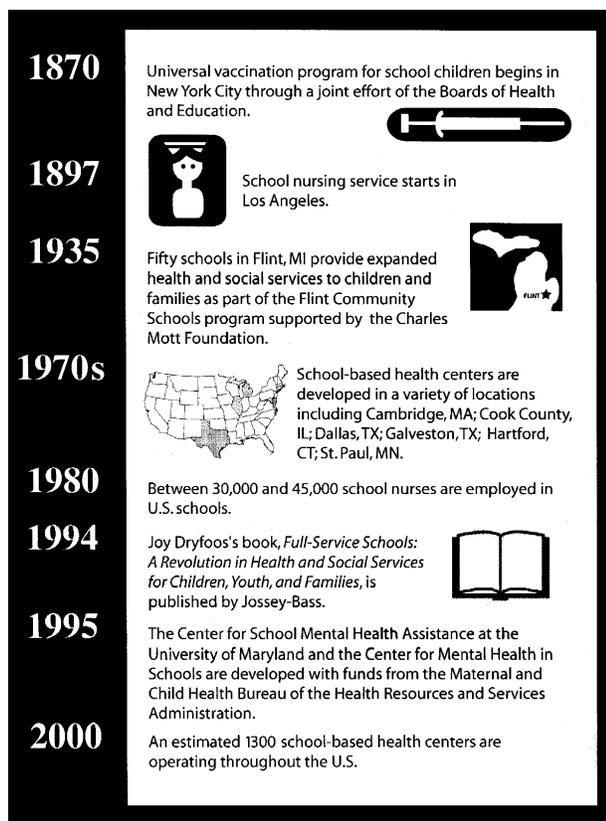


FIGURE 1 History of expanded school-based services.

programs allowing children to be treated for minor health concerns at school rather than being sent home.

The full-service school model represents an expanded version of such cross-disciplinary collaboration in the provision of services designed to enhance children's social, emotional, physical, and intellectual development. Although the idea has yet to take hold in a permanent, universal way, a number of relevant model programs and social experiments have appeared over the years, with fluctuating support from such sectors as state and federal government, private foundations, and professional organizations such as the American Medical Association.

There is considerable variation in what these programs offer and how they are structured. The label "full-service school" has been applied to schools with on-site health clinics, to schools with expanded hours of operation and social services aimed at violence prevention, and to schools that offer mental health services to students and families. Even schools that do not offer these services on-site may be labeled full-service schools if they have well-developed links (e.g.,

referral agreements) with relevant agencies in their communities. In this article, the underlying rationale and approach to service delivery that tie these diverse programs together are articulated. Variations in the organizational structure and function of full-service schools, potential roles for psychologists within these service delivery systems, and implications for in-service and preservice training are also discussed.

## 2. RATIONALE FOR FULL-SERVICE SCHOOLS

### 2.1. Theoretical Background

Rather than focusing narrowly on academic goals, full-service schools are designed to address children's needs holistically. This approach is consistent with the theoretical perspectives of developmental psychopathology and ecological psychology. Developmental outcomes in the physical, social-emotional, and cognitive domains are assumed to be influenced by a variety of interconnected risk and resilience factors. These include characteristics of the individual child (e.g., temperament and cognitive abilities), the contexts in which the child develops (e.g., family, school, and neighborhood), and transactions within and between these systems. For example, a child's temperament is assumed to influence and be influenced by the skills and parenting style of his or her primary caregivers, which in turn may be shaped by a host of environmental factors, including cultural norms, neighborhood characteristics, and workplace experiences.

### 2.2. The Developing Child in Context

At the level of the individual child, development may be influenced by biological factors such as prenatal exposure to alcohol, illegal drugs, and tobacco; birth weight; and genetics. When children grow older, their development is also influenced by their own behavior as they engage in (or avoid) such activities as substance abuse, gang involvement, unsafe sex, dangerous driving, and other health-related risky activities.

Children develop in the context of proximal environments or microsystems, including families, schools, neighborhoods, and peer groups, each of which has the potential to influence outcomes. In the family context, the quality and quantity of children's interactions with their primary caregivers often provide a basis for developmental trajectories leading in desirable or undesirable

directions. For many children, these interactions are limited as a result of increasing divorce rates, large numbers of children growing up in single-parent homes, and increasing numbers of two-parent families in which both parents are employed outside of the home. In some cases, family functioning is further compromised by the presence of parents with mental health or substance abuse problems, illiteracy, family violence, or economic strain associated with unemployment and poverty.

At school, children's social and academic experiences may include overcrowded classrooms, bullying, and peer rejection as well as a range of development-enhancing opportunities. In their neighborhoods, many children must contend with risk factors such as poverty, drug trafficking, exposure to traumatic events, and gang activity. Other systemic influences include mass media, which exposes children to images of violence and risky behavior but may also deliver health-promotion messages; government policies and laws, which often dictate how resources are allocated to needy children and families; and many ideologies and values that underlie much of U.S. society. These macrosystemic influences include institutionalized racism and sexism.

Development is influenced not only by the characteristics of these systems but also by the transactions within and between systems. For example, it has been established that risky behavior among adolescents (including violence and unsafe sex) is most likely to occur during late afternoon and early evening hours. To understand this finding, it is necessary not only to consider the role of the family, the school, the neighborhood, or the peer group but also to recognize the ways in which these systems (along with other, more distal systems) work together to increase (or reduce) children's risk levels.

Related to this is the idea that sets of common risk factors (such as those that produce daily blocks of time during which most adolescents within a community have little to do and little adult supervision) often lead to a variety of divergent emotional and behavior problems. It is becoming increasingly clear, for example, that outcomes as conceptually distinct as adolescent pregnancy, school drop-out, violent behavior, and substance use may represent different end points of very similar causal pathways.

### **2.3. Child Outcomes of Concern**

The consequences of the many risk factors described previously are seen across a variety of domains of children's functioning. Specifically, increasing rates of mental health problems, substance use and abuse, and

dangerous behavior have been observed among young people in recent years. An estimated 20% of youth between the ages of 9 and 17 have diagnosable mental or addictive disorders, and approximately 4 million children and adolescents suffer from major mental illnesses associated with significant impairment in functioning. Additionally, an increasing number of deaths among young people are attributed to risky or dangerous behavior, including homicide, suicide, HIV-AIDS, and automobile accidents.

### **2.4. Full-Service Schools: Comprehensive, Integrated, and Preventative**

Full-service schools address the risk factors that contribute to these worrisome outcomes by providing comprehensive and integrated educational, mental health, medical, and social services. Prevention programs also represent an important response to the wide variety of biological and environmental risks and stressors affecting school-age children and adolescents. Designed to promote positive quality of life and adaptive functioning across settings, and to reduce the incidence of psychopathology, prevention programs fit naturally into the full-service school model.

One important advantage of full-service schools is the opportunity for families to access coordinated, multidisciplinary services through "one-stop shopping." A second advantage, from both an intervention and prevention perspective, is that full-service schools provide access to most (if not all) children and their families in a key setting. Typically, parents may be more comfortable accessing a variety of mental health, medical, and social services in the more familiar setting of a school as opposed to a local community mental health center.

## **3. GENERALIST APPROACH TO SERVICE DELIVERY**

The full-service school model requires a high level of collaboration among diverse professionals, such as clinical, school, and community psychologists; physicians; nurse practitioners; social workers; family educators; and educators. This integrated, multidisciplinary model represents a generalist approach to service delivery, in which children's functioning and needs are viewed holistically. Whereas traditional service delivery systems divide problems into separate,

specialized domains, each assigned to a different profession (e.g., with medical professionals concerned with physical health, educators concerned with learning, mental health professionals responsible for emotional and behavioral functioning, and social workers responsible for family environment factors), the generalist approach emphasizes the interconnections among these domains.

For example, under a more traditional or specialist model, a child who is having trouble learning to read might be referred to a school psychologist for assessment and a reading specialist or special educator for intensive reading services. The student may have received vision screening through a pediatrician and eyeglasses through a social service agency. Contact between these various agencies and professionals would be limited. The generalist model, in contrast, would simultaneously consider health, mental health, family, and community factors that might influence the child's functioning. Such issues as vision, nutrition, family members' literacy (and English language proficiency), and family resources would be considered as possible explanations for the child's difficulty and would be integrated into any intervention plan.

#### 4. THE OBSTACLE OF CATEGORICAL FUNDING

Traditionally, one of the major barriers to the generalist approach to service delivery is the categorical nature of

funding sources. Resources controlled by separate agencies are allocated to separate programs with specialized professionals who address narrow domains of functioning. Proponents of full-service schools have argued that this system results in fragmented, inefficient services that may duplicate efforts, on the one hand, and miss opportunities to influence meaningfully the common factors that underlie many of the devastating outcomes that children experience across domains of functioning, on the other hand. Some foundations as well as federal, state, and local government agencies have responded to this problem by relaxing restrictions on how funds may be used.

#### 5. VARIATIONS IN ORGANIZATIONAL STRUCTURE AND FUNCTION

Full-services schools, also referred to as school-based or school-linked health centers, differ in their degree of comprehensiveness and integration of services, whether services are primarily preventative or treatment oriented, and whether the services are delivered in the school or in locations within the community (Fig. 2). On the surface, it appears that school-based services enjoy certain advantages over school-linked services. These advantages include convenience for the children and their families as well as enhanced opportunities for collaboration between health and

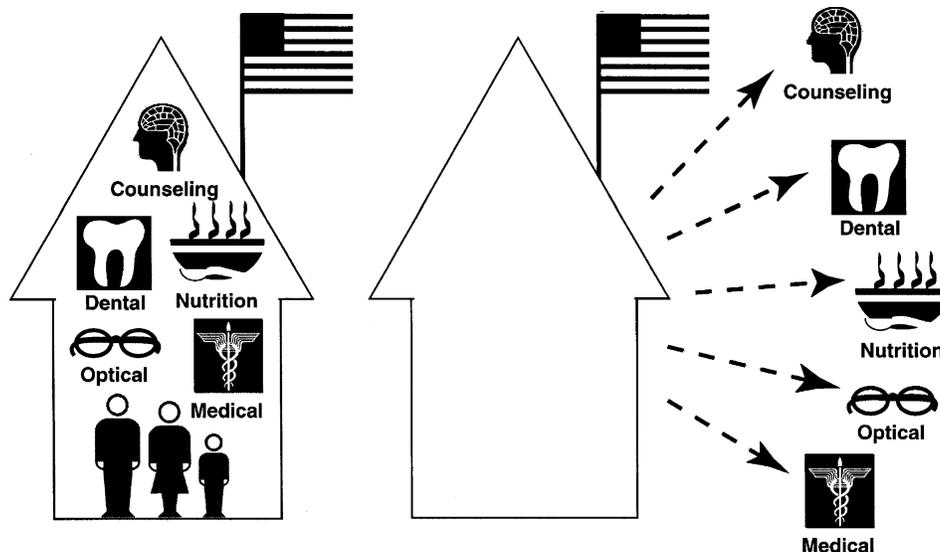


FIGURE 2 School-based vs school-linked models.

social service providers and school personnel, such as teachers, counselors, and administrators.

Despite the intuitive appeal of the distinction between the school-based service models and school-linked models, there is evidence that the location of services may be less important than originally assumed. A nationwide survey conducted by Steven Shaw and colleagues in 1996 found no significant differences between school-based and school-linked services in terms of types of programs offered (acute and preventative medical services, and mental health counseling focused on drugs, alcohol, family, and violence). Neither model provided many social services, such as employment or financial counseling or parole/probation counseling, and both were more likely to offer medical than mental health services. With respect to cross-disciplinary collaboration, it was found that school-based clinics were no more likely to be integrated with education than were school-linked clinics. Thus, despite a large body of literature stressing the importance of a clinic's location as it relates to integration of services, many of the school-based clinics operated autonomously from core educational services.

## **6. PSYCHOLOGISTS IN FULL-SERVICE SCHOOLS**

Psychologists join other professionals in the staffing of full-service schools. Roles for psychologists include providing a range of mental health services, such as screening for conditions such as depression and eating disorders; intervention, including individual, group, and family therapy/counseling; and prevention activities particularly related to substance abuse, pregnancy and sexually transmitted diseases, other high-risk health behaviors, and violence.

Psychologists in full-service schools may also serve as consultants to other professionals, such as teachers, administrators, and medical personnel, with regard to mental health and systemic issues. Psychologists who are skilled in collaborative problem solving and teaming may serve as effective team leaders.

Related to the role of collaborative consultant, psychologists may also function as full-service school administrators and grant writers. Many different funding sources exist for full-service schools, including state and federal mental health, public health, compensatory, and special education funds and also third-party funding

such as private insurance and Medicaid. Funding from private sources, including the Robert Wood Johnson and Hogg Foundations, and public and private hospitals also supports the development of full-service schools.

Researcher/program evaluator represents another potential role for the psychologist. As is true in any new initiative, it is important to document outcomes for students and their families. However, this is particularly challenging when prevention is one of the services offered. Psychologists with expertise in applied research and program evaluation could help fulfill this function.

## **7. IMPLICATIONS FOR CONTINUING PROFESSIONAL DEVELOPMENT OF PSYCHOLOGISTS**

Efforts to expand and reconceptualize current training are called for to help prepare psychologists to offer the broad range of services necessary to perform many of the previously mentioned functions in full-service schools. These changes could occur at the preservice level, including practicum and internship experiences, as well as at the in-service level, including opportunities for continuing professional development. Relevant foci for training include the rapidly growing research base on behavioral health and behavioral management of medical problems. In addition, prevention of academic failure, mental health problems, and health risk behaviors such as smoking and substance abuse as well as promoting healthy development through the use of health education, health services, and guidance are important areas of emphasis.

Because learning is one of the primary missions of the school, psychologists would benefit from knowing how factors such as health, nutrition, substance abuse, and conflict with family or peers influence the learning process. Empirically supported interventions to address these factors at the individual, community, and societal level should also be included in preservice and continuing professional development programs for psychologists who wish to work in full-service schools. Also important are such areas as psychopharmacology (e.g., the effects of various medications on children's learning and behavior), family systems, health promotion, and individual and group therapy. Knowledge of interventions that target multiple microsystems of the child's environment,

including the classroom, the family, peers, the school, and the community as a whole, may be required.

Development of full-service schools will involve significant system reform. In order to succeed, the mission of the full-service school must be integrated into the mission of the school and communicated to decision makers. Thus, information about school reform efforts and initiatives should also be provided as part of the preservice and continuing professional development of psychologists who want to work in full-service schools. Understanding how expanded services fit into the typical organization and structure of the school and familiarity with system change strategies are particularly important for the psychologist. Background in community psychology would also be helpful, as would training in program administration and supervision for the role of program administrator.

Training in group dynamics, collaborative problem solving/consultation, and teaming represents a necessary focus for those who wish to work in the collaborative environment of a full-service school. Areas of concentration include systems-level consultation, coordination of service delivery, obtaining financial reimbursement for services, and overcoming resistance.

Training in the use of technology in providing effective health services to students and their families would be beneficial as well. Technology can be used for data management to enhance and support intervention evaluation activities and to empower students and their families to develop self-help skills and increase self-sufficiency. Technology, including the Internet, can be used to access a wealth of information and resources.

## 8. WHAT IS THE FUTURE OF FULL-SERVICE SCHOOLS?

In the future, schools will likely be challenged to meet the increasingly complex academic, physical, and mental health needs of children and their families. One way to accomplish this may be to expand the network of full-service schools from the current level of approximately 1300 nationally. In order for this movement to reach its objectives, reform efforts are needed to restructure schools beyond instructional and management systems to providing comprehensive, integrated services addressing all barriers to learning and promoting physical and mental health. This includes continuing efforts to make effective coordinated health care available to poor children and children of color.

To achieve the goals of the full-service school, more effective methods of coordinating and integrating existing school and community resources must be developed. Disciplines involved with full-service schools will need to address barriers to more effective coordinated efforts. One way to do this is to provide interdisciplinary training in evidence-based prevention and interventions at the graduate, postgraduate, and continuing education levels to allow health, mental health, and education personnel to work collaboratively and effectively in full-service schools. In this context, psychologists may be able to assume leadership roles in outcomes-based program evaluation activities so as to apply the science of psychology to better inform practice, resulting in implementation of maximally effective full-service schools.

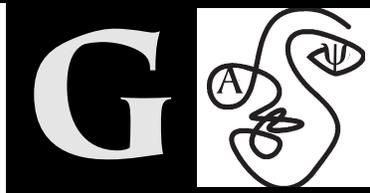
### See Also the Following Articles

Behavioral Observation in Schools ■ Learning ■ Learning Styles and Approaches to Studying ■ Neuropsychological Assessment in Schools ■ School–Community Partnerships

### Further Reading

- Adelman, H. S., & Taylor, L. (1994). *On understanding intervention in psychology and education*. Westport, CT: Praeger.
- Adelman, H. S., & Taylor, L. (1998). Reframing mental health in schools and expanding school reform. *Educational Psychologist*, 33, 135–152.
- Adelman, H. S., & Taylor, L. (2000). Shaping the future of mental health in the schools. *Psychology in the Schools*, 37, 49–60.
- Bronfenbrenner, U. (1977). Towards an experimental ecology of human development. *American Psychologist*, 32, 513–531.
- Brown, M. B., & Bolen, L. M. (1993). School-based health centers. [Special issue]. *School Psychology Review*, 40(2).
- Bucy, J. E., Meyers, A. B., & Swerdlik, M. E. (2002). Best practices in working in full service schools. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 281–291). Bethesda, MD: National Association of School Psychologists.
- Carlson, C. D., Tharinger, D., Bricklin, P., DeMers, S., & Paavola, J. (1996). Health care reform and psychological practice in the schools. *Professional Psychology: Research and Practice*, 27, 14–23.
- Center for School Mental Health Assistance. <http://csmha.umaryland.edu>.
- Dryfoos, J. (1994). *Full service schools: A revolution in health and social services for children, youth, and families*. San Francisco: Jossey-Bass.

- Paavola, J. C., Cobb, C., Illback, R. J., Joseph, H. M., Jr., Torreulla, A., & Talley, R. C. (1995). *Comprehensive and coordinated psychological services for children: A call for service integration*. Washington, DC: American Psychological Association.
- Reeder, G. D., Maccow, G. C., Shaw, S. R., Swerdlik, M. E., Horton, C. B., & Foster, P. (1997). School psychologists and full-service schools: Partnerships with medical, mental health, and social services. *School Psychology Review*, 26, 603–621.
- Shaw, S. R., Reeder, G., Swerdlik, M. E., Maccow, G. C., Horton, C., & Bucy, J. (1996, August). *A national survey of full-service schools: Services provided, professionals represented, and the role of the school psychologists*. Paper presented at the annual meeting of the American Psychological Association, Toronto, Canada.
- Talley, R. C., Short, R. L., & Kolbe, L. (1995). *School health: Psychology's role: A report to the nation*. Washington, DC: American Psychological Association Center for Psychology in the Schools and Education/U.S. Public Health Service, Centers for Disease Control and Prevention.
- UCLA Center for Mental Health in the Schools. <http://smhp.psych.ucla.edu>.
- Weist, M. D. (1999). Improving children's mental health through expanded school-based programs [Special issue]. *Clinical Psychology Review*, 19(2).
- Weist, M. D., Myers, P. C., Danforth, J., McNeil, D. W., Ollendick, T. H., & Hawkins, R. (2000). Expanded school mental health services: Assessing needs related to school level and geography. *Community Mental Health Journal*, 36, 259–273.



# Gender and Culture

Deborah L. Best and Dustin J. Foster

Wake Forest University, Winston-Salem, North Carolina, USA

1. Gender Development across Cultures
  2. Behavioral Differences among Boys and Girls
  3. Development of Gender Roles and Stereotypes
  4. Theories of Gender-Related Learning
  5. Gender at the Adult Level
  6. Relations between Women and Men
  7. Gender Differences at the Cultural Level
  8. Conclusion
- Further Reading

*sex* Term generally used to describe the anatomical and physiological differences between males and females, but it may also be related to social and behavioral differences.

*sex role ideology* Beliefs about the appropriate roles and relationships of men and women. They are often described as varying from traditional to egalitarian.

## GLOSSARY

*culture* A dynamic system of customary rules and expectations shared by a group of people and incorporating the beliefs, attitudes, norms, social practices, and organizations of that group.

*gender* Term used to describe the differences between males and females, often by emphasizing social and behavioral factors.

*gender roles* The social roles (e.g., familial, interpersonal, occupational, and recreational) in which males and females participate with differential frequency.

*gender stereotypes* The psychological traits that are believed to occur with differential frequency between males and females (e.g., men are more “aggressive” and women are more “emotional”).

*masculinity/femininity* The degree to which females and males perceive that they or others possess traits or behaviors that are considered “woman-like” and “manlike” in their culture.

Gender is a term used to describe a set of psychological, behavioral, and physical attributes that characterize members of either sex. Among any group of people, the psychological and behavioral components of gender often lead to various degrees of maleness and femaleness with considerable overlap between the two. Because gender is defined by the beliefs groups of people have regarding what it means to be male or female, gender is a subjective concept and an appealing topic for psychologists who study multiple cultures. Anyone who has visited different countries has surely noticed differences in the way men and women dress, the kinds of work they do, and the way they interact with one another. Such behavioral differences and their corresponding expectations regarding what is proper for men and women within a particular culture are referred to as gender roles. This article discusses cross-cultural research on gender roles; the degree of cross-gender similarity within cultures in comparison to the substantial variability often observed across cultural groups may be surprising. This article also examines gender theories from developmental,

social, and personality perspectives, with particular attention to various means of studying and measuring gender across the life span.

## **1. GENDER DEVELOPMENT ACROSS CULTURES**

### **1.1. Distal Considerations**

Developmental influences can be grouped into two somewhat overlapping categories: distal and proximal factors. Distal factors are those experienced by many or all within a culture. These include socioeconomic (e.g., national economy and type/availability of work), political, and religious factors as well as prevalent societal expectations. Discrepancies exist within cultures to the extent that job opportunities, political roles, governmental regulations, and social norms for proper behavior differ for men and women. By growing up with continuous exposure to these overarching factors, children learn the gender role expectations of their culture.

#### **1.1.1. Socioeconomic Influences**

National standards of living tend to be related to parents' reasons for having children. The nine-nation Value of Children study found that parents in agrarian cultures (e.g., Indonesia) frequently endorsed utilitarian reasons for having children. These values typically indicated a desire for fulfillment of fundamental needs, such as financial support in old age. In contrast, parents in industrialized nations (e.g., Germany) reported reasons for having children that related to fulfillment of emotional needs, such as affection. Parents in agrarian societies tended to bear more children, perhaps in response to environmental conditions. Particularly relevant to gender development, parents in agrarian societies favored boys over girls, but the two were valued equally by parents in industrialized nations. Similarly, males in societies marked by labor-intensive economies tend to be more successful than females at securing jobs and generating income for their families. Gender role expectations for women may limit their ability to contribute financially to the family, often preventing them from working in jobs outside of the home. Additionally, a young woman may marry into another family and cease contributing to her biological family.

#### **1.1.2. Cultural Environments**

Cultures also differ with regard to the religious practices endorsed by the people and/or the political/governmental organization. Religious doctrines vary greatly in the status and freedoms they grant to men and women. The three most predominant religions throughout the world are Christianity, Islam, and Hinduism. In theory, Christian teachings endorse considerable gender equality. However, the Christian God is generally personified as male, and the principal religious servants, ministers, and priests are almost exclusively men. The Hindu pantheon has goddesses as well as gods, and both men and women serve in the temples. In contrast, in Muslim theology, all the significant figures are male and religious practice is controlled by men, as is Muslim society. In nations in which religion and government are mostly separate, religion may not exert a universal influence on the developing child. However, when there is a union between religious doctrine and governmental rule, religious beliefs may strongly influence the relations between men and women as well as child-rearing practices.

Language is one of the most noticeable ways that cultures differ, and indeed this fundamental difference can lead to divergence in gender roles. Languages differ in terms of the frequency and connotation of gender-related words. Recent estimates indicate that approximately 7000 different language systems exist throughout the world. Obviously, the potential for cross-cultural differences in the way children mentally represent and verbally express gender-laden information is substantial. For example, in English it is grammatically correct to use the pronoun "he" to mean he or she, but research has shown that psychologically "he" is not a sex-neutral or generic term and may create a male bias in communication. In inflected languages, such as French or German, which have gender-specific endings for words, it is virtually impossible to communicate without introducing gender. Although such linguistic messages may seem trivial, their prevalence in everyday speech make them a subtle but pervasive influence on sex differentiation practices.

The aforementioned distal environmental influences can interact with biological factors to produce adaptive but distinctive cultural differences. When cross-cultural gender differences exist, they may in part be the result of contrasting environmental triggers that elicit different behaviors. Consistent with this perspective, David Gilmore proposed that the male macho behavior pattern could be an adaptation to divergent economic conditions. This speculation is supported by an

examination of the gender roles of two South Pacific islands, Truk and Tahiti. In the Trukese culture, men are violent, competitive, and sexually promiscuous, whereas women are expected to be submissive. Tahitian culture differs dramatically in that the men engage in submissive behaviors and avoid violent or competitive encounters. In contrast, Tahitian women tend to be sexually active. Gilmore explained these gender role discrepancies by a consideration of each culture's means of obtaining food. Tahitians fish in a protected lagoon where fish are plentiful. However, the Trukese must fish in the open ocean with continuous exposure to life-threatening situations. From this account it appears that the Trukese male gender role is an adaptive response to increase one's chances of survival. This evolutionary theory of gender role development is generally consistent with widely held views of the interaction of nature and nurture. However, many criticisms of the theory exist and research has often failed to support some of its assumptions.

Considering these distal influences, it is undeniable that culture can have profound effects on individual development. Sex differentiation messages are clear and cultural practices that emanate from these determine how babies are delivered, how children are socialized, how women and men dress, what tasks they pursue, what languages they speak, and what roles they adopt. Distal cultural factors are also related to the more immediate, proximal factors in development.

## **1.2. Proximal Considerations**

Proximal developmental factors are more localized than distal ones. They are essentially developmental influences that are relevant to a particular individual or family. In a sense, they are the case-by-case manifestation of the distal factors in one's culture. Although many distal features of a culture are objectively consistent for all members in the culture, subjective experiences of those distal factors can vary greatly. In a Third World country, everyone is not poor, nor are all of the poor people poor to the same degree. Everyone in a particular culture is exposed to and influenced by a faltering national economy, but the impact of the economy on individual lives varies widely. For example, in the Value of Children study, Turkish parents with low income and education valued their children for their ability to contribute financially to the family. However, as income and education increased, Turkish parents valued their children less for old-age security reasons and more for emotional ties. Although the distal economic influence

was objectively consistent for all Turkish participants, their subjective experiences and subsequent responses varied greatly.

### **1.2.1. Biological Factors**

Biological factors are perhaps the most significant contributor to the development of observable gender differences. Although the term biological is often used to refer to genes, in this case sex chromosomes, it also includes an organism's prenatal and postnatal environment. Although these developmental components intuitively appear to be fixed, considerable cultural variation actually exists. Differences in general nutrition and prenatal care can have profound effects on the prenatal experience. In the postnatal environment, biology and environment work together to yield cultural differences in behaviors, such as young children's sleep patterns, and attainment of developmental milestones, such as sitting and walking.

Observable gender differences are influenced by biological factors essentially from birth. Males tend to be more biologically vulnerable from the moment of conception (e.g., male:female ratio at conception of 140:100 versus 105:100 live births) and throughout the life span. However, males tend to be larger and have higher activity levels, higher basal metabolism, more muscle development, and higher pain thresholds than females at the time of birth. Physical sex differences in body structure and hormonal state are minimal throughout the preadolescent years, but the differences that exist are consistent with subsequent development. By adulthood, males tend to be taller, have a more massive skeleton, a higher muscle-to-fat ratio, higher blood oxygen capacity, increased body hair, and different primary and secondary sex characteristics. In part, these differences result from a longer growth period for boys and hormonal changes that occur after age 8. These changes generally relate to greater physical strength and stamina of males relative to females, although males tend to die an average of 6 or 7 years earlier than females. Note that these differences appear within cultures rather than across all cultures and at group rather than individual levels.

### **1.2.2. Socialization of Girls and Boys**

Many attribute the emergence of sex-specific behaviors for boys and girls to differences in socialization. Surprisingly, many of these differences appear to be consistent across numerous cultures. Herbert Barry,

Margaret Bacon, and Irvin Child studied the socialization practices of more than 100 preindustrial societies and found that boys were generally raised with an emphasis on achievement, self-reliance, and independence; conversely, girls were typically socialized to be nurturant, responsible, and obedient. Note that these differences emerged between boys and girls within the same cultural group and not necessarily between cultures. Girls in some societies are raised to be more independent and self-reliant than boys in other cultures.

Studies in North American and Western countries have found that parents more consistently reinforce sex-typed behaviors and use higher levels of physical punishment with boys than with girls. However, cultures differ widely in how frequently physical punishments are used, and the differential treatment of males and females appears to dissipate with age. The infinite human variation in the biological and experiential influences that comprise each person's developmental history allows for many different interpretations of the same experience.

**1.2.2.1. Cultural Differences in Role Socialization** The impact of cultural differences in socialization was seen in the classic six-culture study, an intensive investigation of socialization practices in India, Kenya, Mexico, Okinawa, the Philippines, and the United States. Of central interest was the occurrence of aggressive, nurturant, responsible, helpful, and attention-seeking behaviors among 3- to 11-year-old children. Less gender differentiation was found in cultures in which both boys and girls helped with sibling care and household chores (i.e., Kenya, the Philippines, and the United States). In contrast, more gender discrepancies emerged in cultures in which females participated more than males in sibling care and household chores (i.e., India, Mexico, and Okinawa). Overall, girls spent more of their time in responsible, productive work, and boys spent relatively more time playing.

Further evidence for discrepant patterns of caretaking roles of males and females comes from an examination of the care-taking practices of 186 nonindustrialized societies. Thomas Weisner and Ronald Gallimore found that mothers, female adult relatives, and female children were the primary caretakers of infants. However, as infants reached early childhood, responsibility for child care fell upon siblings of both sex groups. Because young children in some societies spend as much as 70% of each day in the care of their siblings, sibling caretakers constitute an integral part of

the gender role socialization process. Although mothers in such societies spend considerable proportions of their time in nonmothering, family provider roles, children in virtually every culture view mothers as responsible for children.

Fathers often fulfill the role of primary financial provider for the family, and this may limit their contributions to caretaking. Father-child interactions in many cultures are structured around play-oriented activities. Although the father's role may not appear to be essential, research has shown that fathers play an important role in the socialization of children. When fathers are absent for lengthy periods of time, boys may display effeminate behaviors, high levels of dependence, excessive fantasy aggression, and overly masculine behaviors. However, the criteria for judging behaviors as "effeminate" may vary across cultural groups. Indeed, fathers tend to exhibit higher levels of gender role socialization than mothers and they pay less attention to daughters than to sons. These findings suggest that fathers' influence is more critical for sons than for daughters.

In addition to parents, peers play an important socialization role for children in all cultures, and this role tends to increase with age. Eleanor Maccoby identified three gender-linked phenomena in children's social development with peers: gender segregation, dissimilar interaction styles, and group asymmetry (e.g., boys groups are more cohesive, sexist, exclusionary, and separate from adult culture than girls' groups). The six-culture study found that as early as age 3, children tend to seek out same-sex peers and to avoid other-sex children, and this tendency strengthens throughout grade school. In their segregated groups boys strive for dominance, take risks, and are reluctant to reveal weaknesses to each other. In contrast, girls self-disclose more, try to maintain positive social relationships, and avoid conflict. Compared with girls' groups, boys' groups are more cohesive, more exclusionary, and more separate from adult culture. Segregated groups may lead to different activities and toy choices, which may in turn lead to differences in intellectual and emotional development.

**1.2.2.2. Education** Across cultures, educational settings are common sources of differential gender role socialization. For example, observations of fifth-grade students in Japan and the United States revealed that teachers devoted more than half of their attention to boys, but that much of this attention was negative. Research studies have consistently found teacher expectations for student achievement to play a critical role in

predicting discrepancies in classroom performance. As one would expect, parental expectations also play an important role in academic achievement. Robert Serpell confirmed this notion with his observation in Zambia, where education is intended for boys rather than girls. Harold Stevenson and colleagues found that mothers in China, Japan, and the United States expected their sons to outperform their daughters in mathematics and their daughters' reading abilities to exceed those of their sons; however, boys and girls perform equally well in some aspects of both disciplines.

Interestingly, during the second half of the 20th century women's level of education has increased and even surpassed that of men in the United States and in several other Western countries. This is a profound shift considering that in some countries women are not permitted to participate in formal education.

## 2. BEHAVIORAL DIFFERENCES AMONG BOYS AND GIRLS

The aforementioned developmental considerations include some of the potential factors that contribute to different gender role behaviors. A more in-depth discussion of behavioral domains in which males and females show divergent patterns is presented here.

### 2.1. General Patterns of Activity

Cross-cultural consistency can be found through an examination of the characteristics of male and female group behaviors. Daniel Freeman studied 5- to 7-year-olds in eight different cultures: Australian Aboriginal, Balinese, Ceylonese, Japanese, Kikuyu, Navajo, Punjabi, and Taiwanese. He found that, relative to males, young females remained closer to home and engaged in repetitive, conversation-oriented behaviors. Males, on the other hand, frequently played in larger groups, participated in more physical activities, covered more physical space in play, and performed more unpredictable behaviors. In general, boys tend to spend most of their time with other boys, but girls often choose to interact with adults.

Children's drawings also shed light on their segregated gender preferences and gender role-consistent interests. Boys tend to produce more drawings of other boys, vehicles, monsters, and violent themes, whereas girls' drawings more often depict other girls and flowers.

### 2.2. General Behavioral Styles

Cross-cultural studies have generally found preadolescent males to exhibit higher rates of competitive, aggressive, and dominance-seeking behaviors and females to demonstrate higher levels of nurturance. In particular, the six-culture study found sex differences in aggression and dominance, consistent levels of aggression across many age groups, and increased physicality among the eldest boys. Indeed, fathers may play a particularly important role in the socialization of aggressive behaviors in boys. Sex differences have also emerged from a consideration of the manner in which both boys and girls engage in aggressive interactions. Initially, males display more restraint than females but are more violent when they do act. Female aggression is often marked by emotion-laden, verbal tirades.

With regard to nurturant behaviors, because infants elicit more nurturant behavior than do older children, girls who spent more time with infants displayed more nurturance than boys who did not engage in as much interaction with infants. Herbert Barry and colleagues examined socialization practices in more than 100 societies and found that generally boys are reared to achieve and to be self-reliant and independent, whereas girls are reared to be nurturant, responsible, and obedient.

### 2.3. Self-Esteem

Girls have often appeared to be less satisfied with being girls, and boys generally perceive themselves to be more competent than girls. However, some studies suggest that these differences tend to be domain specific and can even be reversed depending on what aspect of self-esteem is examined. David Watkins found that girls in Australia, Nepal, and the Philippines reported lower self-esteem than boys in the areas of physical abilities and mathematics. Additionally, Nigerian boys felt they were more intelligent overall than girls. However, Australian and Nigerian girls perceived themselves to be more competent than boys in reading. As one can see, male and female differences in self-esteem across cultures are not as robust as general behavioral patterns. Perhaps the inconsistencies in self-esteem across genders are due to cultural differences in the appropriateness of various behaviors. Cultures that consider certain behaviors to be more negative than do other cultures may associate them with greater maladjustment, such as low self-esteem.

### 3. DEVELOPMENT OF GENDER ROLES AND STEREOTYPES

Research has consistently shown that gender-specific knowledge emerges early in a child's development. In the United States, children as young as age 2 begin to stereotype objects as masculine or feminine, and by age 3 or 4 they reliably use stereotypic labels with toys, activities, and occupations. Similar patterns can be found in countries as distinct as West Africa, where girls tend to play with dolls and boys often build vehicles and weapons. Although the contextual details of children's play differ across cultures, most general play patterns are not culture specific.

#### 3.1. Sex Trait Stereotypes

Research in the United States revealed that children's understanding of sex trait stereotypes may emerge more slowly than their stereotypic knowledge of toys and occupations. John Williams and Deborah Best developed the Sex Stereotype Measure (SSM) to assess children's understanding of adult notions of sex stereotypes. In this task, children were shown a silhouette drawing of a man and a woman while they were read stories consisting of various sex stereotype traits (e.g., aggressive, emotional, and gentle). Children were then asked to tell which silhouette each story was about. European American children's knowledge of sex trait stereotypes showed a consistent pattern of growth from kindergarten through high school. Scores showed the most dramatic gains during elementary school years and often appeared to plateau after junior high school. African American children's knowledge of the same stereotype traits also increased with age, but at a slower rate. This observation suggests that different nationalities within one country may vary somewhat in their stereotype knowledge base.

#### 3.2. Findings across Cultures

Williams and Best used the SSM with 5-, 8-, and 11-year-old children from 25 different nations. In general, they found that the frequency of stereotyped responses increased from approximately 60% at age 5 to 70% at age 8. At all ages, items consistently associated with males were strong, aggressive, cruel, coarse, and adventurous; weak, appreciative, softhearted, gentle, and meek were consistently associated with females.

Scores were exceptionally high in Pakistan and relatively high in England and New Zealand, indicating that children in these countries had a considerable understanding of sex stereotypes. Conversely, scores were particularly low in Brazil, France, Germany, and Taiwan, suggesting consistent stereotype trait knowledge may develop more slowly in these countries. The general pattern across all countries was characterized by the onset of stereotype learning prior to age 5, accelerated learning during the school years, and functionally complete stereotype knowledge during adolescence.

Boys and girls acquired stereotype knowledge at the same rates, but there was a tendency for both groups to learn about the male stereotype at a faster rate. However, this tendency was reversed in Latin/Catholic cultures (i.e., Brazil, Chile, Portugal, and Venezuela), in which the adult female stereotype was more favorable than that of the adult male. Male stereotype items were better known than female items by both sexes in 17 of the 24 countries studied. Germany was the only country in which there was a clear pattern of greater knowledge of female stereotype items. Like the Latin/Catholic cultures, German female stereotype knowledge at age 5 was greater than male stereotype knowledge. Differences emerged at 8 years due to the persistence of greater female stereotype knowledge among German children but a change to greater male stereotype knowledge among 8-year-olds from Latin/Catholic cultures.

Five-year-olds in predominantly Muslim countries differentiated more between the sexes than 5-year-olds in non-Muslim countries. Furthermore, children in Muslim countries learned the stereotypes, particularly the male stereotype, more quickly than children living in predominantly Christian countries. Because adult stereotypes tend to be less well differentiated (i.e., overlap more) in countries in which a majority of people practice Christianity, children in these countries may acquire stereotype knowledge at a slower rate than children living in cultures distinguished by rigid boundaries that separate the characteristics generally considered to be typical of males or females.

Studying older children, Margaret Intons-Peterson found that 11- to 18-year-olds in Sweden had more similar male and female stereotypes than children in the United States. However, little overlap was found in the Swedish children's ideal occupational choices. Swedish boys were interested in business careers, whereas Swedish girls reported a preference for occupations in the service industry, such as nanny, hospital worker, and flight attendant.

#### 4. THEORIES OF GENDER-RELATED LEARNING

Although most theories of gender role learning were developed in the United States, they consistently emphasize the gender-relevant knowledge that is available within a particular cultural context. The theories differ on many points, but they each accommodate cross-cultural patterns of gender role development.

##### 4.1. Social Learning

From the perspective of social learning theories, gender role development is conceptualized as the accumulation of experience with socialization agents in one's environment. Socialization agents such as parents, teachers, and peers concurrently shape a child's gender-related behaviors by rewarding gender-appropriate behaviors, punishing gender-inappropriate behaviors, modeling gender-appropriate activities, selecting gender-appropriate toys, and generally treating boys and girls in a differentiated, sex role-consistent manner. Parents in the United States, and fathers in particular, have been found to treat their children differently depending on the sex of the child. For example, boys are often given more freedom and independence than girls, and they tend to receive more physical stimulation. However, studies in other countries have shown less differential treatment on the basis of sex relative to that found in the United States.

Deborah Best and colleagues observed parents and their preschool children in public parks and playgrounds in France, Germany, and Italy, and they found variation in parent-child interaction patterns across both gender and culture. In comparison to French and Italian parent-child pairs, German parents interacted less frequently with their children. Furthermore, German children showed and shared more objects with their mothers, whereas French and Italian children engaged in the same behaviors more frequently with their fathers. These observations suggest that German children, relative to those in other countries, may engage in a disproportionately greater amount of interactions with their mothers relative to their fathers, possibly fostering more rapid development of the female stereotype relative to the male stereotype. Similarly, these data suggest the opposite pattern for French and Italian children. This pattern is consistent with the aforementioned findings from the cross-cultural work of Williams and Best, which

indicated that Germany was the only nation in which the female stereotype was learned more rapidly than the male stereotype and persisted at higher levels into later childhood.

It is clear that social factors play an important role in children's learning of gender roles, and there is wide variation in how girls and boys are treated across cultural groups. Surprisingly, differential treatment of boys and girls is not consistently related to gender differences in the behavioral patterns observed across cultures.

##### 4.2. Cognitive Development

Cognitive developmental theory focuses on the way cognitive structures affect the influence of the environmental factors that are often studied by social learning theorists. Determining the role that cognitive structures play in the development of gender knowledge and gender-typed behaviors is a difficult task due to the changing nature of children's cognitive structures across childhood. As these structures change with age, so too does the manner in which children interpret and remember gender-relevant experiences.

Consistent with Jean Piaget's stage theory of cognitive development, Ronald Slaby and Karin Frey identified four developmental stages that characterized children's gender knowledge acquisition in North America. Prior to age 2, children did not distinguish between men and women, but in the second stage by age 2 or 3 they used physical characteristics to differentiate between the sexes. By the third stage at approximately 4 years of age, children understood that gender was stable, and by age 5 or 6 they knew that gender remained constant.

Ruth and Robert Munroe and Harold Shimmin tested the predictions of cognitive developmental theory in a study of children in societies marked by differing degrees of gender differentiation. They predicted that in societies with high gender differentiation (i.e., Kenya and Nepal), children would progress through the cognitive gender stages at a faster rate than children living in less differentiated settings (i.e., Belize and Samoa). Contrary to these predictions, attainment of gender classification capabilities was not related to cultural variations in gender differentiation. Although the degree of gender differentiation varied between these cultures, the discrepancies within each culture may have been sufficient to promote the development of gender discrimination skills. Furthermore, it is possible that the extensive gender differentiation in Kenya and Nepal may have been superfluous for children's learning.

## 5. GENDER AT THE ADULT LEVEL

It is apparent from the research on the development of gender roles and stereotypes that children's cultural environment influences much of what they learn. Children are socialized by adults whose social behaviors, gender stereotypes, and self-perceptions reflect differences between men and women.

### 5.1. Sex Role Ideology

Sex role ideology concerns an individual's beliefs about men's and women's proper role relationships, ranging along a continuum from traditional to modern. The traditional end of the continuum is characterized by beliefs that men are more "important" than women and that women should be subservient to men. Modern ideologies are more egalitarian and are consistent with the notion that women and men are equally important and that neither sex has the right to dominate the other.

Because of the juxtaposition of traditional and modern ideologies among its inhabitants, India has been a fascinating nation in which to study adult sex roles. Khavita Agarwal, David Lester, and Nisha Dhawan asked university students in India and the United States to identify the qualities men and women should possess in their respective cultures. They found that North American students expressed more modern views than Indian students; however, across both cultures, women were more liberal (modern) than men. Rehana Ghadially conducted a more in-depth analysis of Indian women and found that those with the most modern sex role ideologies grew up in nuclear families, had educated mothers, and worked in professional disciplines.

Atsuko Suzuki conducted similar studies in Japan and the United States and found education and job status to be strong predictors of sex role attitudes among women. North American women with jobs held more egalitarian views than women without jobs. In Japan, women with career-oriented professional jobs reported more egalitarian views than all other women.

Judy Gibbons, Deborah Stiles, and Gina Shkodriani studied attitudes toward family roles and gender among adolescents studying in The Netherlands. The students originated from 46 different countries and formed two groups based on origin: individualistic, wealthier countries and collectivistic, less wealthy countries. Students from this second group of countries reported more traditional attitudes than students

from the first group, and across both groups, boys reported more traditional attitudes than girls.

In studies of sex role ideology, the United States has often served as a comparison group, and the beliefs of U.S. adults are generally found to be more liberal than those of adults of other nations. However, this was not the case in a 14-country study of the attitudes of university students by Williams and Best. Northern European countries (i.e., The Netherlands, Germany, Finland, and England) had the most modern ideologies, the most traditional attitudes were found in African and Asian countries (i.e., Nigeria, Pakistan, India, Japan, and Malaysia), and the United States was in the middle of the distribution. With the exception of Malaysia and Pakistan, women reported more modern attitudes than men. However, within a given country, men's and women's sex role ideology scores were highly similar; that is, greater differences existed between cultural groups (men and women combined) than within cultural groups (men compared to women).

In cross-cultural research it is necessary to show that differences between cultural groups are related to other cultural variables before one can conclude that the differences of interest are due to cultural factors. In the 14-country study noted previously, sex role ideology scores were related to socioeconomic development, supporting the role of cultural factors. Specifically, sex role ideology scores were more modern in more developed nations, in countries in which Christianity was more predominant, in more urbanized countries, and in countries at higher latitudes (i.e., countries at higher latitude tend to have greater socioeconomic development).

### 5.2. Gender Stereotypes

Gender stereotypes are the psychological traits believed to be more characteristic of one sex relative to the other. These stereotypes are closely related to sex role ideologies, and they have often been used to account for variations in sex role ideologies across cultures. Williams and Best presented university students in 27 countries with the Adjective Checklist, a standard personality measure consisting of 300 person-descriptive adjectives. Students indicated whether, in their culture, each adjective was more characteristic of men, more characteristic of women, or not differentially associated with either sex. Students generally agreed about the characteristics differentially associated with men and women, but differences between the male and female

stereotypes did emerge. Male and female stereotypes varied the most in The Netherlands, Finland, Norway, and Germany and least in Bolivia, Scotland, and Venezuela. Additionally, greater stereotype variation was found in Protestant compared to Catholic countries, in more developed countries, and in countries higher in individualism.

Analysis of the qualitative aspects of the male and female stereotypes revealed that the male stereotype items were stronger and more active across cultures. Across all countries, items representing dominance, autonomy, aggression, exhibition, and achievement were associated with men, and nurturance, succorance, deference, and abasement were consistently chosen to describe women.

Recent work examined these stereotype findings in terms of the Five Factor Model of personality, an approach based on the assumption that five dimensions can fully capture the aspects of personality. The pan-cultural male stereotype, which consisted of items consistently associated with men across most cultures, was found to be higher on the dimensions of extraversion, conscientiousness, emotional stability, and openness, whereas the female stereotype was higher on the dimension of agreeableness.

In contrast to these cross-cultural generalities, variation between the countries emerged as well. No consistent pattern emerged for stereotype favorability, with the male stereotype items being more favorable in some countries (e.g., Japan, South America, and Nigeria) and the female stereotype items exceeding the male items in others (e.g., Italy, Peru, and Australia). Greater variation in the perceived strength and activity of the male and female stereotypes was found in countries that were less socioeconomically developed and had lower literacy rates and lower percentages of women attending universities. These findings suggested that economic and educational advancements were accompanied by reductions in the tendency to view men as stronger and more active than women. Note, however, that these effects were only reduced, not eliminated.

### **5.3. Masculinity/Femininity of Self-Concepts**

The paired concepts of masculinity and femininity essentially mean manlike or woman-like, and a person can be characterized as masculine or feminine along a number of dimensions, such as dress, mannerisms, or

tone of voice. One aspect of masculinity and femininity relates to self-concepts and the degree to which individuals incorporate traits that are differentially associated by men or women into their self-concepts. Researchers have used a variety of measurement techniques to assess individual differences in the masculinity and femininity of self-concepts. For example, some researchers have used self-descriptive questionnaires, some have analyzed only socially desirable characteristics, and others have examined gender-associated characteristics without considering social desirability. Certainly, a conjunction of the three would provide a more comprehensive examination.

Valid measurement is a continuing concern for researchers studying the similarities and differences between various cultural groups. Measurement problems can arise when a scale is developed in one culture and then applied in other cultures without attention to the validity of the instrument in the new context. A prime example of questionable measurement validity occurred in a cross-cultural study of masculinity and femininity by Robert Helmreich and Janet Spence. They used the Personal Attributes Questionnaire, which contains traits associated with either males or females in the United States, to compare the self-descriptions of men from Brazil and the United States. They found that Brazilian men endorsed more feminine than masculine traits, whereas men from the United States endorsed the opposite pattern. On the surface, these results suggested that men in the United States were more masculine than men in Brazil, and, surprisingly, that Brazilian men may have been more feminine than masculine. However, these accounts do not consider cultural differences in the definitions of masculinity and femininity. It is equally plausible that, from the Brazilian perspective, the self-descriptions of men in the United States are more feminine than those of Brazilian men.

The aforementioned difficulties were addressed in a study by Williams and Best in which the masculinity and femininity of items were determined within each culture. These authors assessed the masculinity and femininity of the self-concepts of university students in 14 countries. Participants selected the items from the Adjective Checklist that they believed described their actual and ideal self-concepts (i.e., who they really were vs who they would like to be). These descriptions were scored according to university student ratings of the masculinity and femininity of each item in prior studies in each of the 14 countries. Therefore, the stereotyped items reflected culture-specific conceptualizations of

masculinity and femininity. As one would expect, the self-concepts of men were more masculine than those of women in each culture. Surprisingly, members of both sexes rated their ideal selves as more masculine than their actual selves. Considering the preceding discussion of the affective features of male and female stereotypes, this trend of desired masculinity may reflect a universal desire for a stronger, more active self.

Variation did exist in the self-concepts of men and women across the cultures studied, but this variation was not related to meaningful cultural comparison variables, such as socioeconomic development. When self-concepts were scored in terms of the relevant culture's standards for masculinity/femininity, no evidence was found to suggest cross-cultural differences in the degree of masculinity among men or femininity among women.

In contrast, the discrepancy between ideal and actual self-concepts varied considerably across cultures, and much of this variation was systematic. In particular, the differences between men's and women's actual and ideal self-concepts were smaller in cultures characterized by greater socioeconomic development, higher rates of female employment outside the home, higher rates of female enrollment in universities, and modern sex role ideologies.

## **6. RELATIONS BETWEEN WOMEN AND MEN**

### **6.1. Mate Preferences**

Social scientists have often considered mate preferences to be culture bound and therefore inconsistent across cultures. However, findings by David Buss and associates suggest quite the contrary. In the most extensive investigation of mate preferences to date, these researchers collected responses from more than 10,000 individuals in 33 countries. Using two lists of potential mate characteristics, participants indicated their preferences by ranking or rating the items. Surprisingly, a high degree of similarity in the preferred characteristics of mates emerged for women and men across cultural groups. For the 13 items considered, both sexes ranked "kind and understanding" first, "intelligent" second, "exciting personality" third, "healthy" fourth, and "religious" last. However, women tended to place a higher value on a potential mate's earning capacity than did men, and men generally valued physical appearance more than women. In a reanalysis of the Buss data, Alice Eagly and Wendy Wood also found

that men valued a good cook and housekeeper more than women.

In contrast to the aforementioned similarities, cultural differences were found for nearly every item. The greatest variation was found for the item "chastity." Groups in China, India, and Iran placed considerable importance on this item, but those in northern Europe viewed it as relatively unimportant. In general, men valued chastity more than did women.

Considering the remarkable degree of similarity across samples, Buss suggested that substantial unity exists in human mate preferences, and that this unity may be regarded as "species typical." However, no sample was identical to another, and each group showed unique tendencies in the ordering of characteristics, indicating at least modest degrees of cultural variation.

Not only are there similarities along with some differences in preferences for human mate characteristics, but the recent International Sexuality Description Project of 16,288 people across 10 major world regions revealed similar findings regarding desire for sexual variety. Universal sex differences were found, with men possessing greater desire than women for a variety in sexual partners, for less time to elapse in a relationship before consenting to sexual intercourse, and for more active seeking of short-term sexual partners. However, for each of these sex differences, there was significant variation by world region. For example, compared with other regions of the world, East Asian men and women were lower in their desire for more than one partner in the next month.

### **6.2. Romantic Love**

Romantic love and intimacy, as with mate preferences, are assumed to be primarily under cultural influences. In general, romantic love is valued in more modern cultures, such as the United States, in which individualism is emphasized and few have extended family ties. However, romantic love is viewed as less essential in cultures in which strong family ties reinforce the family relationships brought about through marriage, such as in Japan. There does appear to be some degree of universality when one considers an analysis of the ethnographies and folklore materials from 166 societies. This review found at least one incidence of passionate love in 147 (88%) of the cultures examined, suggesting that the potential for romantic love may be a human universal that is overwhelmed by other cultural variables.

A study by Josephine Naidoo addressed these issues with Asian Indian immigrants in Canada by examining

generational changes in the attitudes reported toward love and marriage. Although nearly two-thirds of the immigrants had arranged marriages, a substantial proportion of them believed that “love marriages” were an option for their offspring. More than 70% of the second generation believed love should precede marriage and wanted their children to have greater freedom in mate selection than they had experienced.

### 6.3. Harassment and Rape

Kaisa Kauppinen-Toropainen and James Gruber conducted one of the few cross-cultural studies of male harassment and hostility toward women. After examining blue-collar and professional women in Scandinavia, the former Soviet Union, and the United States, these authors found that Americans reported the most woman-unfriendly experiences. Relative to Americans, Scandinavians reported fewer job-related or psychological problems, more autonomy, and better work environments. Professional women in the former Soviet Union reported fewer negative experiences than blue-collar women but more than professional women in the United States and Scandinavia.

Colleen Ward and colleagues conducted a comprehensive assessment of attitudes toward rape victims with university students from 15 countries. Cross-cultural variation in attitudes toward rape victims followed patterns of attitudes toward women more generally. That is, countries with more modern sex role ideologies reported more favorable attitudes toward victims, whereas less favorable attitudes emerged in countries with lower percentages of women in the labor force and lower literacy rates. In particular, more favorable attitudes were reported by university students in Germany, New Zealand, and the United States, and less favorable attitudes were found in India, Malaysia, Mexico, Turkey, and Zimbabwe.

## 7. GENDER DIFFERENCES AT THE CULTURAL LEVEL

### 7.1. Masculinity and Work-Related Values

Using attitude survey data from thousands of IBM employees who were predominantly male, Geert Hofstede compared work-related values in 40 countries. One scale derived by Hofstede focused on the extent to which values of assertiveness, money, and things prevail instead of values of nurturance, quality of life, and people.

This scale was labeled Masculinity (MAS) due to the fact that male employees assigned greater importance to the first set of values, whereas females assigned greater importance to the second set. A MAS index was computed for each of the countries in Hofstede's study; the five highest rated countries on the MAS dimension were Japan, Austria, Venezuela, Italy, and Switzerland, whereas the five lowest rated were Sweden, Norway, The Netherlands, Denmark, and Finland. High MAS countries were characterized by stronger beliefs in independent decision making, stronger achievement motivation, increased job stress, and a view of work as central in people's lives.

The significance of the MAS dimension is unquestionable; however, many researchers have challenged the appropriateness of labeling the scale Masculinity. This label leads to the expectation that variation on the MAS scale would be associated with cross-cultural variation on other gender-related constructs. Contrary to these expectations, Williams and Best have not found relationships between their sex stereotype or masculinity/femininity measures and Hofstede's MAS dimension. Other researchers have also failed to find the expected relationships.

### 7.2. Gender Division of Labor

Conceptions of masculinity and femininity vary widely across cultures, but two universals are plausible: (i) To varying degrees, every society assigns traits or tasks on the basis of sex, and (ii) the status of women is inferior to the status of men in every society. As one would expect based on these generalizations, extensive differences do exist in the work roles of men and women. Examining jobs and tasks in 244 societies, Roy D'Andrade found that men were involved in hunting, metal work, and weapon making and tended to travel further from home than women. Women were responsible for food preparation, carrying water, caring for clothing, and various child-rearing responsibilities. Although women's subsistence activities generally included child-rearing demands, some did hunt in societies in which this activity did not compete with child care. The strong sex segregation for child-rearing duties was mirrored by another study that found that men were significant caretakers in only 10% of the 80 cultures examined. However, both sexes seemed to be flexible enough to adapt to a range of socioeconomic roles.

Today, women account for a substantial proportion of the world's labor force. With decreases in infant mortality and fertility, women now spend less time in child-rearing

roles. Furthermore, technological advances have allowed women in many parts of the world to separate child-bearing from child-rearing and thereby contribute to the family through jobs outside the home.

However, women's increased autonomy has not been paralleled by increased acceptance or equality. For example, in a 56-country study of labor trends from 1960 to 1980, the job market was marked by declines in women's occupational opportunities and increases in sex segregation. When measured by per capita gross national product and women's level of education, modernization was associated with increased segregation of the sexes. Additionally, increased workplace involvement for women correlated with decreases in total fertility rate. Women continue to be disadvantaged in the workplace, most overtly through persistent salary discrepancies that favor men. In addition to women's lower salaries is evidence suggesting that women prefer traditionally female jobs, especially those offering extensive contact with other people. Moreover, these jobs tend to be low paying. On the contrary, men tend to prefer jobs with high income and promotion opportunities.

Even in the countries with the highest proportion of females in the labor force, women continue to face inequality within the home. Studies in several North American and European countries have found that women perform a majority of the housework, regardless of the extent of their occupational demands. Along with children and larger homes comes reduced male involvement in domestic chores. This is surprising in light of the previous suggestion that increases in education and income are associated with more modern sex role views (i.e., equality in the workforce). However, studies suggest that systematic differences in sex role ideology persist in these more modern countries. For example, in the United States, Great Britain, West Germany, and Austria, people with relatively high levels of education and women with employed husbands indicated less support for efforts to reduce gender inequality compared to those with less education and women without employed husbands. Such findings suggest that even the subjective change in perceived life quality associated with improved socioeconomic conditions may be greater for men than women.

## 8. CONCLUSION

Sex differences continue to pose challenging questions for social and biological scientists, and with increased

interest in the role of culture, new questions will certainly emerge. Although males and females are biologically more similar than different, they lead different lives, regardless of their society's tendencies toward traditional or egalitarian views.

Considering the degree of gender stereotype similarity across cultures suggests a pancultural model in which cultural nuances produce relatively small variations around general tendencies. Within this model, the relatively modest biological differences are perceived to set the stage (e.g., women bear children, men have greater physical strength) and subsequently lead to a division of labor such that women are responsible for child-rearing and other domestic duties, whereas men are providers (e.g., hunting) and protectors. From the perspective of this model, gender stereotypes evolved to support this division of labor. Gender stereotypes lead to an implicit assumption that members of either sex have or can develop characteristics that are consistent with one's assigned role. Once they are established, such stereotypes can guide differential sex role socialization for boys and girls and lead boys to become more independent and adventurous, whereas girls become more nurturant and affiliative. Additionally, such boundaries heighten the salience of people who differ from societal expectations of gender-appropriate behavior, possibly creating social difficulties for those people. Although the accuracy of such a model is difficult to test due to its reliance on evolutionary processes, its usefulness is apparent because it demonstrates a means by which people in widely different cultures can come to associate one set of characteristics with men and another with women, with only minor deviations from these central themes. The relatively modest biological differences between the sexes appear to be enhanced or diminished by cultural practices; however, even small differences in gender roles and behaviors can have important cultural implications.

### See Also the Following Articles

Gender and Education ■ Gender Role Development  
 ■ Occupation and Gender ■ Sexual Harassment

### Further Reading

Best, D. L., & Thomas, J. J. (2003). Cultural diversity and cross-cultural perspectives. In A. H. Eagly, A. E. Beall, & R. J. Sternberg (Eds.), *The psychology of gender* (2nd ed.). New York: Guilford.

- Edwards, C. P. (1992). Cross-cultural perspectives on family-peer relations. In R. D. Parke, & G. W. Ladd (Eds.), *Family-peer relationships: Modes of linkages*. Hillsdale, NJ: Erlbaum.
- Hoyenga, K. B., & Hoyenga, K. T. (1993). *Gender-related differences: Origins and outcomes*. Boston: Allyn & Bacon.
- Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.
- Yoder, J. D. (2003). *Women and gender: Transforming psychology* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.



# Gender and Education

Linda M. Raffaele Mendez

University of South Florida, Tampa, Florida, USA

1. What Is Gender Equity?
  2. Gender Equity in Schools: Issues of Concern
  3. Increasing Gender Equity: Strategies for Educators
  4. Summary
- Further Reading

**unified discipline approach** A method of handling discipline that involves partnerships and consistency between schools, families, and communities rather than conflicting or contradictory policies between these groups.

## GLOSSARY

**academic retention** A school's decision to keep a student in a grade for another year rather than having the student matriculate to the next higher grade.

**the boy code** A set of unwritten rules that require boys to act tough, independent, and confident even in situations in which they feel frightened or unsure.

**classroom dynamics** The interactions occurring between individuals in a classroom.

**disorders of body image** Negative perceptions and attitudes about one's own body (often manifest as being too fat among girls or not having enough muscle among boys).

**externalizing behavior** Disruptive behavior, such as non-compliance or aggression.

**gender equity** Girls and boys experiencing equal opportunities in educational settings.

**gender stereotypes** Commonly held beliefs about characteristics deemed appropriate for males and females.

**relational aggression** A form of aggression that damages others' relationships (e.g., social exclusion and rumor spreading).

**self-contained classroom** A classroom serving only students in special education, most commonly because of a perception that the students' needs cannot be accommodated in the general education classroom.

In most industrialized countries, laws have been established to mandate equal educational opportunities for boys and girls. For example, in 1972, the United States passed Title IX, a federal law that mandates that no person be discriminated against on the basis of sex in any educational program receiving federal funds. One might assume, then, that gender does not have much of an impact on education (i.e., that there are no appreciable differences in how girls and boys experience school). In contrast to this assumption, research shows that gender can have a major impact on a child's educational experiences. This article reviews the literature on gender and education with a particular focus on the issue of gender equity.

## 1. WHAT IS GENDER EQUITY?

When applied to education, the term gender equity refers to girls and boys experiencing equal opportunities in educational settings. The term not only refers to academic learning but also applies to a much broader range of activities and experiences that are a part of the educational process. In an equitable educational environment, boys and girls have similar opportunities to

master the academic curriculum, experience a sense of comfort and belonging, have interactions with teachers and peers that promote both competence and confidence, be recognized for their contributions, and participate in extracurricular activities that are enjoyable and build self-esteem. In order to create such an environment, educators must understand and appreciate boys' and girls' unique developmental issues and needs within schools. Educators must also seek to eradicate gender stereotypes that limit opportunities for youth (e.g., that girls cannot be successful in math or that boys should not be interested in cooking or sewing). Funding (e.g., of athletic teams) and access (e.g., to vocational programs that lead to higher paying jobs) also must be given careful consideration in a gender-equitable educational environment.

## 2. GENDER EQUITY IN SCHOOLS: ISSUES OF CONCERN

### 2.1. Concerns for Girls

Much of what has been written about gender equity in education has focused attention on educational inequity for girls. Frequently noted concerns include girls receiving less teacher attention than boys, biased curricular materials presenting women in subordinate roles to men and/or failing to recognize the significant accomplishments of women (e.g., excluding the work of female scientists from textbooks), attitudes toward sexual harassment that accept such behavior as being a normal part of the adolescent school experience, and stereotypes indicating that girls are less able than boys in mathematics and the sciences.

Several publications from the early to mid-1990s did much to raise awareness of gender inequity in schools. For example, David and Myra Sadker, in their book, *Failing at Fairness: How Our Schools Cheat Girls*, summarized research showing that boys stand out in the typical classroom (receiving both more positive and more negative attention than girls), whereas girls quietly do their work and blend into the classroom landscape. It is notable that research has shown that low-achieving girls are less likely to be referred for special education services than low-achieving boys, even when their levels of achievement are similar. This is because low-achieving boys tend to manifest more behavior problems, which makes them stand out more. Also, when girls are in need of a self-contained classroom experience, especially a classroom for children

with emotional or behavioral problems affecting their learning, there tend to be very few appropriate classroom placements since these classes are dominated by boys.

Importantly, many of the types of problems that are more common among girls than boys (particularly in adolescence) and may be related to diminished opportunities for classroom learning (e.g., adolescent depression, eating disorders, disorders of body image, sexual harassment, and relational aggression) often do not come to the attention of teachers or other school personnel because they tend not to disrupt classroom activities. Thus, teachers may not know that girls in their classrooms are underperforming because of unaddressed social or emotional issues. Boys are much more likely to act out their problems than girls and thus are also more likely to come to the attention of those who might be able to provide assistance (e.g., school psychologists and counselors).

### 2.2. Concerns for Boys

Systematic inquiry into the ways in which boys are underserved in schools has only begun to emerge in the past 5–10 years. One important source of concern for boys centers on the fact that most educational environments, while becoming increasingly tolerant of girls engaging in traditionally male activities (e.g., playing physical contact team sports), are still highly stereotyped for boys. For example, William Pollock, in his book, *Real Boys: Rescuing Our Sons from the Myths of Boyhood*, noted that boys are more likely than girls to be criticized by teachers for expressing feelings of fear (e.g., on the first day of school) or affection (e.g., hugging another child). Male peers also are likely to chastise boys for any behavior that is perceived as feminine or as being “like a girl” (e.g., boys whose athletic skills are not well developed might be criticized as “throwing like a girl”). These observations exemplify the fact that boys are still restricted by what Pollock calls “the boy code,” which stipulates that boys must act tough, independent, and confident even in situations in which they feel frightened or unsure. Most educators are much more likely to tolerate cross-sex interests and behavior from girls than boys.

Another important concern regarding educational equity for boys is related to the observation that so many more boys than girls experience academic or behavioral failure in school. Boys are considerably more likely than girls to be retained, to be identified as having special needs and placed in a special education program, and to

experience school discipline (e.g., suspension and expulsion). Whether such gender differences are related to biological or social factors (or a combination of the two) is unclear. Nonetheless, given these outcomes, the possibility must be considered that the types of educational environments currently existing in our schools are not meeting boys' unique needs. Some researchers have noted that many boys begin kindergarten behind their female peers in terms of fine motor skills, verbal skills, attentional capabilities, and other important indicators of readiness for formal learning. This appears to be due to slower overall maturation in boys compared to girls. However, instead of conceptualizing boys as having different needs in early elementary school than girls (e.g., more opportunities for learning through gross motor as opposed to fine motor activities), educators tend to be quick to label boys as deficient, leading to early elementary retention in grade or referral for special education testing. If educational environments gave more careful consideration to boys' unique needs, gender differences in educational failure might be reduced.

### **3. INCREASING GENDER EQUITY: STRATEGIES FOR EDUCATORS**

There is much that educators can do to ensure that children's educational experiences are consistent with their needs. With regard to enhancing gender equity in educational settings, there are three primary areas to be considered: classroom dynamics/instructional style, the curriculum, and behavior management/intervention.

#### **3.1. Classroom Dynamics/ Instructional Style**

As noted previously, David and Myra Sadker's work focused on the fact that, in general, girls receive less teacher attention in the typical classroom than boys. As such, boys experience the benefits of a more intense educational climate, receiving both more attention from teachers overall and more precise attention (e.g., greater feedback when responding to a teacher-posed question) than girls. It is likely that the discrepancy in the way teachers treat boys and girls is related to gender differences in activity level and interactional style, with boys demanding more attention through their behavior than girls. Nonetheless, differences between boys and girls in teacher attention could be reduced if teachers were more aware of these issues.

Thus, one of the primary means for increasing gender equity in classrooms is to educate school staff about the ways in which they may be treating boys and girls differently. Because at least some of this behavior is unintentional, it can be helpful for teachers to have someone such as the principal, school psychologist, or guidance counselor (i) educate teachers about the truths and myths of gender differences (e.g., boys are not inherently better at math and science than girls), (ii) train teachers directly in gender-fair teacher-student interactional practices (e.g., through modeling and/or role-playing gender-fair expectancies, practices, and attitudes), and (iii) observe and provide feedback on the degree to which teachers are engaged in gender-fair teaching methods and strategies. Many teachers may not realize that they never ask a girl to help if they are having a problem with the TV or VCR or that they communicate to boys that they should act tough even if they are hurt or scared. Bringing these issues to awareness is the first step in creating positive change.

#### **3.2. The Curriculum**

A second important issue to address is the curriculum. It is important that educational materials used in schools (e.g., textbooks, posters, media, and Internet resources) represent the diversity of the human experience. In terms of gender, this includes depicting both men and women of various ethnic and racial backgrounds in positions of leadership and achievement. Students should have the opportunity to learn about the important contributions that both women and men have made to society and to see depicted in their educational materials role models who help them to envision possibilities for themselves that defy gender role stereotypes. It also includes balancing the number of books or other educational materials created by men and women as well as the gender of the strong central figure in these materials.

In addition to carefully considering educational materials, there are a number of other ways for school administrators to reflect to students, staff, and the larger community that students at their school will not be limited by traditional gender role stereotypes. These include developing nontraditional career mentoring programs, providing similar levels of financial support for girls' and boys' athletic teams, and establishing awards programs that recognize teachers for excellence in promoting gender equity and students for significant accomplishments in areas traditionally considered to be more appropriate for the opposite

gender (e.g., girls in science or math and boys in writing or the arts). Additionally, at the middle and high school levels, it is particularly important for schools to find proactive ways to address the problems of teenage pregnancy, sexual harassment, and sexual violence. These are concerns that many school administrators (and parents and teachers as well) do not believe should be discussed in schools. Nonetheless, research demonstrating the disproportionate negative impact of these issues on young women's schooling experiences is cause for concern. It also is recommended that schools have a policy prohibiting sexual harassment. It is particularly important that this policy include student-to-student sexual harassment because this is the primary form of sexual harassment that students experience in school.

Finally, with regard to the curriculum, educators must address the question of how well the curriculum at each grade level is meeting the needs of both boys and girls. For example, if twice as many boys as girls are being retained in kindergarten, the question of whether the current kindergarten curriculum is meeting the needs of boys must be raised. One sex should not set the standard to which the other sex is compared. Differences in development and unique needs must be recognized. Similarly, if girls are not opting to take upper-level science and mathematics courses in high school, the reasons behind this should be investigated. Are those courses being taught in a way that optimizes girls' learning potential? Are girls choosing other alternatives because they have been socialized (even inadvertently) to believe that they cannot be successful in these courses? Do they realize that their college options may be limited without this coursework in high school? These types of questions can help educators assess the degree to which they are meeting unique gender needs at different levels of the curriculum.

### **3.3. Behavior Management/ Intervention**

It has been noted that boys are greatly overrepresented with regard to grade retentions, referrals to special education, and school disciplinary action compared to their female peers. In part, this is because many more boys than girls demonstrate problems with externalizing behavior in the classroom (e.g., noncompliance and aggression). Thus, in order to meet boys' unique needs, it is very important, from a preventive

standpoint, to put in place behavioral strategies that attempt to curtail disruptive classroom behavior. Most school suspensions—which have a disproportionate impact on boys—are for classroom disruptions and disobedience. Research has shown that such behaviors can be reduced through preventive programs that involve positive behavior management, social skills instruction, a unified discipline approach (involving partnerships between schools, families, and communities), and shared expectations for socially competent behavior.

It also is important to recognize that in schools, problems that cause disruption in the classroom are those that are most likely to be recognized and further assessed. Thus, the fact that more boys than girls express problems outwardly (e.g., through fighting) while more girls than boys tend to internalize their problems means that boys' problems come to the attention of teachers more often than girls' problems. Boys are thus more frequently referred to the school's student support team for assistance. Educators need to become more aware of the types of problems that are more common in girls than boys (e.g., learning problems that occur in the absence of behavioral problems, depression, eating disorders, and relational aggression), the warning signs associated with these problems, and the necessity of involving members of the student support team in intervening with girls in these situations.

Additionally, it is important to recognize that there are sex differences in the degree to which boys and girls feel comfortable sitting down with an adult to discuss their thoughts and feelings. William Pollock noted that boys have been socialized to believe they should handle their problems on their own rather than reach out to others for help. As a result, it is less likely that boys, compared to girls, will readily open up to a teacher about their feelings when asked about a problem. This may be particularly true when the teacher is female and/or of a different racial background. Pollock's research suggests that in such situations, boys may feel more comfortable focusing on an activity (e.g., games and role plays) and having discussions of thoughts and feelings arise out of the activity rather than simply being asked to share what is on their minds.

## **4. SUMMARY**

This article has described how gender impacts educational experiences and what can be done to

increase gender equity in education. The resources listed in Further Reading give interested readers additional insight into how gender role stereotyping can limit opportunities for children and what can be done to ensure that both boys and girls receive equal opportunities to maximize their own personal potential.

### See Also the Following Articles

Academic Interventions ■ Aggression and Culture ■ Attention Deficit Disorders: School-Based Interventions ■ Educational Achievement and Culture ■ Gender and Culture ■ Learning

### Further Reading

- Gurian, M. (1999). *A fine young man: What parents, mentors, and educators can do to shape adolescent boys into exceptional men*. New York: Tarcher/Putnam.
- Jossey-Bass. (2002). *The Jossey-Bass reader on gender in education*. San Francisco: Jossey-Bass.
- Pollock, W. (1999). *Real boys: Rescuing our sons from the myths of boyhood*. New York: Owl Books.
- Sadker, M., & Sadker, D. (1995). *Failing at fairness: How our schools cheat girls*. New York: Touchstone.
- U.S. Department of Education. (2000). *Trends in educational equity of girls and women* (NCES Publication No. 2000-030). Washington, DC: National Center for Education Statistics. [Available online at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000030>]



# Gender, Overview

Florence L. Denmark

Pace University, New York, New York, USA

1. The Difference between Gender and Sex
  2. Gender Acquisition in Childhood
  3. The Harmful Effects of Gender Stereotyping
  4. Gender's Sexual Component
  5. Marriage and Parenthood
  6. Gender and Old Age
  7. Feminism
  8. Suggestions for Applied Psychology
- Further Reading

## GLOSSARY

**ageism** The discrimination of individuals based on societal stereotypes of age and aging.

**blatant sexism** The transparent harmful and unequal treatment of an individual based on his or her sex.

**covert sexism** The intentional, hidden, and hostile sexism focused on assuring the failure of a member or members of a particular sex.

**cultural feminism** Type of feminism that emphasizes the qualities and characteristics generally applied to women, such as the qualities of caring, nurturing, and concern for the needs of others, which society usually devalues and ignores.

**feminism** The belief that women and men are equal and should be equally valued as well as have equal rights.

**gender** A social construction based primarily on sex, consisting of traits, interests, and behaviors ascribed to each sex by society.

**liberal feminism** Type of feminism that focuses on the promotion of legal and social equality between men and

women; advocates for a revision of values, laws, and mores associated with the inequality of women in society.

**radical feminism** Type of feminism that focuses on men's control over and oppression of women, which remains the primary source of women's oppression economically and socially.

**sex** The biological differences in genetic composition and reproductive structures and functions of men and women.

**sex discrimination** The unequal and harmful discrimination of individuals based on their sex.

**sexism** The experience of discrimination based on sex.

**socialist feminism** Type of feminism that focuses on social relationships and the preservation and promotion of male dominance in social institutions and that seeks to exact economic institutional change in order to achieve equality between men and women.

**subtle sexism** The expression of internalized negative cultural messages regarding the inferiority of members of one sex.

Gender remains an important, and often overlooked, consideration in psychological research, theory development, and practice. The cultural, structural, and cognitive components of gender contribute to the life experiences of both women and men. Cultural pressures, along with gendered cognitions, affect people of both sexes differently throughout the life cycle. Gender stereotypes and expectations regarding gender appropriate behavior bombard men and women from birth. Often these gender role expectations create difficulties for individuals socially, sexually, physically, and psychologically. The concept of feminism, loosely

defined as a belief in the equality of women and men, provides a framework for the treatment and study of individuals of both sexes. The articles contained in this section address a broad range of gender issues, including gender development, cultural impact on gender, gender and sexuality, homosexuality, homophobia, sexual harassment, rape, and issues pertaining to women's health in old age. Sensitivity to the cultural, structural, and cognitive origins of gender and gendered behavior contributes to the field of applied psychology in considering the different experiences of men and women throughout the life cycle.

## 1. THE DIFFERENCE BETWEEN GENDER AND SEX

The terms sex and gender are often misunderstood and misused. The term sex refers to the biological differences in genetic composition and reproductive structures and functions of almost all organisms. While the actual number of sexes existing in non-mammalian species remains controversial (some organisms are asexual, or have both male and female sex characteristics), scientists currently recognize two biological sexes in mammals, male and female.

Human males and females differ in their chromosomal makeup, genetically, hormonally, and neurologically. These differences create confusion and widespread misunderstanding regarding male and female biological processes and the ways in which these processes interrelate. These misunderstandings often create unfounded beliefs and confusion regarding the distinction between the male and female sex. For example, many individuals believe that females alone produce the hormones estrogen and progesterone. In addition, many hold the belief that only males produce testosterone. However, these hormones exist in both sexes to varying degrees depending on individual genetic composition, development, and biological sex.

While species may be divided in terms of sex, only humans are differentiated by both sex and gender. Gender, although based on primarily on sex, remains a social construction. Societies ascribe various gender differences based on traits, interests, and behaviors. Social context, the social sphere in which an individual exists and interacts, remains the greatest determinate of the designation of gender, that of "masculine" and "feminine." Personal history, immediate situation, society, and culture comprise the social context in which

individuals behave and interact. This interaction, the reciprocal activities that encompass being male and female, continues throughout day-to-day existence and society as individuals abide by society's prescriptions for gender.

### 1.1. The Construction of Gender

Gender remains a socially constructed concept that refers to the creation and explanation of the differences between girls and boys and women and men. The socially constructed nature of gender debases the concept of absoluteness in the biological differences between women and men. Therefore, biological differences between the sexes account for very little of the observations of different behaviors and attitudes among women and men. Rather, the social construction of gender and the societal pressures regarding behavior and attitudes deemed gender appropriate contribute to the manifestation of gender in individuals. The acquisition of gender-appropriate behavior occurs through several psychological processes involved in the etiology of beliefs regarding the adoption and maintenance of gender and gender-related activities.

Currently, there exist three approaches to conceptualizing the evolution of the construct of gender: the cultural approach, the structural approach, and the cognitive approach. The cultural approach suggests that the acceptance of culture and society's values and norms occurs through societal rewards and that punishments contribute to the appearance of gendered behavior. Through rewards and punishments, children become socialized to behave according to cultural and societal demands. The acceptance of cultural gender values continues throughout the life cycle and leads to the reinforcement and eventual perpetuation of gender stereotyping.

Another approach, the structural approach, suggests that the constant evolution of structural and social role requirements guide individual experiences in both the societal and family levels. The structural approach places less emphasis on the socialization of gender-typed behaviors and traits; rather, it examines the common positions occupied by certain groups within the existing social structure. For example, researchers found that occupational roles outweighed gender information as a strong determinant of traits assigned men and women.

Stereotyping is another cognitive process by which gendered beliefs become perpetuated and solidified. Stereotyping presents an avenue through which information becomes cognitively categorized in order for

individuals to process unmanageable amounts of complex information. The formation of stereotypes allows people to form manageable categories based on similarities in which newly encountered people and objects fit. Although stereotypes represent highly efficient modes of organizing information, they often become generalized and oversimplified. Stereotypes represent widely shared and socially validated beliefs or cognitions regarding whole groups of people that become applied to individual members of that group. Stereotypes differ from prejudice, as prejudice refers to an attitude toward a group that often manifests itself as hostile and negative. Therefore, a stereotype represents a cognition or thought, whereas prejudice represents an attitude and often a behavior brought about by that attitude. The cognitive process and the inherent danger of stereotyping applies to gender in addition to ethnicity, socioeconomic class, and religion, for example.

## 2. GENDER ACQUISITION IN CHILDHOOD

Gender socialization, or the process of learning gender roles, begins at the onset of human life and occurs through cultural, structural, and cognitive avenues. The first statements at the birth of a child include references to the sex of the infant. Parents and friends alike agonize over the purchase of blue or pink gift items according to the sex of the newborn. Beyond the acquisition of items that are “blue for boys” and “pink for girls” and the inherent information presented by these items, infants encounter different treatment based on their sex. Adults, whether knowingly or not, handle infant girls more gently than infant boys, and comment upon the delicate features of girls but on the strong features of male infants. For example, as soon as 24 hours following the birth of a child, mothers and fathers describe female infants as finer featured, softer, smaller, and more inattentive as compared to male infants. This occurs despite the lack of significant differences between female and male infants in both birth length and weight. This difference in treatment occurs even prior to children’s acquisition of language and before they become cognitively aware of gender-appropriate behavior.

The acquisition of gender identity becomes apparent in children by the age of three. Social learning theory offers an explanation for the development of gender roles and gender-typed preferences at such an early age. It suggests that children learn gender-appropriate

behavior through the observation of others. This occurs through observational learning in which the child watches the behavior of others, parents and other adults alike, and mimics their behavior. In addition, adults reinforce children for acceptable gendered behavior and punish them for violating gendered expectations. For example, if a young boy begins to play with his sister’s doll, his parents may look at him disapprovingly or even remove the doll from his play area. In addition to experiencing reward contingencies directly, children witness vicarious reinforcement. Vicarious reinforcement refers to the observed reinforcement or punishment of parents, friends, teachers, and television characters for their gender-consistent or -inconsistent behavior, which children imitate depending on the anticipated reward. The reinforcement of observed others sends messages to the child regarding the acceptability of the observed individual’s actions. Therefore, children receive information regarding not only the actual mechanics of behaviors but also the social value in imitating them.

The process of gender socialization also takes place bidirectionally, with children participating in their own socialization. Children cognitively categorize the information they actively learn through the reinforcement and modeling of gender-appropriate behavior. Cognitive-developmental theory suggests that children desire to display prescribed gender-appropriate behaviors. While reward contingencies shape some gendered behavior, children also form their own preferences regarding toys, activities, and same-sex individuals. In a cyclical fashion, social and cognitive processes maintain these gender preferences, and both encourage the reinforcement and performance of gender-consistent behavior. In turn, the performance of appropriate gendered behaviors increases the reinforcement of gendered activities and cognitions.

Children often prefer to wear gender-appropriate clothing and play with same-sex peers. Often this occurs due to children’s own inability to remember or categorize same-sex behaviors and traits correctly. Therefore, children use general in-group (individuals perceived as belonging to the child’s own group) and out-group (individuals perceived as not belonging to the child’s own group) favoritisms to guide their own behaviors. For example, a girl may express the belief that her same-sex peers are more friendly, honest, and humorous than the same-sex boys group. However, the boys feel the same way toward their same-sex peers and feel negatively toward the opposing group of girls.

Child peer preference also occurs due to the propensity of parents and child-care facilities to segregate children into same-sex play groups based on gender-typed activities. These activities (such as sports or group projects) encourage bonding among their participants and the formation of in-group and out-group stereotyping and grants children a venue for developing more same-sex relationships. In addition, children, like adults, pay attention to and remember behavior consistent with their gender stereotypes. Conversely, gender-inconsistent behavior passes unnoticed and unassimilated into the child's general gender schema. As a result, children learn that boys and girls normally spend time apart and in different activities.

### 3. THE HARMFUL EFFECTS OF GENDER STEREOTYPING

Due to the classification of gender into “feminine” and “masculine” qualities, traits, behavior, or activities, gender becomes prone to stereotyping. Due to the salience and visibility of sex, it becomes the basis for cognitive categorization (age, race, and ethnicity offer some other examples), and a number of gender stereotypes become founded on the perception of an individual as a member of the male or female sex. Whether stereotypes actually reflect differences between men and women remains debatable. Biology, the environment, or an interaction of the two may actually be responsible for differences in males and females. However, applying stereotypes to individuals presents some danger, despite any apparent truth at the group level. The deeply engrained and subtle nature of gender-associated processes causes most individuals to remain unaware of their impact on behaviors, attitudes, cognitions, language, interpersonal relations, and the structure of social institutions.

#### 3.1. Sex Discrimination and Sexism

Gender stereotyping becomes problematic, especially for women, considering that judgments of female gender are formed in regard to ambiguous and intangible traits such as an individual's “congeniality,” “warmth,” or “interpersonal communication skills.” Research conducted during the 1970s and 1980s indicated that women were portrayed as more expressive and communal. Men were viewed as more competent, agentic, and instrumental. In addition, these gendered differences in

beliefs regarding personality traits superseded beliefs regarding appropriate roles and occupations.

Stereotypes of women also differ based on race, class, or ethnicity. For example, traits such as “warm,” “vain,” “dependent,” and “intelligent” were frequently applied to White women, while traits such as “dirty,” “hostile,” “emotional,” and “superstitious” were generally applied to Black women. In addition, people view middle-class women as more “intelligent,” “ambitious,” “competent,” and “happy.” Lower class women became viewed as “talkative,” “irresponsible,” “dirty,” and “dependent.” Stereotypes of women become especially hazardous considering sex discrimination and sexism.

##### 3.1.1. Sex Discrimination

In general, the term sex discrimination refers to the harmful and unequal treatment of individuals based on their sex. Sex discrimination occurs as a function of cognitive internal processes coupled with the power men exert in invoking anti-female prejudices in everyday practices, resulting in the denial of the same privileges enjoyed by men. Women remain the frequent targets for sex discrimination, which produces greater disadvantage for them. While men experience discrimination based on other social variables (religion, age, race, or sexual preference), their gender neutralizes the experience of these sources of discrimination. However, being female remains the largest predictor of discrimination over any other variable because it is interlaced with other social variables (such as age, race, social class, and ethnicity), which produces a greater disadvantage.

##### 3.1.2. Sexism

The experience, impact, and severity of sexism varies depending on its intentionality, visibility, and ease of documentation. Blatant sexism occurs when women face transparently harmful and unequal treatment. This type of sexism includes physical violence, sexist jokes, degrading language, sexual harassment, and unequal treatment under the law or in employment. Less apparent, subtle sexism remains less visible to others and the targeted individual. More often it lacks the focus of other forms of sexism in its intention, maliciousness, and manipulateness. This type of sexism might occur when, for example, a male friend jokes about women always leaving to go to the bathroom together. The friend's comment lacks the harmful intention, manipulateness, or maliciousness of other forms of sexism. This type of sexism can be easily

documented; however, it often passes unreported due to the lack of a perceived threat.

Sexism in the form of deliberate behavior designed to sabotage women's efforts socially or in the work place represents an intentional, hidden, and frequently hostile attempt to thwart women and ensure their failure. For example, a man may go out of his way to slash a woman's tires, making her unable to arrive at work on time, or he may erase the content of her computer prior to an important presentation. This type of sexism presents difficulties in documentation, as the strategies used remain easily denied.

### 3.2. Suggestions for Minimizing Harmful Stereotypes

The danger of applying gender stereotypes negatively through sex discrimination or through blatant, subtle, or covert sexism raises the question, How do parents and other adults minimize gender stereotyping? Some parents believe that by avoiding the imposition of gender role constraints and pressures their children will develop more fully. There exist different reasons for parental focus on reducing gender constraints on their children, as well as implications of gender role research on parenting practices. Parents give a number of reasons for their desire to decrease their children's exposure to gender-stereotyped roles. First, many parents believe that androgynous characteristics and gender flexibility contribute to mental health. Second, many of these parents believe that the attainment of gender equality depends on the debasement of gender roles in the separate spheres of influence traditionally inhabited by females and males (e.g., the home as the women's sphere of influence and the workplace as the man's sphere of influence). Third, they believe that men and women remain more similar than different. Finally, these parents also believe that traits belonging to both genders increase success in today's technologically savvy world. Parents with such goals sometimes find their efforts frustrated, and psychological evaluations of parenting practices oriented toward decreasing gender-stereotyped cognitions and behaviors remain mixed. This is discussed in the article on gender role development in this section.

Parents who desire to minimize the effects of cultural pressures on gender role identity formation can take several steps. First, to take advantage of children's propensity to display interest in activities preferred by same-sex individuals, parents can give examples to their

children of same-sex individuals participating in those activities. For example, if parents want their daughter to participate in track and field competitions, they should show their child pictures of a female world-class athlete winning an Olympic medal. Second, parents should present their children with a range of activities and playthings with the hope that they prefer some non-gendered objects and activities. In addition, parents should request that other family members give gifts intended for both sexes, or those intended for the other sex, and encourage their use equally. Third, due to the powerful effect of parents as role models, parents should attempt to display gender-inconsistent behavior as often as possible. Parents should also seek out other role models, such as teachers, who value nontraditional gender roles, refrain from segregating groups on the basis of gender, and attempt to minimize children's exposure to traditional gender roles through television and literature. Fourth, parents should give their children gender-neutral toys and encourage their participation in activities that both sexes enjoy, thereby decreasing the amount of time children spend in stereotypic gender play. Fifth, parents should remain consistent regarding children's exposure to gender-inconsistent and gender-neutral material and behavior. Often this includes refraining from discouraging gender-inconsistent behavior in their children, such as a boy playing "dress-up." Finally, as children approach middle childhood (ages 9 to 12), they become more cognitively developed. By teaching children critical thinking skills and challenging their dogmatic gender cognitions, parents may successfully counter the effects of gender stereotypes.

## 4. GENDER'S SEXUAL COMPONENT

Gender roles manifest themselves culturally, structurally, and cognitively in many spheres of human interaction, including sexual interaction. Sexual gender role behavior begins in childhood and develops through life. Gendered sexual behavior becomes learned early in life through biological and social reinforcements ("nature" versus "nurture"). These biological and social reinforcements occur differently for males and females and continue throughout the life cycle.

Early childhood sexual behavior begins as children explore their surroundings and themselves. Children experience pleasure through contact with their genitalia, and this pleasure serves as conditioning for future self-stimulation. Genital pleasure also occurs accidentally through the child's contact with objects that create

stimulating responses. Through exploration children learn about the pleasure that results from masturbation. These biological processes serve a purpose for the developing child, both providing the child with sexual reflexes and assisting in the development of neural pathways in the brain necessary for the experience of sexual pleasure. However, the responses of parents and caretakers to young children's self-initiated stimulation determine the ways in which children feel about sexual behavior and their own sexuality.

Parents differ in their responses to their child's self-stimulation. Some parents adopt a normative view regarding their child's exploration and self-touching. Parents who allow their children to explore their sexuality allow positive Pavlovian conditioning and decrease sexual inhibitions. However, some parents regard self-pleasuring behavior as abnormal, and according to cultural beliefs as dirty or sinful, and often use punishment in response (e.g., spankings, harsh tones, or physical reprimand). Unfortunately, children receiving harsh punishment for self-touching repress their masturbation and experience long-lasting inhibitions. Many parents allow some forms of self-touching but punish others, thereby inhibiting and punishing their child.

Young children encounter many different double standards in regard to sexuality and sexual pleasure. Baldwin and Baldwin, in their article on sexual behavior in this section, discuss parents' unequal treatment of sexual behaviors exhibited by their children. First, parents punish girls for masturbation more often than boys, leaving boys with a greater impression of acceptance of sexual self-stimulation. This causes girls and women to become more uncomfortable with exploring different forms of self-pleasuring. Second, even the language used in naming and referencing of children's genitalia gives evidence to this double standard. While boys genitalia becomes referred to as "pee-pee" or "wee-wee," parents give socially detailed information to their girls in the names they provide, such as "treasure box, to open after marriage." Parents often neglect to inform their children about the uses of their genital organs, especially their use in pleasurable activities.

Throughout their childhood, children learn about sexuality from information gathered through observation of parents, other adults, television, film, and literature. Contrary to adult beliefs, children often remain more attentive to sexual information and learn from observation. Parents often assume that children remain oblivious to the sexual content and innuendos present in media and adult behavior. However, considering children's attention to the abundance of sexually

suggestive material available to them, they are more exposed to sexually explicit material than their parents might think. Some concerned parents offer their children explanations of sexual material, thereby opening a dialog and equipping their children to better process sexual information and to incorporate the parent's values into this process. However, some parents fail to offer advice regarding sexually explicit material and risk their children accepting sexual violence as normal and natural. For example, while sexual harassment and rape represent a manifestation of power inequality between men and women in which men view women as objects designed to fill their sexual desires, learned gender stereotypes and sexual scripts contribute to the view of men as the initiators of sexual contact and of women as the gatekeepers who rebuff sexual contact.

Parental lessons taught to children regarding their sexual self-pleasure and their observations carry social implications for adolescent and adult sexual behavior. Parents tend to offer girls more guidance regarding sexually explicit media as compared to boys, possibly due to the fear of possible detriment to girls resulting from sexually transmitted infections, pregnancy, and internal injuries resulting from forcible sexual activity. Parents, adults, and media messages therefore convey a message of responsible sexual behavior for girls, but neglect to focus this message on boys. This message becomes especially potent during adolescence, when secondary sex characteristics emerge, causing males and females to appear more adult-like. Baldwin and Baldwin suggest that biological differences in males (e.g., their more visible and often spontaneous sexual arousal and orgasms, and the more visible changes in outward genital appearance) cause them to become more curious and interested in sexual activity at an earlier age. However, this level of curiosity and interest differ depending on the inhibitions and positive associations formed in childhood through interactions with parents, other adults, and media influences.

The interaction of nature and nurture during adolescence creates gender-related sexual behavior differences. Adolescent females, for example, become more concerned regarding possible pregnancy, abortion, sexually transmitted infections (STIs), and sexually transmitted diseases (STDs). In addition, certain social stigmas become evident during adolescence. For example, adolescent females become labeled as "sluts," "easy," or "loose" if they participate in sexual behavior with a number of partners. However, male adolescents receive the label of "stud," which often improves their social status or popularity, if they participate in sexual

behavior with many females. As a result, female adolescents develop more inhibitions regarding sexual behavior, and they tend to become more concerned about the number of partners they engage in sexual behavior with.

The social context of double standards in sexual behavior also contributes to the actual content of sexual activity in adolescence and often in adulthood. Adolescent females often take passive roles in sexual encounters, due to the belief that “nice girls” do not initiate or aggressively pursue sexual pleasure. In addition, females tend to feel uncomfortable and worry about how their partner perceives them if they seem too eager or display too much sexual knowledge. Conversely, actively pursuing sexual contact and obtaining vast amounts of sexual experience become the expectation for male adolescents. In addition, men often state their motivation for seeking sexual activity as oriented toward sexual pleasure or enjoyment, whereas women view their motivation for sex as the desire for love and commitment. The biological differences between men and women fail to offer any support for the differences between male and female sexual behavior. Rather, sexual behavior differences may result from the role expectations conferred upon both men and women.

#### 4.1. Cultural Views of Sexuality

In a larger cultural context, sexual behaviors and values remain important aspects of all societies. Societies differ in the various meanings attributed to sexual intimacy according to their sexual permissiveness, the expression of their sexual values, views regarding contraception and birth control, and the health issues involved. Some broad generalities exist within this infinite variety and the individual differences existing within each culture. Cultural, socioeconomic, historical, global, and technological contexts determine the commonalities and differences in beliefs and behaviors occurring in societies. The large amount of information available to people through globalization and technology allow them to engage in social comparison and augment their understanding of what is natural and normal, thereby shaping their attitudes toward sexual issues and the acceptability of different sexual behaviors.

The permissiveness or restrictiveness of the culture in which an individual lives determines sexual behavior for both genders. A culture’s technological advancement, religious orientation, socioeconomic importance placed on marriage, and significance placed on reproduction influence the degree to which it becomes permissive and accepting of sexual behavior. Positive attitudes toward

sexual development as part of healthy social adjustment characterize societies with permissive attitudes toward sexual behavior. Permissive cultures place less restrictions on sexual activities and provide children with sexual information (e.g., through ritual or institutional practice, or through exposure to nudity). Restrictive societies limit exposure to sexual information and attempt to repress sexual expression. These societies often view sexual behavior as dangerous, sinful, or harmful. This view toward sexual behavior often stems from societal values regarding the expected goals of its members. For example, parents in the United States often expect their children to obtain college degrees and marry at older ages. The possible risks of participating in sexually intimate activity (e.g., pregnancy or the contraction of AIDS) run counter to the goals of educational success and economic prosperity. Sexual activity, therefore, becomes risky in these cultures due to the possible threat to an individual’s social status, physical health, or psychological well-being. In restrictive cultures, abstinence prior to marriage is stressed, often for women more than for men.

The risks of sexual activity often dictate cultural views of sexual activity prior to marriage. Agrarian and hunter–gatherer cultures often perceive premarital sexual intimacy negatively due to the lack of availability of birth control options. In contrast, industrialized societies increasingly view premarital sexual activity more permissively due to advances in birth control technology, the increasing time gap between sexual maturity and marriage, and rising economic equality between men and women. Societies also differ in their permissiveness of other forms of sexual expression, such as homosexuality and transgenderism.

#### 4.2. Homosexuality

As with heterosexual sexual behavior, cultures display either permissiveness or restrictiveness regarding homosexual sexual behavior according to the socio-historical and socioeconomic contexts. Generally, homosexuality refers to sexual attraction to members of an individual’s own sex. Cultural differences in the definition of “sex” and the meanings attributed to different sexual behaviors between people of the same sex often lead to contrasting ideas regarding what homosexual behavior entails. Cultures with a narrow definition of sex, such as consisting solely of penile–vaginal intercourse, may accept a great deal of same-sex intimate behavior.

Gender role perception remains linked to conceptualization of homosexuality in many societies. In some

cultures, men participate in homosexual relationships yet refrain from identifying themselves as homosexual. The roles that these men undertake during homosexual activity assist them in avoiding the stigma that accompanies this behavior. These men may remain considered heterosexual due to the adherence to the masculine norms of their gender. For example, males in prison may participate in homosexual sexual activity; however, because they maintain “masculine” gender characteristics, may escape the social label of homosexual.

In other cultures, such as the United States, homosexuality is defined less according to gender roles, but rather becomes perceived as an intrinsic aspect of the individual. In cultures ascribing homosexuality as part of the individual, individuals experience flexibility in the sexual roles they assume. Increasingly, many cultures are adopting Western views regarding homosexuality, especially due to the proliferation of scientific and political information regarding a possible genetic basis for homosexuality. Due to the Puritanical nature of Western society and its history of discrimination directed toward homosexual individuals (including hate crimes), the shift in values creates a difficult atmosphere for those with other sexual orientations.

### 4.3. Transgendered Individuals

Some individuals adopt the gender roles of the opposite sex. These transgendered individuals prefer the gender roles or identities of the opposite sex and either continue to behave in accordance to the roles granted them by society or outwardly display the identity and roles they perceive as their true gender. Some transgendered individuals choose to adopt the dress and mannerisms of the other sex, or pursue sex reassignment surgery. Transgenderism does not necessarily apply to cross-dressing, because the motivation for cross-dressing does not always entail a strong preference for opposite gender roles and identities.

### 4.4. Homophobia

The tendency of other cultures to adopt Western views regarding homosexuality becomes increasingly problematic due to the reflection of Puritan attitudes regarding sexual behavior and gender roles in the United States. Individuals fundamentally threatened by homosexual attractions and relationships claim that traditional morals and family values become eroded by the acceptance of homosexuality. These attitudes rest at the root of discrimination based on sexual orientation. Individuals

identifying themselves as homosexual, or even appearing to display homosexual “traits,” experience discrimination (e.g., in the workplace, regarding club membership, in school attendance, in obtaining housing and credit, and in obtaining services at places of business) despite the outlaw of discrimination and violence directed toward individuals based on sexual orientation. Discrimination based on sexual orientation, while illegal, occurs frequently in Western society.

On the individual level, part of the discrimination against homosexuals, bisexuals, lesbians, gays, transsexuals, or transgendered individuals occurs due to homophobia, or anti-gay attitudes or behaviors. The term homophobia implies an intense fear of homosexuals (being touched, looked at, or approached by a homosexual) due to the use of the Greek word “phobia,” meaning fear. However, while some individuals exhibit an intense fear of homosexuals, the term best describes a heterosexual bias in which heterosexuality is viewed as the norm and superior compared to homosexuality. Often this bias causes individuals to ignore other forms of attraction and attitudes, focusing solely on heterosexual attractions and attitudes. The same prejudices found to contribute to the discrimination faced by women also apply to the discrimination faced by homosexual individuals. Homosexual individuals become evaluated or judged as a group toward which many members of society express negative attitudes and beliefs. Heterosexuals who adhere to a more “traditional” ideology of family and gender, who hold strong religious beliefs, and who maintain friendships with people expressing anti-homosexual attitudes tend to express stronger anti-homosexual beliefs and attitudes. Heterosexuals who experience more contact with homosexual individuals express fewer negative attitudes toward homosexuality as compared to those with little contact with homosexual individuals.

Homophobia affects homosexual individuals’ social well-being as well as their emotional and psychological well-being. Internalized homophobia occurs when homosexual, lesbian, gay, bisexual, transgendered, or transsexual individuals feel negatively about themselves based on their sexual orientation. Homosexual individuals also risk adjustment difficulties arising not only from internalized homophobia, but also from the social stigma attached to homosexuality. In addition, very pronounced fears regarding the AIDS pandemic and its labeling as a “gay” disease caused many homosexual individuals to resist their sexual orientations in order to avoid contraction of the virus. Although the stigma of AIDS as confined to the homosexual and

intravenous needle-using population abated and this belief became replaced by more informed views, many homosexual individuals still experience the stress of anti-gay sentiment.

Homosexual men and women experience high levels of stress due to their marginalization in society. Historically, the mental health field confronted homosexuality as a form of psychopathology; however, this view evolved somewhat and many mental health professionals remain supportive of gay and lesbian identities. Although both the mental health field and society are moving toward a more accepting view of homosexuality, greater attention is needed in addressing the social stressors contributing to the developmental difficulties faced by homosexual individuals.

## 5. MARRIAGE AND PARENTHOOD

### 5.1. Marriage

Marriage represents another social context in which gender plays a large part. In most societies, marriage remains an antecedent of the nuclear family, consisting primarily of a husband, wife, and children. The larger, extended family consists of those related by blood to the members of the nuclear family, for example, the grandparents, aunts, uncles, and cousins. The socio-historical context of the family frames the gender roles occurring within its influence.

During most of human history, people perceived marriage as essential due to the needs of both men and women. The labor of caring for and maintaining a family became shared by both the man and the woman, usually with the man undertaking the role of the hunter and breadwinner and the woman being responsible for the care of children. Both men and women needed the product of the other's labor, not only for their individual survival, but also for the survival of the next generation. In addition, in many cultures marriage remains beneficial in regulating sexuality, providing legal ties between women, men, and their children, and creating economic and social stability.

Women receive the greatest amount of social pressure to marry. The importance of the family remains paramount in most modern societies, despite the existence of many opportunities for women to work and participate in social life, politics, and religion. From a very young age, girls receive education regarding their eventual role as wife and mother. Women who remain unmarried, or elect to forgo marriage, face social stigma

(e.g., they receive the name “spinster”) and pity unless they take on a socially accepted role such as a nun. As a result of the social stigma attached to remaining unmarried, many women experience a feeling of inadequacy or deprivation. Consequently, families attempt to ensure marriage for their daughters (in some cultures, with arraigned marriages).

For many married women, large segments of daily life remain occupied by the roles of wife and mother, with family obligations at the forefront, despite other obligations. For example, many women in the United States work and care for their families, often maintaining the role of primary caretaker for their children as well as full-time employment. In most Western societies, household maintenance remains unequally allocated between husband and wife, often with the wife assuming the lion's share of daily domestic activities such as cooking, cleaning, laundry, and child care. Evidently, the increase of time-saving technology, the changing social landscape in regarding the increase of women in the workplace, and the rise of gender equality ideals have failed to fully evaporate traditional gender roles.

Evidence of the asymmetry in traditional marriages occurs in various facets of marriage customs, including the assumption of the husband's last name. This also occurs in the conceptualization of marriage in regard to occupations. The expectation of the wife as a “junior partner” in a man's occupation occurs presently in many societies. For example, the wife of a baker becomes the “baker's wife,” or the wife of a wealthy banker becomes the “banker's wife.” The wife becomes incorporated into the husband's occupation, as an extension of himself. The “baker's wife” may not appear on the payroll; however, she participates in the daily activities of the bakery. The “banker's wife” may appear as the social face of her husband's occupation, contributing to charities and attending social functions. Wives of politicians provide another example; society expects them to place their aspirations aside in order to participate and contribute to their husbands' endeavors. To a large extent, societal expectations of gender roles in marriage dictate the experiences of married women.

In some societies, wives remain subservient to their husbands. In fact, in most parts of the world women's rights and entitlements remain secondary to those of their husbands. A woman may lack control over finances, lack legal rights to inheritance, be considered the property of her husband, and in some societies become the ward of her children in the event of her husband's death. In past extreme cases, women in India were expected to burn themselves alive following

the death of their husband (suttee). The Indian government outlawed the practice, and the actual current incidence remains questionable. The *Code of Manu* (a list of life rules well established by 200 B.C.E.) stated, "She must on the death of her husband allow herself to be burnt alive on the same funeral pyre; then everybody will praise her virtue" (Noss & Noss, 1984; p. 181). Although suttee represents an extreme situation, many women still face societal judgment based on their ability to live up to gender role stereotypes such as maintaining a cleanly and ordered household.

In the past, societies expected individuals to marry for greater financial stability, to reproduce through engaging in socially approved forms of sexual activity, and to form family alliances. In many societies, marriage forms the foundation of families and basis of society. However, other political, social, and economic institutions in industrial urban societies often debase the perceived social function of marriage. Improved technology and economic organization decrease the need for the division of labor based on gender. However, the perception of marriage and the inherent expectations that accompany the institution contribute to the perpetuation of gender role stereotypes and expectations. In addition, due to continued inequalities for women, such as unequal pay, women still maintain strong financial and social motivations for marriage. Same-sex marriages bear mentioning here. While some countries allow same-sex marriages, homosexual couples in the United States struggle for the permission to marry. (Massachusetts is the only state in the United States currently permitting same-sex marriage.) Same-sex couples desire the social, financial, and legal security afforded to heterosexual couples. Therefore, the motivations behind same-sex couples' desire to marry remain the same as the motivation behind heterosexual marriages.

## 5.2. Parenthood

While many couples choose to marry later and improve their quality of life prior to the birth of their first child, the decision to have children often accompanies a shift toward traditional gender roles in the family. This occurs in part because of the perception of motherhood as the ultimate fulfillment of womanhood. For example, until recently, pregnant women left the workforce in anticipation of children and became full-time homemakers and mothers.

The expectation of women assuming the sole role of motherhood following childbirth grew out of the view that mothers maintained the most important influence

over the child's development. However, further research suggested that children with fathers who play an active role in child rearing develop higher intellectual capacity and higher levels of empathy toward their peers. This affects fathers as well; they display higher self-esteem in general, greater satisfaction in their role as a father, and higher confidence in their parenting abilities. Fathers sharing in household responsibilities and childcare also experience greater marital stability.

Despite the benefits of father participation in household and childcare responsibilities, the mother remains more likely to remain home with the child after birth, at least part-time. The mother, therefore, participates more fully in meeting the child's needs, primarily because she remains in the household while the father continues to work. Often this causes the woman to undertake the management of the family's emotional life (including worrying about the state of the marital relationship), the child's everyday needs, and social relations with extended family members.

Yet women in Western societies often return to work, either following the termination of their maternity leave or after their children depart the home. However, parenting demands remain stronger for women. Even with provisions for childcare, working mothers often experience conflicting demands of their jobs and their children, sometimes to the detriment of their mental health. This conflict often becomes exacerbated in the case of single-parent households.

## 5.3. Single Parenthood

Single parents face a different set of difficulties, as they assume both gender roles. Single mothers and fathers may experience role overload as they attempt to ensure their family's economic survival in addition to managing the household and finances. Employers are often insensitive to the conflicting needs of work and home, expecting full commitment from their employees. Single-parent difficulties become compounded due to the antiquated design of the workplace, which best suits men with wives who manage the household. These issues become especially difficult for single mothers, who face the same discrimination experienced by other working women; however, the situation becomes more desperate considering their status as the provider for the family.

While some single parents encounter social difficulties fitting in with married couples and encounter difficulties in dating, as the number of single parents increases, more opportunities for forming social support networks

are arising. In addition, some divorced women express a sense of accomplishment and fulfillment, while many single mothers enjoy their autonomy and independence. Single fathers also make caring for their children an important priority. Despite the difficulties faced by single parents, single parenthood offers the custodial parent an opportunity to experience both gender roles.

#### **5.4. Gay and Lesbian Parents**

Gay and lesbian parents, either single or in committed relationships, find themselves in a unique position. These parents face many problems resulting from the social stereotypes regarding their abilities as role models for children. Despite the concerns put forth by society, many homosexual parents have children from previous heterosexual relationships, adopt children, obtain the assistance of a surrogate, or undergo in vitro fertilization. Contrary to public concerns regarding gay and lesbian parents' abilities to raise well-adjusted heterosexual children, studies indicate that children raised in households with gay and lesbian parents enjoy outcomes similar to children raised in households with heterosexual parents. The raising of children by homosexual couples does not increase the likelihood of children becoming homosexual later in life. While these results show promise, homosexual parents still struggle to protect their children from the negative public opinion of gay and lesbian parenthood.

### **6. GENDER AND OLD AGE**

#### **6.1. Gender and Medical Concerns**

Men and women experience the aging process differently, both physically and psychologically. Both medical and psychological literature contain gender bias, as researchers often falsely assume either equalities or differences between the sexes. Many researchers conduct their research using only male samples and then generalize the results to females. For example, researchers receive promising results when using vitamin therapy to decrease hair loss in men and advocate the therapy for use by both men and women. However, the vitamin therapy may not work as well, or at all, in women. While this is a fictitious example, some researchers approve treatments for women regardless of their biological differences, sometimes with detrimental results. In addition, researchers also overlook the incidence of diseases, defects, and illnesses by labeling them as either

“men’s” or “women’s.” For example, many prevention strategies target women for osteoporosis. Television commercials advertising calcium consistently refer to women’s bone loss and the necessity of increased calcium consumption. However, although women have a higher rate of osteoporosis (one in two women experiences osteoporosis), men also experience bone loss (one in eight experiences osteoporosis). This may lead men to ignore the signs of bone loss or seek treatment later after serious damage occurs.

In addition to the research and prevention strategies for illnesses and disease, men and women experience different illnesses based on their biology. Women have different medical concerns throughout their lifespan (during pregnancy, menstruation, contraception, and menopause, for example). In addition, men and women differ in the incidence and severity of mental illness, due to their differing physiology and environmental stressors. Researchers and other health professionals now seem to recognize that both similarities and differences exist between females and males. In fact, the recognition of women’s health as a distinct entity attests to the growing acceptance of quantitative and qualitative differences in the effects of health-related issues on both men and women.

#### **6.2. Gender and the Psychological Experience of Aging**

Women and men also experience aging differently psychologically. Women often view aging negatively as the loss of youth characterized by the loss of attractiveness, and the loss of the ability to produce and nurture children. Men, however, view the graying of hair and wrinkling of skin as adding character. The appearance of age does not threaten the traditional gender stereotypes of men as competent, autonomous, and self-controlled; in fact, the appearance of age often serves to “add character.” This creates a double standard: as women become viewed as less attractive due to the appearance of age, men become viewed as more attractive for the very same age indicators (e.g., some extra pounds, wrinkles, and gray hair). Youth-oriented society plays a large part in reinforcing views of aging and its experience, especially for women. One study found that the promise of “youth, youthful appearance, or the energy to act youthful” appeared in 57% of monitored television advertisements. However, senior citizens in

these same advertisements dealt with loose dentures, constipation, and backaches.

### 6.3. External Perceptions of Aging

The emphasis placed on youth and the appearance of youth in culture also influences the ways in which aging individuals are perceived. Researchers found that older women were perceived as more active in the community, more personally acceptable, less self-centered, more nurturing, and more sensitive as compared to older men. However, an older woman perceived as lacking competence is judged more harshly and is perceived as experiencing cognitive decline.

### 6.4. Ageism

Because women often outlive their male counterparts and make up the majority of the elderly population, they necessitate greater medical and psychological consideration. Older men usually are cared for by women, making them less susceptible to the effects of ageism. Often women do not share this luxury. The lack of any buffers against ageism causes women to become the subject of more age-related jokes. This becomes especially problematic if the loss of her husband and the constraints of living on a restricted budget require an elderly woman to seek employment. Age discrimination occurs in large part due to the view that women's responsibilities as child caregivers decline and end in middle age, therefore placing them at the end of their responsibilities and accomplishments.

## 7. FEMINISM

Feminism represents one approach to addressing the necessity of different perspectives on gender, as opposed to a purely male-oriented perspective, in research, theory, and practice. The term feminism refers to a belief in and commitment to the social, legal, and economic equality of women. This includes the belief in the equal value of women as well as their equal rights. The broad-based perspective and the incorporation of different ideologies provide feminism with a unique frame of reference. There exist different kinds of feminism (socialist, radical, cultural, and liberal); however, each viewpoint and theory intersects and maintains commonalities with other theories of feminism.

Socialist feminism focuses on social relationships and the preservation and promotion of male dominance in

social institutions. The necessity of change in the economic system remains the primary focus of socialist feminism, as economic and sexist oppression reinforce each other in society. Inequalities stemming from race, socioeconomic class, and ethnicity also become the focus of research under this theory. This type of feminism stresses the unity and integration of all aspects of women's lives in order to produce unified feminist theory. Social feminists desire the restructuring of social relationships and social institutions to create economic equality, and therefore social equality, between men and women.

Radical feminism focuses on men's control over and oppression of women, which remains the primary source of women's oppression economically and socially. This theory advocates for the destruction of existing class and gender inequalities and the creation of a new culture and society founded upon shared power between the sexes. Radical feminists believe that women-centered beliefs and systems remain essential to the end of women's oppression. In contrast to other feminist groups, radical feminists reject all forms of feminism based in male-developed theories. In their view, women must create and develop feminism and root it in their own experiences. Radical feminist research often examines the relation between men's societal power and sexuality and violence.

Cultural feminism emphasizes the qualities and characteristics generally applied to women, such as caring, nurturing, and concern for the needs of others, which society usually devalues and ignores. Cultural feminism focuses on increasing the acknowledgement and honor of women's contributions (usually unpaid) to society and culture.

Liberal feminism, or "equal opportunity feminism," focuses on the promotion of legal and social equality between men and women. This theory advocates for a revision of values, laws, and mores associated with the inequality of women in society. In addition, the concept of males and females as more similar than different remains central to this theory, as liberal feminists believe that women and men will exhibit similar behavior if afforded equal opportunities. Liberal feminists call for equal consideration regardless of gender; they do not seek special privileges. Liberal feminism claims responsibility for revisions in language, media representation of women, and changes in religious practices, which formerly contributed to stereotypical views of women in society and inequality based on gender. Due to the stigma attached to the term feminism, many people avoid classifying themselves as feminists.

However, many individuals express beliefs in gender equality consistent with liberal feminism.

## 8. SUGGESTIONS FOR APPLIED PSYCHOLOGY

The adoption of feminist ideals regarding gender equality in scientific research, theory, and practice would serve to increase sensitivity regarding the needs of both men and women. These ideals serve to exact change in the cultural, structural, and cognitive views of gender, gender roles, and gender-appropriate behavior. The expansion of ideologies pertaining to gender role development, behavior, and attitudes to include often-neglected alternative views (such as those inherent in homosexuality and transgenderism) might serve to lessen the incidence of the discrimination of historically marginalized groups. In addition, a closer understanding of gender stereotypes and their formation along with social and cultural ideological change might contribute to fewer incidences of medical and psychological problems associated with gender discrimination, sexual harassment, and rape.

### See Also the Following Articles

Gender Role Development ■ Homophobia ■ Rape Prevention  
 ■ Risk-Taking and Sexual Behavior ■ Sexual Behavior  
 ■ Sexual Behavior and Culture ■ Sexual Harassment  
 ■ Women's Health

### Further Reading

- Benokraitis, N. V., & Feagin, J. R. (1995). *Modern sexism: Blatant, subtle, and covert discrimination* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues*, 28, 59–78.
- Carter, B. (1992). Stonewalling feminism. *Family Therapy Networker*, 16, 64–69.
- Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, 29, 80–85.
- Deaux, K., & Kite, M. (1993). Gender stereotypes. In F. L. Denmark, & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (pp. 107–139). Westport, CT: Greenwood Press.
- Demo, D. H. (1992). Parent-child relations: Assessing recent changes. *Journal of Marriage and the Family*, 54, 104–117.
- Denmark, F. L., Rabinowitz, V., & Sechzer, J. (2005). *Engendering psychology: Women and gender revisited*. Boston: Allyn and Bacon.
- Gilbert, L. A. (1994). Reclaiming and returning gender to context: Examples from studies of heterosexual dual-earner families. *Psychology of Women Quarterly*, 18, 539–584.
- Harre, R., & Lamb, R. (1986). *The dictionary of personality and social psychology*. Cambridge, MA: MIT Press.
- Hochschild, A. (1989). *The second shift: Working parents and the revolution at home*. New York: Viking.
- Keshet, J. K. (1980). From separation to stepfamily: A subsystem analysis. *Journal of Family Issues*, 4, 146–153.
- Kimmel, D. C. (1988). Ageism, psychology, and public policy. *American Psychologist*, 43, 175–178.
- Kite, M. E., Deaux, K., & Miele, M. (1991). Stereotypes of young and old: Does age outweigh gender? *Psychology and Aging*, 6, 19–27.
- Lamb, M. (Ed.). (1987). *The father's role: Cross-cultural perspectives*. Hillsdale, NJ: Erlbaum.
- Lamb, M., Pleck, J., & Levine, R. (1987). Effects of increased parental involvement on fathers and mothers. In C. Lewis, & M. O'Brien (Eds.), *Reassessing fatherhood* (pp. 108–125). London: Sage.
- Lott, B. (1995). Distancing from women: Interpersonal sexist discrimination. In B. Lott, & D. Maluso (Eds.), *The social psychology of interpersonal discrimination* (pp. 12–49). New York: Guilford Press.
- Noss, J. B., & Noss, D. S. (1984). *Man's religions* (7th ed.). New York: Macmillan.
- O'Connell, A. N., & Rotter, G. (1979). The influence of stimulus age and sex on person perception. *Journal of Gerontology*, 34, 220–228.
- Offen, K. (1988). Defining feminism: A comparative historical approach. *Signs: Journal of Women in Culture and Society*, 14(11), 119–157.
- Palmore, E. (1971). Attitudes toward aging as shown by humor. *Gerontologist*, 11, 181–186.
- Richmond-Abbott, M. (1984). Sex-roles of mothers and children in divorced, single parent families. *Journal of Divorce*, 8, 34–46.
- Rossi, A. (1979). Transition to parenthood. In P. Rose (Ed.), *Socialization and the life cycle* (pp. 132–145). New York: St. Martin's Press.
- Rubin, J., Provenzano, F. J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry*, 44, 512–519.
- Rubin, L. (1983). *Intimate stranger: Men and women together*. New York: Harper & Row.
- Silverstein, L. B. (1996). Fathering is a feminist issue. *Psychology of Women Quarterly*, 20, 3–37.
- Simon, R. W. (1992). Parental role strain, salience of parental identity and differences in psychological distress. *Journal of Health and Social Behavior*, 33, 25–35.
- Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The personal attributes questionnaire: A measure of sex-role

- stereotypes and masculinity/femininity. *JSAS Catalog of Selected Documents in Psychology*, 4, 43.
- Stockard, J., & Johnson, M. M. (1992). *Sex and gender in society*. Englewood Cliffs, NJ: Prentice Hall.
- Tong, R. (1989). *Feminist thought: A comprehensive introduction*. Boulder, CO: Westview Press.
- West, C., & Zimmerman, D. H. (1987). Going gender. *Gender and Society*, 1, 125–151.



# Gender Role Development

Joan C. Chrisler

Connecticut College, New London, Connecticut, USA

1. Toward a Modern Individualist Society
  2. Learning to Conform to One's Gender
  3. Acquisition of Gender Identity
  4. Active Participation in Gender Socialization
  5. Shaping Future Women and Men
- Further Reading

This article describes psychological, social, and cultural factors that lead to the development of gender roles. It provides advice for parents who wish to encourage their children to be more individualist and less gender stereotyped in their behavior.

## 1. TOWARD A MODERN INDIVIDUALIST SOCIETY

### GLOSSARY

**gender** A term used to classify individuals based on psychological and behavioral characteristics, social activities, and cultural expectations that are often based on beliefs about people's sex; the terms girl/woman/feminine and boy/man/masculine are gendered classifications.

**gender identity** A term used to describe an individual's own belief about his or her gender classification (e.g., "I am a girl," "I am a boy"); by 3 years of age, most children's gender identity is fixed, and in most cases, it matches their biological sex.

**gender role** Social norms for the behavior of girls/women and boys/men; these are considered to be meta-roles in that they also influence expectations of people's behavior in subroles.

**sex** A term used to classify individuals based on biological characteristics related to reproductive capacity (e.g., chromosomes, hormone levels, gonads, genitalia); most cultures maintain a binary classification of individuals as either female or male, although some have long recognized the presence of intersex individuals by providing three or more categories of sex.

Parents who describe themselves as modern, liberal, feminist, and/or nontraditional often say that they want to raise children who are free from gender role constraints and, thus, are able to develop into full human individuals who can be "themselves." These parents set that goal for various reasons, including some (or all) of the following: (a) the belief that the flexibility of androgyny is good for people's mental health; (b) the conviction that women and men will never reach political and economic equality so long as traditional gender roles maintain the expectation of separate spheres (i.e., that women belong in the private sphere and should exercise power in the home, whereas men belong in the public sphere and should exercise power in business and government); (c) the belief that women and men are more alike than they are different and that, therefore, similarities (rather than differences) should be emphasized; and (d) the fact that technology (e.g., contraceptives, power tools, computers) has changed modern life to such an extent that the specialized talents (e.g., nurturance, physical strength) that gender

roles describe are no longer essential for success in modern life and that, instead, both women and men need combinations of feminine and masculine traits to succeed. These parents tell their children that they can be or do anything they choose, provide them with both traditional and nontraditional toys and games, and try to model androgynous behavior themselves, for example, by deliberately showing their children that both women and men can be both competent and kind.

How successful are these parents in reaching their goal? The results are mixed, and parents find that it is especially difficult to encourage androgyny in boys and in both girls and boys when they are young. The parents become frustrated and confused, and sometimes they abandon their goal. They ask psychologists questions such as the following: "Why are my sons macho and my daughters girly despite my efforts to provide them with all sorts of opportunities? What am I doing wrong?" The answer is that the parents might not be doing anything wrong; in fact, they might be doing everything right. Powerful messages from peers and popular culture can override parents' best efforts.

## 2. LEARNING TO CONFORM TO ONE'S GENDER

Gender socialization (i.e., the process of learning one's gender role) begins as soon as babies are born. The hospital staff members announce, "It's a boy!" or "It's a girl!" As soon as the newborn is cleaned up and weighed, it is wrapped in either a pink or a blue blanket. Relatives and friends of the new parents rush to the store to buy congratulatory cards, and they find themselves confronted with an array of blue or pink cards containing gendered messages about how active a baby son is and how sweet a baby daughter is. One must search long and carefully to find a unique nongendered card, which might be yellow and green and discuss how much happiness a new baby can bring. Before babies can even begin to understand spoken language, adults and older children greet them with gendered exclamations such as the following. "What a big boy!" "How's my little girl?" "He has such a strong grip!" "She has beautiful eyes!" Adults handle baby girls more gently and play with baby boys more vigorously. Some of these actions are done deliberately because people believe that it is proper to treat boys and girls differently. Other actions are done without thinking

about their repercussions. The phrases "big boy" and "sweet girl" just seem to go together because people have heard those phrases uttered so often.

## 3. ACQUISITION OF GENDER IDENTITY

Children acquire their gender identity by 3 years of age, and they begin to exhibit gender-typed preferences even earlier. In 2002, Poulin-Dubois and colleagues demonstrated that 2-year-olds spent more time looking at photographs of people behaving in gender role-inconsistent ways (e.g., a man putting on makeup) than they spent looking at photographs of people behaving in gender role-consistent ways (e.g., a woman ironing). This suggests that even at that young age, the children found the gender role-inconsistent photographs to be surprising or puzzling. Social learning theory has often been used to explain the development of gender roles and gender-typed preferences. The theory posits that children learn about gender by watching how others behave (i.e., by observational learning or modeling), that children are rewarded by others when they behave according to gendered expectations, and that children are punished by others when they do not (i.e., reinforcement contingencies are applied).

Although parents are the most powerful models that young children observe, they are certainly not children's only important models. Grandparents and other relatives, teachers, coaches, neighbors, religious leaders, television characters, and other children also are influential. In addition to modeling gendered behavior, all of these (with the exception of the fictional characters) are also able to reward children for exhibiting gender role-consistent behavior and to punish them for exhibiting gender role-inconsistent behavior. Rewards take forms such as smiles, nods, hugs, kisses, praise, and gifts. Punishments take forms such as frowns, head shakes, withdrawal of affection, verbal corrections, and even spankings. For example, "Susie" may notice that her older sister has a doll collection, whereas her older brother does not. Susie's grandmother may give her a doll for her birthday and exclaim, "I'm so glad that you like it, sweetheart!" In preschool, Susie may notice the teacher frown when she joins the boys in playing with the transportation toys, and her girlfriends may offer a behavioral correction by calling out, "Susie! Come play dolls with us."

### 3.1. Media Influences

The media are very strong influences on children's behavior, and they teach lessons of all kinds about the world and people's roles in it. Researchers have found that 3- to 5-year-old North American children watch an average of 15 to 20 hours of television per week. By 7 years of age, this drops to an average of 15 hours or less, mainly because the children are spending more hours per week in school. The youngest children may watch primarily educational programming, such as *Sesame Street* and reruns of *Mister Rogers' Neighborhood*, which tend to emphasize the uniqueness of individual children. However, older children watch primarily cartoons, action/adventure shows, sports, and situation comedies, all of which portray gender role stereotypes. On average, boys watch more television than do girls. Children who watch television learn many things about women/girls and men/boys. They learn that boys are noisy and active, that girls are quiet and passive, that women are victims who need to be rescued by men (heroes) from other men (villains), that boys and girls play with different toys and games, that women and men pursue different hobbies and careers, that girls should be pretty, and that boys should be strong. Children's books and movies contain similar messages, and these messages are repeated frequently as children reread and rewatch their favorite books and videos.

### 3.2. Division Along Gender Lines

Children's worlds are divided along gender lines. Toy stores are clearly demarcated by gender, with "girls' toys" on one side and "boys' toys" on the other. An adult who enters a toy store and asks a clerk for help in choosing a gift for a 5-year-old will immediately be asked, "Are you shopping for a boy or a girl?" Children who enter the store generally know exactly where to go because they have seen girls and boys modeling toy use on television commercials, and they head for the area of the store that features the "appropriate" toys. If there is any doubt about which aisles are for them, the colors of the toys, the packaging, and sometimes even the walls and shelves of the store will serve as clues, with bold colors (especially blue and red) being for boys and pastel colors (especially pink and purple) being for girls. Children's clothing is similarly color coded and is often decorated with gender-typed emblems, for example, a football or firetruck for boys and a flower or rainbow for girls. Girls' bibs, socks, and hats are often

lacey, and stores sell barrettes and headbands for use even with newborn girls who have very little hair. When children start to attend day care, preschool, or other organized activities, they will find themselves addressed as "boys and girls" far more often than as "children," a pattern that will continue through the elementary school years. Although the phrase may be more habitual than deliberate on the part of the adults who use it, it communicates powerfully to children that people come in two "types."

## 4. ACTIVE PARTICIPATION IN GENDER SOCIALIZATION

Gender role socialization is not a passive process. Once children realize that people come in two types and figure out which type they are, they tend to categorize by gender everything they possibly can. Cognitive developmental theory posits that children want to behave in a gender role-consistent manner; therefore, they begin at an early age to gather information about what boys and girls are supposed to be and do. The modeling and reinforcement contingencies described previously are sources of information for children to categorize. However, external rewards and punishments might not be the only forces that shape gendered behavior. Children's own internal contingencies are also at work as they come to state their preferences for gendered toys and activities and even for same-sex people. Thus, children's inner desire to act like girls or boys may be responsible for the dynamics that causes their parents to complain to psychologists as in the following ways: "I prefer to dress my daughter in pants for her warmth and comfort, but she insists on wearing dresses to school." "I want my son to have a variety of toys and games, but all he wants for his birthday are trucks and guns."

Social and cognitive processes help to maintain children's gendered preferences. Children spend a lot of time in gender-segregated play groups, that is, all-boys groups or all-girls groups. This segregation is, in part, both a cause and an effect of children's preference for same-sex peers. This preference is maintained in part by children's tendency to misremember, or miscategorize, people's traits and behavior so that same-sex traits and behavior are more highly valued. For example, boys believe that boys are nicer, smarter, and funnier than girls, and girls believe that girls are nicer, smarter, and funnier than

boys. In-group/Out-group processes also help to maintain gender stereotypes and preferences for same-sex others. Any groups whose members bond together (e.g., gender groups, ethnic groups, sports teams, artificially created groups in a social psychology experiment) look for ways in which to distinguish themselves (the in-group) from others (the out-group). They exaggerate the differences between the in-group and the out-group, form a strong preference for in-group members over out-group members, and believe that in-group members have more variability in traits and behaviors than do out-group members, who are all thought to be alike. Furthermore, once traits and behaviors are categorized by gender, children tend to notice and remember instances that are consistent with gender categories, whereas they might not notice or might forget instances that are inconsistent. These processes may help to explain the logic of a boy who values soccer, thinks that it is a boys' game, and believes that girls do not like soccer even though he has seen some girls play it and knows that not all boys like it.

Gender segregation during childhood also occurs because children are assigned gender-stereotyped tasks (e.g., "Boys, go help your father with the yardwork," "Girls, I want you to help mother with the holiday baking") and because children engage in gendered activities. For example, children learn from older children, as well as from television and books, that girls take ballet lessons and boys play Little League baseball. Of course, one can find boys dancing ballet and girls playing baseball, but they are few in number. Children who choose gender-typed activities will generally find themselves in segregated groups. These activities provide the opportunity to make more same-sex friends and strengthen children's beliefs that it is normal for girls and boys to spend a lot of their time apart. Maccoby's research in 1998 showed that spending so much time with same-sex peers leads boys and girls to develop different play styles. Boys play in larger groups, and their games are often described as rough and competitive. Girls play in smaller, more cooperative groups, and their games are often based on fantasies about becoming attractive (even glamorous) grown-ups. Once these play styles become well established, it is even more unusual for a child to join a cross-sex play group. Boys are more likely than girls to enforce gender-segregated groups; they often actively exclude girls from their games and tease other boys who voluntarily play with girls.

So, what's a parent to do? Is it impossible to raise androgynous children? Should parents give up and accept the status quo? Parents should not give up,

but they probably need to be more realistic about what can be achieved, especially at early ages. Until greater social, cultural, and political changes are achieved, parents will face an uphill battle because children will receive dozens of stereotypical messages for every stereotype-inconsistent message parents may provide. However, once children are old enough to have achieved a high level of cognitive development, parent may have greater success.

## 5. SHAPING FUTURE WOMEN AND MEN

This section contains some advice for parents who wish to raise androgynous children as well as some reasons for these parents to remain hopeful. First, researchers have shown that children show more interest in unfamiliar objects and activities when they are told that the objects and activities are preferred by same-sex people. For example, if parents want their son to take violin lessons, they should show him a picture of Pincus Zuckerman playing his violin and say something about the many famous men who play the violin so expertly. If parents want their daughter to join a soccer team, they should show her a picture of Mia Hamm playing soccer and tell her about the number of girls who are playing in the local youth soccer league. If the first message children receive about something novel is that it is gender appropriate, they may maintain interest in it long enough to find out whether they like it. If they like it, it may become part of the in-group variability discussed previously. A later comment from a peer that "violins are for girls" may be countered by "No, they're not! Lots of boys like to play the violin."

Second, recent studies indicate that both children and adults are more likely to apply gender stereotypes to other people than to themselves. Although this phenomenon is not going to aid in promoting social, cultural, and political change, it does allow individuals to exercise more freedom of choice in their own activities and preferences. Thus, parents should expose their children to a range of toys, games, and activities in hopes that some nongendered objects and activities may become preferred. If children ask only for gendered toys for their birthdays, family members can agree to fulfill some of the requests and to provide some nongendered gifts as well. It would not hurt if the father or grandfather reacts enthusiastically when a boy unwraps a nongendered gift by exclaiming,

“Wow! Candyland [or finger paints]—I loved that when I was a little boy.” Researchers have found that boys are less likely to play with a feminine or gender-neutral toy if they think that their fathers would disapprove.

Third, parents should continue to model gender-inconsistent behavior as often as they can. Parents are powerful role models for their children, and children do pay attention to what their parents do, although sociocultural pressures may cause them, at least at younger ages, to deny that gender inconsistency occurs or that it is good. Some studies have shown that children from single-parent homes, where one parent must be both nurturant and agentic, display less gender-stereotypic behavior and fewer gendered preferences than do children from two-parent homes. Also, children who grow up in two-parent homes where the parents are nontraditional in their own gender roles display less gender-stereotypic behavior than do children whose parents are traditional in their role enactment. Furthermore, researchers have found that children in nontraditional preschools spend half as much time in gender-segregated groups as do children in traditional preschools. Parents can also enroll their children in day care centers and preschools where the teachers value and reward nontraditional gender roles, and parents can try to limit the amount of exposure young children have to books and television shows that reinforce traditional gender roles.

Fourth, parents can encourage children to play with gender-neutral toys and to engage in activities that both girls and boys enjoy. Gender-stereotypic thinking is, at least to some extent, context dependent. The more time people spend in gendered activities, the more likely they may be to maintain rigid gender categories. A focus on what all children have in common may help to reduce in-group/out-group divisions. Research on in-group/out-group phenomena suggests that breaking up the existing groups or challenging them to work together to attain a goal will also reduce divisions. Therefore, instead of dividing children in a classroom or at a party into a boys' team and a girls' team for a spelling bee or a game of tag, gender-integrated teams could be formed and given color or animal names (e.g., the reds vs the greens, the dogs vs the horses).

Fifth, parents should be careful not to send contradictory messages about gender expectations, for example, by rewarding traditional behavior one minute and punishing it the next. Parents must remember that they themselves have been exposed to millions of gender-stereotypic messages during their lives; thus, they need

to take care to pass on as few as possible. For example, people often interpret the same behavior differently depending on whether it is demonstrated by boys or girls. A child who cries with frustration may be thought to be angry if a boy or upset if a girl, and a parent who applies one or the other of these terms to a child's behavior is teaching a powerful, if unintended, lesson about emotionality. Many parents find it more difficult to be accepting of feminine behavior in their sons than to be accepting of masculine behavior in their daughters because “tomboyism” is not considered by most people to be a problem, whereas “sissy” behavior usually is considered problematic. Parents' concern that boys who play “dress up” will become gay men, or that boys who do not like sports will be unpopular with other boys, can override the encouragement that they give their children to try gender-inconsistent activities.

Finally, as children reach middle childhood (ages 9–12 years), parents can begin to contradict children's dogmatic gendered statements and can actively teach critical thinking skills to counter gender role stereotypes such as in the following ways. “What makes you think that boys shouldn't take violin lessons?” “Oh, look at the skinny girls in this magazine ad. I don't think they look healthy, do you?” It is appropriate for parents to share their values with children who are old enough to understand. This is the time for parents to say, “We want you to dream your own dreams, to follow your heart. We believe that you can succeed at anything you want to try. Whatever you decide to do, we will support you with pride.” Children at this stage often become more flexible in their beliefs about, and attitudes toward, gender roles. With encouragement, they may blossom into the androgynous individuals that their parents always hoped they would be.

### See Also the Following Articles

Gender and Culture ■ Homosexuality ■ Risk-Taking and Sexual Behavior ■ Sexual Behavior ■ Sexual Behavior and Culture

### Further Reading

Furnham, A., & Mak, T. (1999). Sex role stereotyping in television commercials: A review and comparison of 14 studies done on 5 continents over 25 years. *Sex Roles, 41*, 413–437.  
Houston, A. C., & Wright, J. C. (1996). Television and the socialization of young children. In T. M. MacBeth (Ed.),

- Tuning in to young viewers: Social science perspectives on television* (pp. 37–60). Thousand Oaks, CA: Sage.
- Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.
- Murnen, S. K., Smolak, L., Mills, J. A., & Good, L. (2003). Thin, sexy women and strong, muscular men: Grade-school children's responses to objectified images of women and men. *Sex Roles, 49*, 427–437.
- Poulin-Dubois, D., Serbin, L. A., Eichstedt, J. A., Sen, M. G., & Beissel, C. F. (2002). Men don't put on make-up: Toddlers' knowledge of the gender stereotyping of household activities. *Social Development, 11*, 166–181.
- Powlishta, K. K. (1995a). Gender bias in children's perceptions of personality traits. *Sex Roles, 32*, 17–28.
- Powlishta, K. K. (1995b). Intergroup processes in childhood: Social categorization and sex role development. *Developmental Psychology, 31*, 781–788.
- Powlishta, K. K., Sen, M. G., Serbin, L. A., Poulin-Dubois, D., & Eichstedt, J. A. (2001). From infancy to middle childhood: The role of cognitive and social factors in becoming gendered. In R. K. Unger (Ed.), *Handbook of the psychology of women and gender* (pp. 116–132). New York: John Wiley.
- Raag, T., & Rackliff, C. L. (1998). Preschoolers' awareness of social expectations of gender: Relationships to toy choices. *Sex Roles, 38*, 685–700.
- Turner-Bowker, D. (1996). Gender stereotyped descriptors in children's books: Does "curious Jane" exist in the literature? *Sex Roles, 35*, 461–488.



# Gerontology, Overview

Peter A. Lichtenberg

Wayne State University, Detroit, Michigan, USA

1. The Merging of Applied and Clinical Geropsychology
2. The Increasing Cultural Heterogeneity of the Aging Population
3. Knowledge Relating to Mental and Behavioral Health Disorders in Older Adults
4. Increasing Use of Applied Geropsychology in Major Life Transitions
5. Principles to Guide Future Activities in Applied Geropsychology  
Further Reading

Over the past 15 to 20 years, applied geropsychology has undergone vast and fundamental changes. This article provides an overview of the excellent articles on the topic of applied geropsychology that are included in this encyclopedia, the four major directions in which the field has grown, and the three primary principles that will guide future growth and development. The four major changes that are touched upon by multiple articles in this section are (1) the emergence of applied and clinical geropsychology tools to prevent and treat health problems, (2) the increasing cultural heterogeneity of the aging population, (3) the explosion of knowledge relative to mental and behavioral health disorders in older adults, and (4) the increasing use of applied geropsychology in understanding and assisting with major life transitions.

## 1. THE MERGING OF APPLIED AND CLINICAL GEROPSYCHOLOGY

In the recent past, applied and clinical geropsychology did not intersect. Applied geropsychologists examined issues relevant to healthy aging and the healthy aged, and the basic process of aging. Clinical geropsychologists examined issues relevant to those older adults needing intervention for aspects of pathological aging, e.g., dementia or depression. The National Institutes of Health (NIH), and in particular the National Institute on Aging (NIA), envisioned the coming together of these seemingly disparate fields of inquiry. By focusing its funding on applications of cognitive aging and other basic knowledge for healthy living, the NIA stimulated the basic understanding of aging with applications of primary prevention and health promotion. At the same time, clinical geropsychology and related disciplines extended their knowledge into the earliest manifestations of pathologies related to aging. For example, the focus on mild cognitive impairment and its relation to dementia has extended early detection from the clinical arena to recognizing that cognitive changes in otherwise healthy aging may be early manifestations of Alzheimer's disease.

The confluence of applied and clinical geropsychology is illustrated in four articles that are included in this section of the encyclopedia. Gould examines aging, cognition, and medication adherence, and how basic cognitive processes of aging are used in designing intervention tools. Ball, Wadley, Vance, and Edwards examine how cognitive skills can be maintained and enhanced. Their

review of the links between cognitive skills and functional skills and the promising results of intervention trials forecasts a continued growth in the area of cognitive training. Laffeche and Verfaellie analyze the syndrome of amnesia and the problems of clinical patients suffering from one of many disorders that cause amnesia. Cowart's article provides an excellent description of dementia as a clinical syndrome and the use of conceptual models to guide prevention, early detection, and treatment methods. All four articles address the usage of cognitive theories to guide interventions, and demonstrate the ever-increasing convergence between healthy and pathological aging, and the ability to prevent and detect pathology and intervene in clinical syndromes.

## **2. THE INCREASING CULTURAL HETEROGENEITY OF THE AGING POPULATION**

Heterogeneity in aging has traditionally been represented in approaches to healthy versus pathological aging. More recently, it has another, broader meaning in the aging field in terms of the tremendous differences in cultural and ethnic backgrounds among the nation's older adults. Changes in older adult immigration patterns have contributed to a much broader diversity of ethnicity and culture within the elderly in the United States. These dramatic changes are directly examined in the articles by Antonucci and Akiyama and by Brown and Jackson. Antonucci and Akiyama note that cultural experience refers to the interaction between immigration and acceptance. They also emphasize that a life span perspective helps in understanding how individuals become more different from each other as they develop over time, and that the differences among individuals from unique cultural backgrounds may be heightened as they grow older. Brown and Jackson use a life course framework to examine socioeconomic and health disparities between minority groups and European Americans who reside in the United States, and vividly illustrate the heterogeneity in aging between and within minority populations.

## **3. KNOWLEDGE RELATING TO MENTAL AND BEHAVIORAL HEALTH DISORDERS IN OLDER ADULTS**

Applied geropsychology has witnessed an incredible explosion of knowledge in mental and behavioral

health relative to epidemiology, measurement tools, treatment and relapse prevention, and the importance of family education. The articles in this section relating to these topics demonstrate the increased sophistication regarding chronic health problems and their assessment and treatment. Related articles on psychotherapy and personality illustrate the growth of knowledge and its application with older adults.

Knight and David's article on psychotherapy presents a conceptual framework that incorporates the similarities of psychotherapy with older and younger adults, while illustrating the unique features of psychotherapy with older adults. Alea, Diehl, and Bluck examine continuity versus the changing nature of personality across older adulthood and the role of emotion in later life. Both frameworks guide their interpretation and thinking about major issues, and provide information that can be used in working with older adults in community and clinical settings.

Depression and anxiety are the two most common forms of mental health disturbance in older age. Late-life depression, which has been the focus of much research in older adults, is examined in the article by Edelstein and Shreve-Neiger. These authors review conceptual models that identify unique aspects of depression in older age and mental and physical health comorbidities. They also examine instruments for use in depression assessment and the interventions used by geropsychologists. Anxiety disorders in older age is examined by Carmin, Mohlman, and Buckley.

Although more prevalent than depression, anxiety lags behind as an area for investigation, possibly due to the confusing array of types of anxiety disorders, each with its own unique diagnostic criteria. Anxiety disorders include a group of syndromes under a single umbrella, such as social phobia, which is clearly distinguished from generalized anxiety disorder. Despite being grouped together, anxiety disorders represent a wide variety of symptom constellations. In contrast, although depression is known to have many different possible etiologies, this disorder is nevertheless defined by a single constellation of symptoms. Consequently, knowledge about anxiety in older adults has lagged behind depression. Carmin, Mohlman, and Buckley examine the contributions that geropsychology is making to the assessment and treatment of anxiety.

Tools from geropsychology have strongly impacted populations such as those with alcohol problems and advanced dementia. People in both groups can easily find themselves isolated and neglected. The unique aspects of alcohol misuse in older adults have been

more clearly identified in recent years. New and broader treatment approaches have also been introduced and used successfully in a variety of settings. Cohen-Mansfield's article on assessment and interventions for behavioral disorders in dementia documents the effective multi-modal approaches developed in geropsychology.

Rybarczyk's article on behavioral medicine issues in late life is a likely precursor to a huge growth of interest in this area of applied geropsychology. His review of pain and sleep disorders and their treatment in older adults reflects the increasing expertise in geriatric rehabilitation that accompanies chronic diseases. Rybarczyk also examines the concept of frailty, which is defined differently by individual researchers. Most work on frailty examines the experience of illness and the adequacy of functioning in the oldest-old. Given the growing numbers of oldest-old, this area will, no doubt, be greatly enhanced by geropsychology in the coming years.

#### **4. INCREASING USE OF APPLIED GEROPSYCHOLOGY IN MAJOR LIFE TRANSITIONS**

Applied geropsychology is becoming ever more important in contributing to greater understanding and more effective assistance with major life transitions. This is an exciting area in which society will increasingly look to geropsychology for advances and contributions. Each of the four distinct areas of applied geropsychology that are covered in this encyclopedia reflects increased activity and growing awareness. Individual articles focus on the interrelated topics of caregiving and competency in older persons and end-of-life abuse. Caregiving involves the stressful experience of caring for an older adult with chronic dementia and/or disease, whereas competency focuses on the advanced stages of illness in which an older individual may need help with basic decisions, usually from a caregiver.

Applied geropsychology has developed a much broader understanding of caregiving in the past two decades, including the normative experience of stress, as well as the impact of stress felt by caregivers. Interventions for caregivers continue to proliferate and are badly needed. Burton, Kwak, and Haley examine the ever-growing phenomenon of caregiving for an older relative, which challenges applied geropsychology to develop multiple methods for intervention. Moyer discusses the growing knowledge about capacity

and competence and reports on research findings and on ethical issues connected to competency. Given the growing knowledge in assessment and measurement, applied geropsychology offers considerable expertise for dealing with issues relating to competency and capacity disputes. The legal system is not well versed in age-related diseases and how they impact specific and broader capacities.

Elder abuse and end-of-life issues in applied geropsychology are two areas of late life in which exploration, to date, has lagged. Elder abuse is an important aspect of caregiving; it is increasingly prevalent and needs greater vigilance. Rodriguez reviews research on elder abuse and examines relevant ethical issues, noting that there are no typical descriptions of abusers or victims of abuse. He also provides clear guidelines for assessment, evaluation, and successful intervention. Allen, Phillips, and Payne review the emerging role of geropsychology in end-of-life issues regarding advanced care planning, palliative and hospice care, and other evolving end-of-life treatment approaches.

#### **5. PRINCIPLES TO GUIDE FUTURE ACTIVITIES IN APPLIED GEROPSYCHOLOGY**

The three principles listed below are relevant to all of the articles in the emerging field of applied geropsychology. These principles are also related to and can be used to anticipate future developments.

##### **5.1. Principle 1: Cohort Considerations**

Researchers and clinicians must continue to consider the ever-expanding interactions between the physical effects of aging and the psychological effects of cohorts. There are several major developments that are prominent for the first time in research on cohorts. The current cohort of older adults witnessed the great promise of medicine, for example, the discovery of vaccines to prevent polio and other devastating and contagious diseases. These discoveries have contributed to optimism in regard to conquering disease, and to healthy living throughout the lifespan.

Compared to previous cohorts, older adults today are more knowledgeable about the effects of aging and are living longer and healthier lives. Consequently, both clinicians and researchers must be prepared to adapt

when working with more knowledgeable and diverse populations. The growth of the older population, both today and in the future, will impact our society on many levels. Older adults will become a more visible group and will be more often targeted by government, media, and businesses.

In addition to being more numerous, current and near-future cohorts are also characterized by greater diversity in their cultural practices and ethnic backgrounds. The effects of aging are also more diverse than ever before due to these differences. Because there are numerous differences between the current elder cohort and previous and later cohorts, the importance of considering cohort effects, along with the effects of aging, cannot be overstated.

### 5.2. Principle 2: Conceptual Frameworks

In the previous decade, applied and clinical research developed in parallel and in a relatively noninteractive fashion. There were few applications and limited theoretical understanding (e.g., it was only after 40<sup>+</sup> years of research that Schaie's longitudinal data were applied to clinical conditions). These new theoretical frameworks are reflected in the contributions of applied geropsychology to this encyclopedia, and are being applied to healthy and pathological aging. Behavioral theory is now used in assisting with treatment adherence among well elders and for decreasing agitation in demented elders. Hopefully, late-life depression theories, such as late-onset and vascular depression, will lead to early detection methods and better depression treatment outcomes.

In the past two decades, clinicians have brought meaningful conceptual frameworks to the study of aging and older adults. Applied geropsychology advances most when theoretical and conceptual frameworks are used in the study of age-related syndromes, and in the application of assessment tools and treatment techniques. These enhanced theoretical frameworks

provide a more logical rationale for evaluating and building upon the results of empirical research results.

### 5.3. Principle 3: Comorbidity

People are living longer lives, which makes them vulnerable to multiple chronic diseases (comorbidities). Among the ever-increasing number of persons surviving one or multiple strokes, there is a concurrent increased risk because their systems are more vulnerable to developing depression, dementia, and delirium. Recent epidemiological evidence suggests that chronic conditions during mid-life, such as high cholesterol, hypertension, and diabetes, place people at much higher risks for developing dementia in later years. Thus, issues of comorbidity are often central to applied geropsychology and need to receive increasing attention.

### See Also the Following Articles

Age-Related Issues among Minority Populations ■ Aging and Competency ■ Aging, Cognition, and Medication Adherence ■ Anxiety Disorders in Late Life ■ Behavioral Medicine Issues in Late Life—Pain, Sleep, etc. ■ Cognitive and Behavioral Interventions for Persons with Dementia ■ Cognitive Skills: Training, Maintenance, and Daily Usage ■ Dementia in Older Adults ■ Depression in Late Life ■ Elder Caregiving ■ End of Life Issues ■ Personality and Emotion in Late Life ■ Psychotherapy in Older Adults

### Further Reading

Curry, L., & Jackson, J. (2003). *The science of inclusion*. Washington, D.C.: Gerontological Society of America.

Kane, R. L., & Kane, R. A. (2000). *Assessing older persons*. Oxford, UK: Oxford University Press.

Lichtenberg, P. A. (Ed.) (1999). *Handbook of assessment in clinical gerontology*. New York, NY: Wiley.

Lichtenberg, P. A., Murman, D. L., & Mellow, A. (2003). *Interdisciplinary handbook of dementia: Psychological, psychiatric and neurological perspectives*. New York, NY: John Wiley & Sons.



# Gestalt Therapy

Stephen G. Zahm and Eva K. Gold

Pacific University School of Professional Psychology, Forest Grove, Oregon, USA,  
and Gestalt Therapy Training Center–Northwest, Portland, Oregon, USA

1. Description of Treatment
  2. Theoretical Bases
  3. Applications and Exclusions
  4. Empirical Studies
  5. Case Illustration
  6. Conclusion
- Further Reading

*foreground* As distinguished from background, that aspect of experience that is the focus of attention at any given point.  
*gestalt* A German word that has no direct translation in English. “Configuration,” “structure,” and “whole” are used as translations, but none of these capture the complete meaning. “Figure” and “gestalt” are often used interchangeably.  
*organismic self-regulation* The process whereby homeostasis is maintained via the figure formation/destruction process.

## GLOSSARY

*background* As distinguished from figure or foreground, that which is not uppermost in awareness; provides the context and is the source from which figures emerge and become prominent in awareness.

*closure* The result of the natural movement toward the completion of perceptual or experiential units.

*contact* The experience at the self/environment boundary that leads to assimilation and growth. Good quality contact involves awareness and excitement.

*contact boundary* The dynamic relationship at the meeting point of self/other or self/environment; where experience occurs, and the focus of therapeutic intervention.

*creative adjustment* A person’s best adaptation to dealing with feelings and needs in response to environmental inadequacy in meeting the needs or competing environmental requirements.

*figure* What stands out from the background as a focus of interest and attention; generally related to a need or desire.

*figure formation/destruction process* The process whereby a figure of interest emerges, becomes a focus of attention, and is acted upon, resulting in the dissolving of that figure.

Gestalt therapy has roots in psychoanalysis, Gestalt psychology, Reichian character analysis, and the work of the early phenomenologists. It was developed in the 1940s primarily by Frederick Perls, Laura Perls, and Paul Goodman. Gestalt therapy is humanistic, holistic, and experiential. It is a process-oriented and relational approach based on a theory of healthy functioning. The focus in therapy is on disruption of the natural process of self-regulation. The goal of treatment is the restoration of awareness, which allows for increased choice and flexibility in all aspects of living.

## 1. DESCRIPTION OF TREATMENT

### 1.1. Philosophical Foundations

Gestalt therapy method is guided by three philosophical foundations. The first of these is field theory, a concept that derives from physics. The “field” is a dynamic interrelated system, each part of which influences every

other part. Nothing exists in isolation. Gestalt therapy is a field theoretical and process-oriented approach, which means the therapist attends to the total field, including content and subject matter, as well as the here-and-now process occurring in the moment. This involves things such as the patient's tone of voice, style of communication and interaction, facial expression, physical gestures, posture, breathing, sensation, and affect. The therapeutic field also includes the patient's history and current situation, the therapeutic relationship, feelings experienced by the patient and the therapist, and their mutual impact on one another.

The second philosophical foundation is phenomenology. The Gestalt therapist attempts to bracket off preconceived biases, beliefs, theories, and interpretations and attend as much as possible to what is actually understandable through the senses, with the focus on the patient's subjective experience. The therapeutic stance is one of meeting the patient where he is, and the therapist describing what she observes, rather than explaining or interpreting. The patient is encouraged to do the same. The therapist also brings her own feelings, sensations, and awareness into the interaction with the patient when useful for the treatment.

The third philosophical foundation is dialogue. A dialogic stance requires that the therapist be "present," confirm the patient, and be available for an I-Thou way of relating. Presence is the therapist bringing all of himself to the encounter in the here and now. Confirmation involves seeing and accepting patients for all they are and all they are capable of being. The I-Thou mode of relating has the qualities of immediacy, directness, presence, and mutuality. These qualities create a therapeutic relationship that is not hierarchical and a meeting between patient and therapist that is in and of itself healing.

### **1.2. Goal of Treatment/Theory of Change**

In Gestalt therapy, the goal is increased awareness. Although the patient presents specific symptoms, such as anxiety, depression, stress-related physical complaints, relationship difficulties, or character/personality issues, the Gestalt therapist does not focus only on symptoms or behavior change. Rather, Gestalt therapy provides a holistic perspective in which symptoms and problematic behaviors are seen as the person's best attempt to deal with conflicting needs or needs conflicting with environmental requirements. Psychological

problems, symptoms, and difficulties in interpersonal relationships result from what were originally creative adjustments that have become rigid, reflexive, and unaware, restricting the experience of self and possibilities. By focusing on moment-to-moment process in the therapy session, the Gestalt therapist works with how the person creates and maintains her particular experience of self and other and how this experience impacts the current situation.

Change then occurs spontaneously as a result of this increased awareness. Gestalt therapy's change theory is paradoxical in the sense that change occurs not as a result of attempting to change (one part of the self trying to control and dominate another part) but as a result of the patient attending to, investing more fully in, and living more completely the actual current experience—what is. Doing this enables new responses to emerge that better fit current needs. Increased awareness leads to choice and opens up new options for behavior and interaction.

As aspects of the personality that have been disowned are reidentified with, an expanded sense of self is also restored. For example, the patient increases her capacity to experience and express grief, anger, or tenderness, which expands her ability to live authentically and to connect more with other people. Gestalt therapy does not focus directly on problem solving but rather on what limits the patient's capacity for creative solutions. As the saying goes, "Give a person a fish and he eats for a day; teach a person to fish and he eats for a lifetime." A Gestalt therapist is interested primarily in the latter.

### **1.3. Therapy Method**

By attending to observable behaviors that indicate blocked awareness and interruptions to acting on needs, desires, or interests, Gestalt therapy captures the essence of a person's existential position in the world, which impacts all aspects of self-experience and relationships. For example, the patient speaks softly, apologizes frequently, becomes slightly tearful but cannot cry, changes the subject, or shuts down if strong feelings threaten to emerge. Can the patient become aware of how she clamps down on herself? What makes this shutting down necessary? How might holding back or not allowing herself to take up space be a theme in this patient's life? How many areas of her experience and relationships are impacted by it? How is it related to her presenting symptoms of anxiety or depression or to the dissatisfactions in her life?

The therapy process enables increased awareness of these blocks and how they are maintained and also examines why they were initially developed. For example, if the experience or expression of sadness has been criticized, punished, or even ignored by the parent, the child learns to cut off the feeling, its expression to others, or both. If a child's needs have been met with resentment, criticism, or disgust, the child's creative adjustment is to learn to block awareness of needs or to learn to avoid showing or expressing them to others.

Through focused attention, the patient can become aware of how he blocks feelings/expression, for example, tightening muscles to hold back tears and also internal messages such as "Don't be weak" or "Men don't cry." The patient will discover the etiology—what made this necessary in his development. Symptoms then resolve as the patient reconnects with his authentic self and experience. For example, anxiety may result from a fear that certain "unacceptable" needs or feelings will surface in a particular situation. Depression may result from holding in anger or grief or from poor self-esteem from directing frustration or criticism toward the self that needs to be expressed to another. Relationship problems may result from an inability to open up and reveal feelings or difficulty needing another person.

Gestalt therapy's method is experimental and experiential. Experiments can be diagnostic as well as serving to highlight current experience and uncover new possibilities. For example, an unassuming person who does not make eye contact and qualifies most of what she says complains that people do not listen to her or take her seriously. In therapy she might be asked to experiment with lecturing the therapist on the proper way to do something that is in her area of expertise. She might then show and experience the side of herself that is a knowledgeable expert. She may begin to speak more directly and forcefully. In this process she can get more in touch with this aspect of herself as well as increase her awareness of her reluctance to show this aspect of herself to others or her anxiety in doing so. Conversely, the patient may speak to the therapist as if she is not an expert, when in fact she is. Her language might be filled with reflexive use of qualifying phrases. She might speak without conviction. In either scenario, the patient can see and understand how she contributes to others not giving her the respect she desires.

The patient also learns what made this way of being necessary. She could have been told "Children should be seen and not heard," or one of her parents might

have become angry or silently disapproving if she expressed an opinion of her own. She might have been told she was stupid or wrong. Prior to therapy, she may not have remembered these incidents, she may have dismissed them, or she may have minimized their impact. She may not even have been aware that she presented herself as timid and unsure of herself or why it was important to do so. With the new awareness, she regains the possibility of a wider range in experience and behavior.

Gestalt therapy uses experimentation to increase awareness. Meaning for the patient then comes from experienced awareness rather than being imposed by the therapist's interpretation. The therapist tries to have no goal, agenda, or desired outcome in proposing an experiment, other than to observe and work with what happens next. An experiment can also be used to highlight an aspect of the patient's way of communicating. For example, a person who qualifies much of what he says by use of words such as "maybe," "possibly," or "I guess" could experiment with either exaggerating the use of qualifying words or eliminating them entirely and adding after every sentence "and I mean that." The therapist encourages the patient to pay attention to his experience as he does an experiment and to express any reluctance in doing it.

Experiments can also enrich what a person is saying and transform it from "talking about" to a lively present encounter. For example, the patient might imagine the person she is talking about is present in the room and express her feelings directly to that person. This use of the empty chair or two-chair experiment can also be effective in highlighting an internal conflict. The patient might imagine a part of herself in the empty chair and create a dialogue between the conflicting aspects of herself, such as the part that wants to be more creative and take more risks and the part that is more logical and conservative. This type of experiment can also explore an area of impasse or stuckness, for example, suggesting the patient develop a dialogue between the part of herself that wants to leave her marriage and the part that wants to stay.

As an example of some of the previously discussed concepts, a patient who learned to stifle her feelings and needs and to nurture herself with food enters treatment depressed and with a poor self-concept, complaining that she has "tried everything" to control her eating and has been unsuccessful. Turning to food to meet emotional needs was at one time her creative adjustment; it was the best option available given the needs she had and the lack of available gratification

possibilities—in this case, a mother who was stern and unaffectionate and sent the patient to her room if she was “too emotional.”

The Gestalt therapist observes this patient’s responses in the moment and helps to facilitate her awareness of them. For example, the therapist notices that the patient speaks more quietly and holds her breath when she talks about a hurtful or frustrating experience. The therapist can either make an observation (“It looks like you’re tightening your jaw” or “I notice you holding your breath”) or ask the patient to attend to what she is currently aware of (“What is your experience right now as you tell me this?”). The patient becomes aware that she is trying to control herself. The therapist attends to her need to control herself. He also attends to what she is controlling—verbalizing feelings and showing anger or tears. Through this process the patient becomes aware that there is also a part of her that would like to express itself and feels held back. Both aspects are explored.

The working-through process involves the patient becoming aware of the importance of holding back feelings to prevent reexperiencing the rejection she felt from her mother or the anger she felt toward her mother when she had feelings or needs. She might remember a forgotten traumatic experience. She might also discover that she expects the therapist to have a similar response. This patient learned over time to identify with her “nonfeeling, nonneeding” self. Her eating behavior may be one of the only remaining vestiges of her “feeling, needing” self. In therapy, she can also experiment with giving voice to the part of her that wants to eat and reidentify with her needs and desires, instead of only wanting to suppress these needs. For example, she might realize that although part of her wants to stop overeating, part of her—out of her awareness—also wants to continue eating as a way to take care of feelings and to comfort herself. This part might say something like “Don’t take food away from me, it’s all I have.” Change becomes possible as she invests more fully in being where she is—feeling those needs she translates into a desire for food—and may take the form of learning to nurture herself in other ways or to get her needs met in relationship with others.

## 2. THEORETICAL BASES

Many central concepts of Gestalt therapy theory are based on laws of perception discovered and studied by Gestalt psychologists in the early 1900s. Gestalt therapy

applies these laws more broadly to all aspects of experience and psychological functioning. Importantly, these Gestalt psychologists discovered that we are not simply passive recipients of perceptual stimuli, but that we are active in organizing our perceptual field. Laws discovered by the Gestalt psychologists include the natural tendency to perceive a figure against a background as a way of organizing experience and a natural tendency toward completion or closure.

Gestalt therapy takes its theory of healthy functioning from a biological concept called organismic self-regulation. This describes the organism’s process of taking in from the environment what is needed (food and oxygen) and expelling into the environment that which is not required (waste products and carbon dioxide) in order to maintain balance or homeostasis. Applied to psychological functioning, this theory states that in health, people will naturally go to the environment to get emotional and psychological needs met—the need to be listened to and the need for support, sex, physical comfort, and social contact—and will discharge as needed by talking, expressing feelings, crying, touching, doing, and creating. Gestalt therapy does not view the individual as separate from the environment but rather considers the individual/environment field.

According to Gestalt therapy theory, such self-regulation occurs via the figure formation/destruction process. This process is central to any human functioning and as such is a core characteristic of people. In all functioning there is an ongoing process of the formation of a figure of interest and the eventual gratification and dissolution of this need or interest. This cycle has several stages: (1) awareness, (2) clarification of need/interest, (3) scanning self-environment, (4) action, (5) contacting, (6) assimilating, and (7) withdrawing. For example, I am on my way to my favorite restaurant for lunch and see an old friend across the street. I become aware of sensations of excitement and increased energy (stage 1). I recognize my interest in talking to my friend (stage 2). I understand that walking across the street will facilitate meeting this need (stage 3). I walk across the street (stage 4). I say hello and give my friend a hug, and we talk (stage 5). As we say good-bye, I tell my friend and acknowledge to myself what this meeting meant to me (stage 6). I walk back across the street with a feeling of satisfaction, resolve to see my old friend more often, and my focus shifts to what I want to eat for lunch (stage 7).

In healthy functioning, a person will maintain balance by intake and discharge as required for the

satisfaction of emotional and psychological needs by this continuous process of figure formation/destruction. Assuming there is no interference in this process (a disruption that causes the person to misperceive her needs, an environment that is hostile to the needs, or both), what becomes figural or foreground is based on the most pressing need at any given moment. After a need is met, this figure recedes (closure) and is replaced by whatever is next in the hierarchy of needs.

When there is a disjuncture between needs and what is available in (or required by) the environment, a person adapts. A creative adjustment is the best possible accommodation she can make at the time, given her experience, perceptions, and limitations. Over time, these creative adjustments may become obsolete, rigid, and maintained out of awareness, limiting a person's potential for satisfaction and growth. Psychotherapy involves examining and understanding how the natural self-regulatory process has become disrupted, preventing closure on units of experience.

Disruption can occur at any stage of the figure formation/destruction process and diminish a person's experience of himself and his ability for contact. A person who is experiencing disruption interferes with his own development of a strong and clear figure (stages 1 and 2) or with the ability to maintain the excitement and energy necessary for action (stages 3 and 4) or to move into satisfying contact (stage 5), which allows for assimilation (stage 6) and closure (stage 7). Either the interference with the development of a strong figure or the blocking of excitement and action can interfere with good quality contact. A Gestalt therapist is trained to observe both where in the cycle the disruption occurs and how it occurs. Often, the focus is on how the disruption dilutes or distorts interpersonal contact.

In health, a person will perceive, experience, identify with, and act on needs and desires. Healthy functioning requires the ability to readily identify "This is me" or "This is for me" versus "This is not me" or "This is not for me." When disturbances exist in perceiving, experiencing, identifying with, and acting on needs, this loss of functioning is observable by the trained therapist at the contact boundary, which refers to the dynamic relationship at the point where the person and environment meet and interact.

The five processes discussed next are the "how" of any disruption of contact. These processes can occur at any stage of the figure formation/destruction process. It is via these processes, maintained rigidly

and out of the person's awareness, that the experience of self and the ability for contact that includes satisfaction of needs and closure is diminished.

Introjecting is experiencing something in the environment as if it is part of the self. An introject is an idea that has been "swallowed whole" without the "chewing" necessary to assimilate it and make it truly a part of the self. Introjecting is particularly problematic if it is a person's primary way of dealing with the world and relationships (such as doing things for others and ignoring one's own needs out of a sense of duty or obligation) or when the introject is destructive (such as a parental introject "You're good for nothing").

Projecting is disowning a feeling, behavior, attitude, or trait of one's own and attributing it to another or the environment. For example, a person who is not able to be angry—due to an introject—may inaccurately perceive another person as angry at him.

Retroreflecting is when energy or action that could be directed toward the environment is directed at oneself. Retroreflecting is of two types. In the first type, a person directs toward herself a feeling or action meant for someone or something in the environment. For example, the person who is unable to experience anger toward another person—again due to an introject—might end up blaming herself. Any unaware suppression or shutting down is also accomplished by this type of retroreflecting. The second type of retroreflecting is doing for/to oneself what one would like another person to do for/to oneself. For example, a person might soothe herself, buy herself gifts, or hug herself instead of getting these needs met by interacting with another.

Confluence is the attempt to deny the existence of a self/environment boundary. Confluence requires a lack of discrimination and articulation of points of difference or otherness. Although there is the possibility of a "healthy" confluence during which two people feel so joined that they seem "as one" (such as at the moment of orgasm) and there is momentary dissolving of the boundary, prolonged attempts to maintain confluence prevent satisfying contact.

Deflecting is behavior that dilutes or reduces the intensity of contact. Examples are avoiding eye contact, laughing off what one says, circumlocution, and understating one's true feelings. There is disagreement among Gestalt therapy theorists as to whether this is to be considered a separate self-regulatory boundary process or whether it is the behavioral manifestation of the others—that is, a result of introjects, retroreflection, projection, or desire to maintain confluence.

### 3. APPLICATIONS AND EXCLUSIONS

Because Gestalt therapy is not an adjustment therapy and strives to be value-free, with a focus on understanding each unique individual and the individual's subjective experience as opposed to an "objective" standard for psychological health, there is no population that would be excluded from treatment. The phenomenological and field theoretical stance, whereby the therapist is striving to understand the patient's subjective experience rather than impose his own, make this orientation valuable for different cultures and minority groups. The Gestalt therapist does not claim to know what is right for the patient, nor does he desire any particular outcome: The only goal is increased awareness for the patient that leads to more choice.

Gestalt therapy can be used with all patient populations (children, adolescent, adult, geriatric, physically/mentally challenged, prison inmates, and inpatient and outpatient psychiatric patients) as well as in a variety of modalities (individual, couples, family, group, and organization). It is applicable to a wide array of problems [e.g., eating disorders, substance abuse, affective disorders, personality disorders, posttraumatic stress disorder (PTSD), adjustment disorders, loss, and grief]. It is an approach that can be used in both long- and short-term therapeutic work. Of course, not every therapist is equally effective with all patient populations, modalities, and diagnoses. The experienced and well-trained Gestalt therapist would be aware of personal limitations and areas of expertise and refer those patients who fall outside of these limits to other practitioners, just as a responsible therapist of any orientation would.

### 4. EMPIRICAL STUDIES

Research on the effectiveness of Gestalt therapy is scant and primarily confined to specific aspects of the approach. Most research has been focused on one technique that derives from Gestalt therapy method—the empty chair (or two-chair) experiment. It is important to note that use of this or any other single technique does not actually constitute a Gestalt therapy approach, without application of the three philosophical principles described in the treatment section. Although no single technique can be said to represent the essence of the Gestalt therapy approach, a technique has the advantage of being easily defined and thus accessible to the manipulations of empirical research.

The majority of research used the empty chair technique. Subjects were asked to have a dialogue with either a part of themselves or someone else relevant to the issues with which they were grappling. Most of the studies that used the empty chair technique found it to be more effective than other forms of therapy or control groups. It was also found to be more effective in reducing anger and producing positive attitudes than either intellectual analysis or emotional discharge. It was more effective than empathic reflection or focusing for creating a greater depth of experience. For resolving "unfinished business," it was more effective than an attention-placebo condition.

In one study, K. M. Clark and L. S. Greenberg found the empty chair technique to be more effective than either a cognitive decision-making task or a waiting list control group for resolving decisional conflict. During the process of using this technique, the original decision the person had wished to make often unfolded into a deeper related decision.

In summary, studies have shown this technique to result in an increase in affect, depth of experience, and the resolution of related emotional issues. Results have included a decreased sense of decisional conflict, reduced anger, increased positive attitude, lower systolic blood pressure, and the resolution of unfinished business.

Two studies found no benefit in using the empty chair technique for resolving grief issues or dealing with mild depression. In a study conducted by N. P. Field and M. J. Horowitz, the intervention consisted of placing subjects in a room alone and asking them to talk to their deceased spouses for 5 minutes through the direction of a taped recording. Nothing like this would ever be done in an actual Gestalt psychotherapy, which requires a focus on moment-to-moment process and the presence of the therapist. This technique was not found to be effective in reducing symptomatology 6 months later. Another study found the empty chair technique to be no more effective than time for decreasing mild depression. The researchers stated that the lack of positive results may have been due to the short duration of treatment, gravitation toward the mean over time, or experimenter bias.

Focused expressive therapy (FEP) was used in several research studies. FEP borrows from Gestalt therapy method and uses such procedures as directed fantasy, two-chair dialogues, and awareness exercises. It emphasizes the importance of emotional insight and the magnification of internal experiences, but it also includes a high level of authoritative guidance and

confrontation that is inconsistent with a Gestalt therapy approach. FEP was not found to be as effective as either cognitive group therapy or self-directed therapy. However, the description given for self-directed therapy was actually more similar to Gestalt therapy as it is practiced than the more authoritarian FEP method used to represent Gestalt therapy in this study.

Few studies have examined a more complete Gestalt therapy approach. Most of these yielded positive results, such as an increase in levels of self-actualization and more positive attitudes toward body image. Gestalt therapy was found to increase internal locus of control and the assumption of personal responsibility in prison inmates. A Gestalt therapy approach resulted in more behavior change than empathic reflection. Gestalt therapy was found to help Vietnam veterans with symptoms of PTSD. In couples therapy, it was effective in replacing power struggles with self-definition, limit setting, and intimacy, resulting in greater clarity, acceptance, and understanding.

## 5. CASE ILLUSTRATION

The following case example illustrates the treatment process and application of the theoretical concepts that have been presented. The therapist in this case was S. Zahm.

### 5.1. Kim

Kim, a 44-year-old successful professional woman, was referred by her physician due to recurrent periods of depression. She had been depressed off and on for much of her life. The bouts of depression had worsened recently in both duration and severity. She had not been able to tolerate antidepressant medications, and two previous experiences with therapy, both brief, had not yielded lasting improvement.

Kim presented as a smart, attractive, stylishly dressed woman. Her wit and humor were evident despite her somewhat depressed affect. Her facility with language, keen observations, and easy manner made it clear why she had achieved success professionally. During the initial session, Kim asked perceptive questions about my experience, my therapy orientation, and how I thought therapy could help her. She had been disappointed in her previous therapy experiences. When asked what she hoped to get from therapy, she answered "Not just a Band-Aid. I'd like to get to the bottom of what this depression is about."

Kim described herself as "an overeducated over-achiever" who had been successful in school and in

her profession. She worked long hours but wondered what was the point of all her hard work because she took little enjoyment from her accomplishments. Kim saw her life as lacking in meaning, with little that gave her pleasure. She said she felt "down" a lot, woke up in the morning dreading the day ahead, and said she was just "going through the motions" of social interaction.

Kim described her relationship with her husband Bill as "pleasant enough" but lacking a sense of emotional connection or passion. She reported little interest in sex and stated that their infrequent lovemaking was "more for him." She said "it's like I'm dead inside, or just numb." Kim reported that most of her friends were work acquaintances, and that she had only one "good friend" but did not really confide in her. Kim and her husband had no children. Kim and I agreed to meet for weekly individual psychotherapy sessions and to track whether she felt the sessions were useful and whether she was getting what she needed.

In describing her history, Kim reported that her parents had divorced when she was 8 years old. Her father abandoned the family, moving to another state to avoid paying child support. Kim, the oldest of three, became responsible for helping her mother with her younger siblings and doing household chores so that her mother could go back to work to support the family. Kim also became her mother's confidant and emotional support. She reported that she never really "got to be a kid" after that, and that her mother "never had time" for her. Kim's mother died approximately 8 months prior to our first meeting. She minimized her feelings about her mother's death, saying she thought she had "dealt with it."

Kim's therapy themes emerged right away in her relationship with me. In our initial sessions, her fear of showing or revealing vulnerable feelings to me stood out. For example, when Kim talked about her mother's death she would reflexively shut down. I noticed her retroreflections and deflections, including choking back tears, changing the topic, and looking away. I pointed these out in an attempt to sharpen this figure, to clarify what was most important to her at that point. I observed that although she had strong feelings about this loss, it was also important to her not to show these feelings or talk about them to me right then.

This process helped Kim focus on her reluctance to open up and share vulnerable feelings. The theme of having a variety of "softer" feelings (sadness, loss, and emptiness) and the reluctance to express them to me was something we worked with over the course of the

therapy. The essential components involved introjects such as “Be strong” and “Don’t show weakness” and the belief that no one would be interested in her feelings.

Within the first few months of therapy, Kim became more aware of these internal mandates and saw how they were connected to her relationship with her mother. Kim had learned not to reveal her feelings and emotional needs but rather to focus on taking care of her mother and siblings and deal with any of her needs by herself. Initially, it was a major step for Kim to identify that she felt a need and to also feel her difficulty and discomfort in expressing it to me.

I suggested Kim experiment with looking at me and imagining allowing herself to let down and cry with me. She imagined me having a critical and rejecting response. As she worked with her reluctance over a period of weeks, and felt supported by me in doing only what she was comfortable with, her reluctance moved into the background and her organismic desire and need to experience and express her feelings became more figural. At that point she was able to let down into her feelings of loss and grief, thus beginning the process of getting the closure she needed with regard to her mother’s death.

Afterwards, Kim was surprised to discover that this felt OK. It was still difficult for her to ask me how I now felt toward her. Her strong pull was to be satisfied with not knowing how I felt and what my reaction had been to her crying. I pointed out how not asking me about that left her with a blank space that might then be filled with her own fears. She decided to ask the question, “How do you feel about me now?” I was touched by her courage in asking the question and told her so, and I gave an honest response to her question. This was a pivotal moment in the therapy because Kim could take in and assimilate my genuine caring for her.

As we did this work and Kim opened up to deeper levels of feeling, another theme became apparent to both of us. Kim became fearful of relying on me and afraid that I would leave her if she needed me. She had a first-class radar system that detected the slightest sign of my lack of interest, lack of energy, or attention wandering. For example, if I looked at the clock she became silent and withdrew. When I told her of planned vacations, her response was to joke that she wondered if I would want to come back. When we explored this, it became clear that my going away evoked fear that I would abandon her or lose interest in her.

We worked with this theme in a number of ways. As Kim focused on feelings and sensations, she was able

to remember particular events with her father, such as his leaving the house in a flurry of loud words and emotion and yelling things such as “I never get any peace in this house!” One time, Kim tried holding on to him to keep him from leaving and he pushed her away, saying something like “Why can’t you kids leave me alone?” The meaning Kim extracted from these experiences and his ultimate abandonment of the family was that it was her fault that her father had left and that she must have been unlovable and “too much” for him.

As we worked with these awarenesses, Kim could feel a little girl part of her that “had always been there” that she tried to ignore. I asked Kim to focus on this sense of herself and to talk to me as the little girl. What emerged from this was that she felt shameful and desperately wanted to hide or disappear. She went back to her adult perspective and told me she wished that little girl would disappear or at least “grow up.” I suggested a dialogue between the adult Kim and the little girl. As the dialogue developed, the adult Kim began to feel sad and realized that rather than wanting the little girl to disappear, she wanted to hold and comfort her. She did not see her as “unlovable” or “too much” but instead felt compassion for her loss and pain and was able to tell her so. This newfound ability to experience compassion toward herself was a key in helping Kim integrate her softer “little girl” feelings with her “adult.” This occurred over a number of months as she continued to work with these aspects of herself.

When we worked directly with Kim’s father she was able to, in fantasy, tell him she was hurt and angry that he had not loved her enough to stay in her life even if he had to leave her mother. She finished her tearful dialogue with him by saying, “It wasn’t my fault you left. You failed me, not the other way around.” She also became aware that she had never been able to grieve the loss of her father and that her mother would not tolerate her tears and told her she had to be strong and that crying “wouldn’t do any good.” At that point, she was able to grieve the loss of her father with me and have more of a sense of closure.

In our sessions Kim came to see how the belief “I’m too much” affected major aspects of her emotional and interpersonal existence. This was especially evident in how she experienced her relationships with men and the type of contact she could have with them. She saw that with her husband, she made no demands, rarely got angry, and focused her attention on his needs to the exclusion of hers. Our work

included Kim experimenting with new ways of interacting with her husband, such as being more assertive with her needs. Initially, this made her anxious. As we worked with the anxiety and she realized the responses she feared were not likely ways that her husband would respond to her, she became less anxious and more excited and playful with the possibilities of being bolder about asking for things she wanted. This affected all aspects of their relationship, including their sexual relationship.

Slightly more than 1 year into the treatment, Kim reported that she “couldn’t remember the last time” she had felt depressed. She said she felt better about herself, her relationships, and the possibilities for her life than she ever had. She began to spend less time working and signed up for a beginning ballet class, which was something she had “always wanted to do.” She developed closer relationships with a few women friends, revealing more personal things and accepting support from them.

I asked Kim to reflect on where she was now and where she had been when she entered treatment. Kim reported that she felt more at peace with her parents, even though she still felt sad about her childhood. She said that it had been “life changing” to risk opening up to another person and allowing herself to need and depend on me. Kim saw the connection between her process of shutting down on her feelings—her self-criticalness, rejecting the little girl part of herself, avoiding feeling her anger and grief, and not allowing herself to need anything from others—and the depression she had experienced for so long.

Kim and I agreed that she was ready to terminate therapy and we allowed 1 month for this process. Given Kim’s issues of abandonment and loss, it was important to set an ending date and not dilute the fact that we were ending by meeting less frequently. In this way, Kim had sufficient support and a structure for dealing directly with the loss of her relationship with me. We celebrated her gains together and cried together. At our last session, we said good-bye in a complete way that Kim said she had never had the opportunity to do with anyone else.

## 5.2. Outcome

Growing up, Kim experienced a lack of awareness of and responsiveness to her emotional needs from both of her parents. Her father abandoned Kim completely, and her mother was unavailable for Kim’s needs, requiring Kim to attend to her needs instead. Kim’s

creative adjustments involved both the awareness stages and the action/expression stages of the figure formation/destruction process. She learned to cut off the experience of certain feelings and to avoid showing and expressing them. The meaning she made of her father’s leaving (“I’m unlovable” and “I’m too much”) and introjects such as “Be strong” created her retroflective process of cutting off and alienating aspects of herself, focusing on the needs of others, and not showing or expressing her own feelings and needs. This process resulted in unfinished emotional situations in which she had not been able to get closure. It resulted in the sense of emptiness, lack of meaning, and numbness that she described at the beginning of treatment as well as in unsatisfying interpersonal relationships that lacked depth. It also contributed to her career success, which was in part based on her skill in attending to the needs of others and her willingness to work hard so that she would not be rejected or abandoned.

All these issues emerged and were worked on in the context of the therapy relationship. As Kim’s awareness increased, she was able to work through her fears and reluctance and to feel the need to express and show her authentic feelings. She came to feel compassion for a part of herself she had previously cut off and wanted to be rid of, and through this experience of compassion she was able to reidentify with and integrate this part of her that included needs and softer feelings. She was able to redirect her self-criticism and see that it was not her being “too much” that caused her father to abandon her and her mother to be unable to attend to her needs. She came to feel that her little girl had deserved to be better taken care of.

In Kim’s relationship with me, she could experience my genuine caring and also see that when I was not there for her (e.g., when I went on vacation) it was not because she was unlovable or “too much.” She learned to express disappointment with my limitations, or anger at me for leaving, in a way that she had never had the opportunity to do with her parents.

The therapy carried over to and influenced every aspect of Kim’s life and interactions. Her relationship with her husband deepened, and she began to consider what she wanted to do in her life besides working. However, it was Kim’s relationship with herself that changed most dramatically. Kim no longer criticized herself or felt she was weak if she had feelings and needs. She no longer cut off and numbed a major part of who she was. She had greater range of emotional experience and expression. Kim understood that her

lifelong depression was the result of ignoring, denying, and cutting off her feelings, which led to an inability to identify and act on her needs. This resulted in a lack of closure on difficult emotional situations as well as unresolved issues of grief and loss.

## 6. CONCLUSION

The practice of Gestalt therapy is based on the philosophical foundations of field theory, phenomenology, and dialogue. Gestalt therapy is a process-oriented approach that focuses on the person/environment field. The goal in Gestalt therapy is increased awareness, and change occurs through focused attention on “what is” and not by attempting to achieve a particular goal or agenda. Because the therapist brings himself or herself to the meeting with the patient, a type of contact that is in itself healing becomes possible. Gestalt therapy is based on a theory of health, and the focus is on how the patient interrupts his or her natural self-regulation process. The method is experiential and experimental, relying on meaning emerging from experience rather than interpretation. Gestalt therapy has wide applicability with a variety of patient populations, problems, and treatment modalities.

Although limited, research indicates its usefulness for a variety of populations and issues.

## See Also the Following Articles

Ideological Orientation and Values ■ Mental Measurement and Culture ■ Stress ■ Values and Culture

## Further Reading

- Beisser, A. (1970). The paradoxical theory of change. In J. Fagan, & I. Shepherd (Eds.), *Gestalt therapy now*. Palo Alto, CA: Science and Behavior Books.
- Breshgold, E. (1989). Resistance in gestalt therapy: An historical/theoretical perspective. *Gestalt Journal*, 12.
- Jacobs, L. (1989). Dialogue in gestalt theory and therapy. *Gestalt Journal*, 12.
- Kepner, J. (1993). *Body process*. San Francisco: Jossey-Bass.
- Latner, J. (1988). *The gestalt therapy book*. Highland, NY: Gestalt Journal Press.
- Polster, E., & Polster, M. (1993). *Gestalt therapy integrated*. New York: Brunner/Mazel.
- Wheeler, G. (1991). *Gestalt reconsidered*. New York: Gardner Press.
- Yontef, G. (1993). *Awareness dialogue and process: Essays on gestalt therapy*. Highland, NY: Gestalt Journal Press.
- Zinker, J. (1997). *Creative process in gestalt therapy*. New York: Brunner/Mazel.



# Gifted Students

John F. Feldhusen

Purdue University, West Lafayette, Indiana, USA

1. Introduction
2. Defining Gifted
3. Why Separate Students?
4. Identification and Nurturing
5. Worldwide Support
6. Literature Regarding Gifted Students
7. Current Research  
Further Reading

## GLOSSARY

**aptitude** Specific ability in an academic area such as mathematics, spatial tasks, science, or verbal activities.

**elitism** An attitude critical of special programs for gifted and talented youth because they provide benefits not available to all students.

**enrichment programs** Special services for gifted youth that avoid the regular school program and curricula and that often offer projects and creative activities. Students are often pulled out of regular classes during the school day for these programs.

**gifted** Youth who have high general mental or artistic abilities.

**intelligence** General mental ability as often represented by gifted students.

**live-in experience** A service for gifted, talented, and/or precocious youth in which they leave their homes and go to live at the site where a special program is offered, often a college or university campus.

**precocity** Achievement or performance that surpasses the average performance of youth at a given age level. The term precocious is often used as a synonym of gifted.

**talent** Specific ability in an applied area such as mathematics, science, leadership, music, art, or writing.

Gifted, talented, and precocious students can be identified with intelligence and aptitude tests, rating scales, achievement test scores, and teacher observations. Once identified, schools can offer special classes, programs, and services to facilitate the growth of their gifts and talents. Colleges and universities also provide special programs and services in the forms of early admission to highly precocious youth as well as special weekend and summer programs.

## 1. INTRODUCTION

Gifted, talented, precocious, high aptitude, high achieving, and creative are all terms used to designate children and youth of high or superior ability. There is little agreement about test instruments, rating scales, and observational procedures used to identify these children, or cutting levels on the various test instruments that are used. However, there is some agreement that test scores 2 standard deviation above the mean or percentile scores at or above 90% are appropriate identification criteria. There is also agreement that it is desirable to gather two or more measures to identify candidates for special programs.

## 2. DEFINING GIFTED

The field of gifted education has generated a large research base, beginning with the work of Lewis Terman. Shortly after his development and publication in 1916 of the first intelligence test in the United States, the Stanford–Binet, he began a major longitudinal study of 12-year-old children with IQs of 135 or higher who were followed throughout their lives, with special interest in their long-term achievements. Another pioneer and researcher in the field was Leta Stetter Hollingworth, whose work focused on highly gifted youth and their education. Terman’s research established the initial focus on high IQ as the major means of identification, and his following research established that those who go on to high-level creative production in life profited from early grade advancement or acceleration and motivation to achieve at a high level. Hollingworth’s work established the critical need of highly gifted youth for challenging and motivating instruction and the problems they face in regular grade-level classrooms, often with teachers who do not understand them and their special needs.

## 3. WHY SEPARATE STUDENTS?

The major purpose of identification of children for gifted programs or educational services is, or should be, to determine if they need a special regimen beyond the services of a regular classroom program and what particular services or resources might be needed. Decisions about a child’s special education needs can best be made through the assessment of the child’s special aptitudes and talents or areas of high achievement. Then, if both aptitude and achievement tests reveal high mathematical ability and the need for higher level mathematics instruction, a decision may be made to place the child in the next higher grade level for mathematics instruction or in a special mathematics enrichment program. Alternately, if the child’s achievement test scores are all above the 95th percentile, after conferences with the child and his or her parents, a decision may be made to move the child to the next higher grade level for all instruction.

In some schools or communities, gifted or precocious children are grouped together in special classes or even in special schools for gifted children only. Teachers of such classes often have received special training and certification for work with gifted children and youth and use curricula designed especially for gifted and precocious children.

Several states have established schools for precocious secondary students, often with major academic focus on mathematics and science and offering a full-time experience. These schools are comparable to the special day schools offered in several cities such as New York and Chicago, again stressing mathematics and science. Admission may be strongly competitive, but the instruction and curriculum as well as association with other academically superior students make these schools ideal programs for academically precocious youth.

## 4. IDENTIFICATION AND NURTURING

Much of the attention given to gifted and talented youth has turned diagnostically, in the past decades, to identification and nurturing of their special talents. Rather than offering a general, homogeneous regimen and curriculum for all gifted youth, the new programs offer a variety of classes and services focused on their special and emerging talents. Thus, a child who has been identified as having exceptional art talent can participate in an after-school arts class, another who shows strong leadership ability may enroll in Saturday morning leadership classes, and yet another who is precocious in science can be matched up with a mentor-professor at a community college and spend one afternoon a week at the college working with the mentor-professor in a science lab. Talent-oriented programs use a wide variety of services and resources in the school and community to foster the development of children’s special talents.

Saturday and summer programs are also now widely available to gifted and talented children with or without school identification or endorsement. Although some of these programs make no pretense of identifying and serving the gifted, many of them explicitly aim to identify and serve gifted and talented youth. These programs often offer talent-oriented classes taught by experts in that field. For example, an actor or director may teach a class in dramatics, a computer professional may teach a class on computers, a math teacher may teach an advanced math class, an attorney may teach a class on the law as a profession, and a local novelist may teach a creative writing class. These classes supply a knowledge base in the field of the talent, offer a motivating challenge for talented youth, and provide mentors and models

in areas related to youth talents. Some programs, such as one at Purdue University, also offer mentoring experiences as a follow-up to the classroom in areas of youth talent.

## 5. WORLDWIDE SUPPORT

The field of gifted and talented education is now fostering identification and special education for gifted and talented youth in countries throughout the world through the auspices of the World Council for Gifted Education. Its major publication, *The International Handbook of Giftedness and Talent* (2000), is an excellent resource, with chapters written by the leading experts in gifted education from throughout the world. In the United States, the major professional organization with focus on gifted and talented youth is the National Association for Gifted Children, based in Washington, D.C. Most of the U.S. states also have professional organizations focused on gifted and talented youth.

## 6. LITERATURE REGARDING GIFTED STUDENTS

There are a number of publications devoted to the nature and educational needs of the gifted and their teachers, notably *The Gifted Child Quarterly*, the major publication of the National Association of the Gifted Children, Washington, D.C.; the *Gifted Child Today*, Prufrock Press, Waco, Texas; the *Journal of Secondary Gifted Education*, Prufrock Press, Waco, Texas; *Roeper Review*, the Roeper Institute, Bloomfield Hills, Michigan; the *Journal for the Education of the Gifted*, the Association for the Gifted, of the Council for Exceptional Children, Reston, Virginia; *High Ability Studies*, the journal of the European Council for High Ability, Bonn, Germany; *Gifted and Talented International*, World Council for Gifted and Talented Children, College of William and Mary, Williamsburg, Virginia; and *Gifted Education International*, AB Academic Publishers, Oxon, United Kingdom.

## 7. CURRENT RESEARCH

The field of gifted and talented education is diverse in the nature of its research and the direction of

educational services. Research runs the gamut from studies of the nature of talent and abilities to personality and attitudinal aspects of giftedness, methods of testing and identifying gifted youth and the appropriate test instruments, creativity and problem solving abilities of gifted and talented youth, gender differences in abilities, models' and mentors' roles in the development of gifts and talents, and nature versus nurture in the development of giftedness. Research on gifted and talented youth and in the field of gifted education is widely reported in the journals of the field and at conferences.

The field of gifted and talented education is also diverse in its conceptions and identification of giftedness, and in its educational services and programs for gifted and talented youth. It is sometimes viewed as elitist by mainstream education, and there is some opposition to special classes and services for these youth. Funding for research in gifted education is limited, but several centers at universities carry on excellent programmatic investigations, notably at the University of Connecticut, the University of Iowa, and Purdue University. It appears that educational programs for gifted and talented youth are now well established in many cities, U.S. states, and other countries, and they will continue to grow and provide valuable educational experiences for gifted youth.

### See Also the Following Articles

Intelligence and Culture ■ Intelligence Assessment  
■ Intelligence in Humans

### Further Reading

- Ellingson, M. K., Haeger, W. W., & Feldhusen, J. F. (1986). The Purdue mentor program: A university based mentorship experience for gifted children. *The Gifted Child Today*, 9(2), 2–5.
- Feldhusen, J. F. (1991). Full-time classes for gifted youth. *Gifted Child Today*, 14(5), 10–13.
- Feldhusen, J. F. (1995). Talent development vs. gifted education. *The Educational Forum*, 59(4), 346–349.
- Feldhusen, J. F. (2002). Guidelines for grade advancement of precocious children. *Roeper Review*, 24, 169–171.
- Feldhusen, J. F., & Koopsman-Dayton, D. D. (1987). Meeting special needs of the gifted through Saturday programs: An evaluation study. *Gifted International*, 4(2), 89–101.
- Feldhusen, J. F., Rehorn, L. S., & Sayler, M. F. (1998). Appropriate design for summer gifted programs. *Gifted Child Today*, 11(4), 2–5.

- Feldhusen, J. F., & Sayler, M. F. (1990). Special classes for academically gifted youth. *Roeper Review*, 12(4), 244–249.
- Gagné, F. (1999). *Tracking talents: Identifying multiple talents through peer, teacher, and self nomination*. Waco, TX: Prufrock Press.
- Hansen, J. B., & Feldhusen, J. F. (1994). A comparison of trained and untrained teachers of gifted students. *Gifted Child Quarterly*, 38(3), 115–123.
- Holahan, C. K., & Sears, R. R. (1995). *Genetic studies of genius: The gifted group in later maturity* (Vol. 6). Stanford, CA: Stanford University Press.
- Hollingworth, L. S. (1926). *Gifted children: Their nature and nurture*. New York: Macmillan.
- Terman, L. M. (1916). *Measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale*. Boston, MA: Houghton Mifflin.
- Terman, L. M. (1925). *Genetic studies of genius: Mental and physical traits of a thousand gifted children*, Vol. 1. Stanford, CA: Stanford University Press.
- Van Tassel-Baska, J. (1992). *Planning effective curriculum for gifted learners*. Denver, CO: Love Publications.



# Goal Setting and Achievement Motivation in Sport

Joan L. Duda

University of Birmingham, Birmingham, United Kingdom

1. Goal Setting in Sport
2. The Meaning of Goals
3. Additional Considerations and Future Directions  
Further Reading

## GLOSSARY

**discrete goal** A particular performance standard or objective.

**ego-involved goal** A goal of revealing superior competence relative to others.

**failure-oriented goal-setting style** Individual difference reflecting a concern about not demonstrating low ability compared to others and thus a tendency to set very easy or extremely difficult goals.

**goal orientation** Dispositional tendency regarding achievement goal focus.

**goal proximity** Having both short-term and long-term goals.

**goal-setting style** An individual difference factor capturing an athlete's preferences regarding the type of discrete goals that are set in training and/or competition.

**moderately challenging goals** Standards that are of an optimal difficulty (i.e., not too easy and not beyond one's capabilities).

**motor endurance task** A physical task that requires muscular or cardiovascular endurance for successful performance.

**outcome goal** A goal based on a competitive or comparative performance standard.

**performance goal** A goal to secure some discrete standard or end product that is self-referenced (e.g., a "personal best" achievement).

**performance-oriented goal-setting style** Individual difference reflecting a tendency to set goals related to learning, performance improvement, and increasing one's self-referent ability.

**pre-competitive anxiety** The cognitive (e.g., degree of worry) and somatic (e.g., heart racing) anxiety experienced by an athlete before a competitive event.

**process goal** A goal of trying to meet some standard of proficiency concerning technique or employment of strategy.

**success-oriented goal-setting style** Individual difference reflecting a tendency to set competitive outcome goals or strive for standards that entail social comparison.

**task-involved goal** A goal that entails an emphasis on demonstrating personal competence in a self-referenced way and that centers on energetic attempts at task mastery and/or improvement.

Any perusal of applied sport psychology texts, examination of prevailing research topics in major sport psychology journals, or conversational engagement with those involved in sport usually points to the relevance of goals, goal setting, and motivation in this domain. Indeed, investigations of youth, intercollegiate and Olympic athletes, and coaches have supported their propensity for the setting of goals in competition and training. Interestingly, though, this descriptive research also

indicates that athletes and coaches consider their goal-setting efforts as only marginally effective. That is, although the formulation of and striving for some target(s) or objective(s) appear to be endemic in the athletic context, the employment of goal setting in sport seems less productive than one would readily assume. A major premise of the present contribution is that the implementation of goal setting will remain inconsistent and the efficacy of such strategies unconvincing and restricted in scope unless the meaning of the goals and their link to variations in motivation processes are considered.

In this article, the underpinnings, directions, and major findings of research on goal setting in sport will be summarized. Limitations of this work will also be highlighted. We then move into a discussion of the concept of achievement goals. In particular, attention will turn to the ways in which the constructs and assumptions embedded in con-temporary achievement goal frameworks provide further insight into the types of goals that individuals strive for within the athletic milieu. Moreover, the implications of these different goal perspectives for athletes' motivation to achieve and the overall quality of their sport engagement will be addressed. In essence, it is argued that a consideration of variability in achievement goals helps complete the picture in terms of when goal setting will be effective and what contributes to motivationally sound and adaptive goal setting. The article concludes with a discussion of future directions and considerations regarding work linking goal setting and achievement goals.

## **1. GOAL SETTING IN SPORT**

In the view of Locke and colleagues (1981, p. 126), goals capture what "an individual is trying to accomplish; it is the object or aim of an action." Weinberg (1992) takes this definition a step further by proposing that meeting a goal means "attaining a standard of proficiency on a task, usually within a specified time limit" (p. 178). According to Locke and Latham, we typically are consciously aware of our goals, but, on some occasions, goals can reflect unconscious intentions or purposes. Particularly in terms of the former, it is held that goals initiate and regulate behavior. That is, goals are conceptualized in Locke's model to be the antecedent that is most proximal to performance.

Locke and Latham further argue that goals can have a direct and indirect effect on performance. It is assumed that goals directly influence performance outcomes by

centering our attention on task-relevant cues (and attenuating our attending to cues that are not task-relevant) and encouraging persistence and appropriate levels of exerted effort. With respect to the proposed indirect consequences of goal setting, resulting strategic planning and problem solving are held to be mediators of the goal setting and performance relationship. Clearly, the latter mechanisms are especially essential when individuals are trying to meet complicated, challenging, and long-term objectives. However, sometimes focusing on a task, trying one's hardest, and persevering are not enough to secure goal attainment. Before any of these direct or indirect influences are likely to come into fruition, it is suggested that individuals must be committed to the goal(s) at hand and possess the necessary ability or skills. Feedback in relation to goal progress is another proposed moderator that is viewed as being essential to goal-setting effectiveness.

Almost exclusively, research on goal setting in sport has been grounded in Locke's goal-setting theory. The major assumptions of this theoretical framework, particularly in terms of the existent sport-related work, are described next. All in all, these predictions indicate which type of goal is most likely to have the most substantial positive impact on performance.

### **1.1. Basic Tenets of Locke's Goal-Setting Theory**

There are three central propositions in Locke's model of goal setting as it has been tested in sport; namely, that (1) hard goals result in higher levels of performance than do less-demanding goals, (2) specific goals give rise to superior performance when contrasted with no goals, very general goals, or a vague goal such as "do your best," and (3) using short-term goals in conjunction with long-term goals will lead to better performance than using long-term goals exclusively. Overall in the goal-setting literature, and specifically in terms of research conducted within organizational settings, evidence has been garnered indicating that difficult and specific short- and long-term standards result in superior performance when contrasted with more vague goals or no goals. Indeed, a number of large-scale narrative reviews and meta-analyses have heralded the effectiveness of explicit, difficult, short- and long-term goal setting. What is especially impressive about this body of work is that

the value of specific and challenging goals has been supported in the field and in the laboratory, with respect to diverse settings, across different tasks, in the case of various populations, and in terms of short- and long-term performance indicators. Moreover, the proposed direct and indirect mechanisms by which goals are held to influence performance have been generally sustained in the general and organizational psychology literatures.

## 1.2. Testing of Locke's Theory in Sport/The Physical Domain

In an attempt to encourage more systematic investigation of the efficacy of goal setting in the sport realm, Locke and Latham proffered that the impact of goal setting on sport performance would be even more striking than what is observed in work-related or organizational situations. Further, they proposed a number of hypotheses regarding when, where, and why goal setting functions in the physical domain. With respect to research in real-life sport settings or laboratory experiments involving sport or physical-based tasks, however, the literature on the effectiveness of goal setting is a bit more conflicting and slightly less compelling than has been the case in work settings. In the available sport-focused narrative reviews, it is typically concluded that research has provided only qualified support for the specific standard versus "do your best" hypothesis in Locke's framework, and, on the whole, the findings in terms of goal proximity (i.e., having short-term along with long-term goals) and goal difficulty (i.e., more difficult goals result in better performance) have been equivocal.

In 1995, Kyllö and Landers published a meta-analysis that provided a slightly more optimistic but still rather constrained evaluation of studies on the effect of goal setting on physical task performance. Overall, in terms of the observed relationship between goal setting and performance witnessed in 36 studies, a mean effect size of .34 was revealed. This result is considerably more modest than what has been reported in omnibus meta-analyses not specific to the physical domain (i.e., the latter have found overall effect sizes ranging from .42 to .80). Providing only partial support for one of Locke's propositions, Kyllö and Landers also found moderately challenging goals to be associated with a greater effect size (.53) than difficult (.09) or easy (.07) goals. Consonant with Locke and Latham's hypothesis, combined short- and long-term goals were coupled with a

larger effect size (.48) than short-term (.38) and especially long-term (.19) goals only.

In their more recent appraisal of 56 published goal-setting studies in sport and physical activity, Burton *et al.* reported a 78.6% effectiveness rate. As this rate was greater than what was reported in an earlier version of the review chapter in question (in particular, 7 years previously, 66% of the published studies reviewed revealed a positive overall goal-setting effect), Burton and colleagues suggested that more support for the consistency of goal-setting effects is accruing as more research has been undertaken. This is a promising conclusion, but caution is warranted when such a deduction comes from a sample of published papers. That is, published research is more likely to show significant and expected effects when contrasted with non-published studies. Further, it is important to keep in mind that the conclusions of Burton and associates did not stem from a systematic, meta-analytic review.

There has been diversity in the methods used to test the veracity of Locke's theory when applied to sport. Particularly during the earlier work on this topic, more laboratory-based experiments flourished. The trend for such investigations was to focus on a simple, motor endurance-type task (e.g., sit-ups), recruit physical education or sport science majors as subjects, have the experimenter assign the goal(s), and then contrast a specific goal versus a general and/or vague "do your best" goal with respect to the impact on performance, often over a short time period. Sometimes a control group that was assigned "no goals" was incorporated in the design. Other investigations in the sport milieu have been field-based studies and/or interventions employing single-subject designs to examine the efficacy of goal-setting programs.

Another contemporary extension of work on goal setting that emanated particularly within the sport and exercise psychology literature is the distinction between outcome, performance, and process goals. When focused on a process goal, the athlete is trying to meet some standard of proficiency concerning her or his technique or employment of strategy. For a performance goal, the athlete is reaching to secure some discrete standard or end product that is self-referenced (e.g., trying to better one's sprint time). An outcome goal is based on a competitive or comparative performance standard, such as winning a contest. On the whole, sport research suggests that athletes focus on all three types of goals, and they may serve different purposes in terms of maximizing skill development and performance.

### 1.3. Limitations of Sport Goal-Setting Work

When trying to explicate the mixed results and typically more moderate effects when goal setting is tested in regard to sport or physical task performance, researchers have offered a number of methodological artifacts and design limitations for consideration. These factors, which to some degree can be considered interrelated, include: (1) the inclusion of individuals who tend to be highly motivated in the physical domain as study participants, (2) the tendency for study participants in the control or “do your best” groups to spontaneously set goals, (3) the likelihood of competition being manifested between study participants, and (4) the failure to consider study participants’ personal goals with respect to task performance. Without a doubt, such methodological shortcomings pose a serious threat to the internal validity of this research. Moreover, not systematically taking account of such factors in goal-setting research diminishes our understanding of why, when, and how goal setting may work in sport.

Characteristics of the tasks used in sport goal-setting research have also been pointed to as explanations for the compromised effectiveness of goal setting in this setting. Specifically, it has been suggested that physical tasks in and of themselves tend to provide feedback. An individual can see and feel how one is doing and how hard one is trying (especially in the case of those sit-up studies!), regardless of whether one is focused on a more-or-less specific or a more-or-less difficult goal. Both the process of performing and performance outcomes per se are more overt in the physical domain when contrasted to tasks in other settings, such as industrial or academic environments.

According to Hall and Kerr, the anomalies in the sport goal-setting literature hold ramifications for the employment of goal setting in applied sport situations. Addressing the methodological and task-related problems identified above in subsequent research may indeed enhance the internal validity of future goal-setting studies. What would be the cost, however, in terms of external validity? That is, if we controlled for spontaneous goal setting and personal objectives, competition, the probability that those involved in the usually voluntary activity of sport are often motivated to participate and perform, and the nature of the task, what does such research have to say to the real world of athletic engagement?

A central conceptual ambiguity that plagues status quo sport goal-setting work based on Locke’s model is the premise that goals are merely end states or external standards that energize and direct behavior. Such mechanistic assumptions about the nature and connotations of goals do not provide insight into what contributes to differences in the goals that individuals adopt, whether assigned by the experimenter, set by the participant, and/or spontaneously generated in “no goal” or vague goal conditions. Moreover, such a perspective on the purpose of goals does little to inform us about the implications of such goals for variation in ensuing motivation-related responses and achievement behaviors.

Although Latham and Locke considered goal setting to be a “motivational technique” that is effective, a conceptual (or indeed, to some degree, even empirical) consideration of motivational processes is sorely missing in the goal-setting literature stemming from Locke’s contributions. For example, commitment is assumed to be an important moderator of the goal setting–performance relationship in Locke’s framework, but what contributes to variability in commitment to goals, both at the onset and throughout the goal-setting enterprise? Adequate ability is held to be fundamental for translating goals into desired performance, but goal-setting theory does not consider potential discrepancies in how individuals judge their level of competence or suggest when or why low ability would prove problematic. In Locke’s goal-setting theory, difficult goals are presumed to foster effort, and this enhanced exerted effort is held to be a central way by which difficult goals contribute to superior performance. It is important to recognize, however, that people do not always try when the situation or task becomes challenging. Indeed, they may even intentionally “hold back” in such circumstances. Why might this occur? In short, such broader motivational issues, which have important implications for the degree to which discrete goals would lead to better performance and enhanced task or activity engagement, are not adequately addressed in Locke’s initial conceptualization or later treatises on goal setting.

## 2. THE MEANING OF GOALS

A contemporary theory of motivation that emphasizes the meaning of personal goals and the consequences of this meaning for corresponding motivational processes

and achievement behaviors is achievement goal theory (AGT). Achievement goals are held to reflect different ways of construing one's competence in terms of the demands and standards placed on the individual and the different ways of subjectively defining successful performance and performance progress. It is assumed, based on AGT, that the demonstration of high competence (or avoidance of demonstrating low competence) is at the heart of striving for achievement and serves as the core of people's views regarding goal accomplishment. In the prevailing achievement goal conceptualizations that have dominated sport research over the past 10 to 15 years, two major achievement goals are deemed to be operating. One is a task-involved goal, which entails an emphasis on demonstrating self-referenced competence. In this case, witnessing personal improvement, learning, and/or task mastery provide a sense of subjective success. If centered on an ego-involved goal, however, revealing superior competence relative to others is fundamental to perceptions of goal accomplishment. Perceptions of adequate (or inadequate) ability are normatively referenced when ego-involved and, as a result, performance attempts stemming from high effort and reflecting personal progress are insufficient for feeling successful. Thus, although AGT considers the level of one's perceptions of ability (high/low) to have motivational significance, the hallmark of this framework is that we must bear in mind how the person judges whether he or she has exhibited high versus low ability when engaged in a task.

Drawing from AGT, it is proposed that a focus on task-involved goals, regardless of whether the person in question is confident or perceives her or his ability to be less than outstanding, will result in a positive and adaptive achievement pattern. That is, a task-involved person is expected to be more likely to perform up to her or his ability level, try hard and be invested in the activity, prefer challenging tasks, and persevere, even when the circumstances become difficult. As long as things are going well (i.e., the person is experiencing successful goal accomplishment and perceives his or her ability to be high), an individual striving to meet ego-involved goals should be committed to the activity at hand and perform well. In this situation, the person would not shy away from challenge and would be willing to work—as long as that level of exerted effort does not take the spotlight off or lead to questions regarding her or his level of superior competence. Short- and especially long-term motivational problems are hypothesized when there is a centering on ego-involved goals and the individual has reached the

conclusion that her or his competence is inferior relative to relevant others. Such a person would be expected to reduce the commitment to the activity and rescind effort (perhaps in an initial attempt to "save face," but eventually, if the situation does not change, as a step toward withdrawal from this endeavor). Optimally challenging goals would be especially worrisome for an individual who is trying to meet ego-involved standards and who has doubts about her or his competence. Rather, the preference would be to work toward easy goals (so the probability of failing is reduced) or set goals that are much too difficult (so that failure can be attributed to task difficulty rather than one's inadequacy).

As Hall and Kerr eloquently argued, a consideration of achievement goals provides insight into what may be the meaning of assigned goals in typical goal-setting protocols and what lays the foundation for the personal goals that study participants might set in goal-setting investigations. Variations in achievement goals are also relevant to what type of feedback is sought and how the feedback provided is understood and used in studies of goal-setting effectiveness. Achievement goals are pertinent to the extent that competition is created among study participants in goal-setting studies, the reasons for competing, and how the competitive experience is construed (e.g., as a challenge or a threat). Finally, as achievement goals are assumed to act like an interpretive lens impacting how we view the task and ourselves in relation to the task (e.g., does our task performance have implications for self worth?), differences in achievement goal emphases should correspond to variations in commitment to the discrete goals that are set and the effort given to reach those standards, especially over time. The focus placed on task and ego goals also has repercussions for individuals' level of anxiety, strategy use, degree of confidence or self-efficacy, and other psychological and emotional mechanisms that lead to differential performance. In short, achievement goals color our approach to and the behavioral, cognitive, and affective consequences of the setting and striving for discrete performance standards.

## **2.1. Basic Tenets of the Achievement Goal Framework**

Achievement goal theory holds that, when performing achievement-related tasks, individuals can fluctuate in their state of involvement directed toward task or ego

goals. That is, they can be more or less task- and ego-involved at any point during task engagement. The likelihood of being task- and/or ego-involved is assumed to be influenced by dispositional tendencies regarding these states of goal involvement; i.e., people (once they have reached a particular level of cognitive development in late childhood and beyond) can vary in their degree of task and ego orientation. It is important to point out that these orientations tend to be orthogonal. Thus, with specific reference to goal setting, we are interested in how individuals who are high task/high ego, high task/low ego, low task/high ego, and low task/low ego might differ in the goals set, their processing of goal-related feedback, and the corresponding motivational processes and resulting behaviors. In the sport domain, two instruments with established validity and reliability have been developed to assess dispositional goal orientations, namely the Task and Ego Orientation in Sport Questionnaire and the Perceptions of Success Questionnaire.

Whether someone is striving to meet task- and/or ego-involved goal criteria during the goal setting process is also assumed to be dependent on the situationally and task-emphasized goals or the perceived motivational climate. In particular, the tasks involved and social environments surrounding individuals as they undergo goal setting vary in the degree to which they are more or less task- and/or ego-involving. With respect to the goal-setting process, these overriding climates are created by the type and level of the goals assigned, the nature of the feedback given, the evaluation criteria used to judge goal accomplishment, and the manner in which that goal accomplishment is recognized. If goal setting was undertaken in a task-involving climate, the goals set would be primarily improvement-focused and effort-dependent (i.e., process and performance goals), the feedback provided would be informational, self-referenced, and task-centered, the person himself/herself could be involved in the evaluation of goal accomplishment, and successful goal completion would be primarily tied to intrinsic satisfaction. The individual's level of commitment, perspective on goal challenge, and input would be strongly recognized in a task-involving goal-setting program.

Goal setting in an ego-conducive manner, on the other hand, would entail a preoccupation with normatively referenced goals (the majority would probably be outcome-oriented), the feedback given would be more judgmental and marked by social comparison, evaluation of goal accomplishment would make the person particularly aware of and concerned about his or her

relative ability, and the sources and nature of the recognition offered when goals are met (or not met) would help couple the individual's sense of self to successful (or unsuccessful) goal completion. Further, personal volition and the participant's contribution to the process would most likely take a back seat in an ego-involving goal-setting program.

Pertinent to sporting contexts, assessments have been developed to tap the perceived motivational climate created by the coach (e.g., the PMCSQ-2) and parents (e.g., the PIMCQ), in the view of athletes. Researchers have also pulled from the TARGET taxonomy formulated by Epstein to analyze the dimensions underlying the climates manifested in sport. The particular structures encapsulated in this taxonomy are the nature of the task, source of authority, bases and form of recognition, structure and focus of groups in the setting, criteria for evaluation, and use of time.

## 2.2. Achievement Goals, Discrete Goals, and Goal-Setting Styles

One source of confusion concerning goal setting in the achievement goal literature is the tendency to mistakenly equate task and ego goal orientations with differential goal-setting styles or discrete goal foci. As indicated above, contemporary work on goal setting acknowledges the existence of different types of discrete goals, i.e., process, performance, and outcome goals. Researchers and practitioners in sport psychology, when making their case that outcome goals and ego orientation should not be considered hazardous to achievement-related health, have implied that a pronounced ego orientation means that one predominantly aims to meet outcome goals. In contrast, it is suggested that a strong task orientation translates into a marked orientation toward process (and perhaps performance) goals. Other researchers and practitioners in the field have challenged these presuppositions. For example, Duda has argued that a person can interpret process-related, performance, and outcome standards in a task-involved as well as an ego-involved fashion.

According to Burton, the construct of goal-setting style refers to athletes' preferences regarding discrete goals during training or competition. Burton has proposed that there are three distinct goal-setting styles, namely performance-oriented, success-oriented, and failure-oriented. Performance-oriented athletes are assumed to have a pronounced task orientation.

Success- and failure-oriented athletes are held to be strongly ego-oriented. However, in the former instance, perceptions of ability are high and, in the latter, perceptions of ability are deemed inadequate. In Burton's competitive goal-setting model, athletes' dispositional goal orientations in concert with their perceived competence are hypothesized to result in the focus on particular discrete goals of varying difficulty (e.g., a performance orientation will likely lead to the setting of challenging performance goals, while a success orientation will most probably result in an emphasis on moderately difficult outcome goals). Once the discrete goal is selected, the direct and indirect effects embedded in Locke and Latham's framework should then come into fruition. Thus, achievement goals are presumed to primarily come to bear in terms of goal selection in this model.

In addition to these conceptual distinctions between goal orientations, discrete goals, and goal setting styles, Hall and Kerr pointed out that there is no compelling evidence supporting the errantly assumed linkages between these constructs (e.g., that highly task-oriented athletes do not set outcome goals). Moreover, discussed above, empirical studies of the goal-setting practices of athletes indicate that many athletes set process, performance, and outcome goals and focus on these divergent goals for different reasons. The achievement goal framework would offer the following to discussions regarding the variety in and relevance of discrete goals operating in sport: the prediction and understanding of an athlete's cognitive, affective, and behavioral responses over time are primarily a function of whether he or she is task- and/or ego-involved, not whether the athlete is currently striving to meet a process-, performance-, or outcome-based standard. That is, what such diverse standards mean to the athlete and the potential significance of meeting (or especially not meeting) the standard fluctuates depending on achievement goal emphases. It is these latter appraisals that color ensuing motivational processes.

With respect to research and applied work, therefore, it is important to keep in mind that athletes' goal orientations offer insight into their concerns when approaching, engaged in, or concluding involvement with achievement tasks. These goal orientations also provide awareness into different perspectives on the connotations of goal accomplishment (or the failure to meet goals). They can be conceptualized as a type of cognitive schema that influence the discrete goals that we may set in the physical domain as well as every

other facet of goal-setting and goal-striving processes. On the other hand, as stated by Hall and Kerr (2001) goals as considered in the goal-setting literature are "end states that regulate human action by specifying an aim or an objective standard for a specific task" (p. 227).

Perhaps it is best to illustrate these qualitative differences with a metaphor. Discrete performance standards are the star we envision visiting and the impetus for us to jettison into space. They are the dream or aim that comes closest to the landing on the star. Goal orientations, however, capture the reason for going into the "last frontier" in the first place. They guide the features and resiliency of the spaceship we decide to take, the characteristics of the long ride to that distant star, the reactions when we reach our destination, and the ramifications of such an achievement for the entire space program. That is, goal orientations and discrete goals are not one and the same.

### **2.3. Potential Mechanisms Influencing the Goal Setting- Performance Relationship**

What might be the mechanisms by which achievement goals impact the goal setting- performance relationship? The contemporary sport psychology literature contains a plethora of studies examining the interplay between achievement goals and various cognitive, affective, and behavioral responses in the sport domain. It is beyond the scope of this article to adequately review and do justice to this body of work. As they seem particularly applicable to the goal-setting process, though, it is important to highlight that variation in achievement goals (whether dispositional and/or situationally induced) have been found to correspond to (1) perceptions of ability, (2) differential beliefs about the causes of success, (3) anxiety responses and ways of coping with stress, (4) learning and performance strategy use, and (5) preferences regarding performance feedback.

In terms of individuals' perceived competence, both task and ego orientation (in cross-sectional investigations) tend to be positively associated. The correlations in these studies, however, tend to be low (typically,  $r = .2-.3$ ), which tells us that at any particular point in time, we can have highly task- and/or ego-oriented individuals who feel their competence to be high, moderate, or far from stellar. In research on the interdependence

between environmental achievement goal emphases and perceptions of ability, task-involving situational cues have been found to correspond to higher perceptions of ability. Ego-involving situational cues, however, seem more likely to lead to a deterioration of a sense of adequate competence, especially if perceptions of ability were already depleted or if the circumstances (e.g., experience of failure) are leading to questions about one's ability level. These findings make sense when one considers that a task goal emphasis means that one construes personal competence in a self-referenced way and centers on effortful attempts at task mastery and/or improvement. Such a perspective should contribute to a sense of demonstrated competence on numerous occasions. In the case of an ego goal emphasis, witnessing one's superiority with respect to others translates into high perceived ability. One's confidence to do what needs to be done would be more likely to "take some hits" when competence and success are construed in that manner.

Locke and Latham's theory of goal setting holds that a person must feel that he or she has what it takes for the setting of a discrete goal (especially a difficult one) to result in heightened performance. Whether one's capability is deemed to be sufficient to meet the standard set, it seems, is dependent on the achievement goal(s) manifested.

With respect to views about what leads to achievement, task orientation tends to be linked to the belief that hard work and training is essential. It is reasonable to suspect that such an outlook on the determinants of success should translate into a resilient tendency to not give up during the "failures" and continuously exert effort during the goal-setting process. Ego orientation, on the other hand, has been found to relate to the belief that the possession of high ability is necessary for success to occur. One cannot help but wonder, then, if someone is pursuing discrete goals in an ego-involved manner, what would be the impact of this corresponding belief when goal accomplishment has not been forthcoming?

A plethora of studies in the field of sport psychology have explored the anxiety-performance relationship. When stress levels go beyond what is optimal for the task and person at hand, performance debilitation ensues. Although much more work needs to be done on this topic, previous research has indicated that task orientation and a perceived task-involving climate tend to be negatively or unrelated to pre-competitive anxiety, whereas a positive association between dispositional or situationally induced ego goals (especially in the case of individuals with low perceived ability) and anxiety responses has received support. In terms of

coping responses, the literature typically points to a positive relationship between task goals and problem-focused strategies (which, in the short- and long-term, have been found to be effective). In contrast, ego orientation and a perceived ego-involving climate have been found to correspond to emotion- or avoidance-focused coping (which may be deemed beneficial at the moment, but does not correspond to long-term indicators of coping effectiveness).

Finally, preliminary work indicates that the achievement goals that are adopted have relevance for strategy use and preferences regarding performance feedback. In the studies to date, task goals have been coupled with more adaptive training and competition strategies such as trying to understand the coach's instructions and engaging in deep-processing when trying to learn a motor task. When athletes are high in ego orientation and/or perceive the climate as highly ego-involving, they have been found to be more likely to view avoiding practice as a viable strategy and to employ superficial learning strategies. Cury and his colleagues have conducted a series of interesting investigations regarding the type of feedback desired following task completion. Their research suggests that highly ego goal-focused individuals who doubt their ability tend to reject task-related or objective performance-related feedback. Clearly, it is difficult to see such people effectively enhancing their perceived or actual ability level with such feedback preferences. Strongly ego-oriented individuals who have high perceived ability seem more likely to seek out normative information (i.e., how good am I compared to others?) and reveal a low preference for objective performance-related feedback. Although seemingly not as debilitating as the feedback preferences of their high ego/lower perceived ability cohorts, such inclinations regarding preferred feedback are probably not optimal for meeting complex and difficult goals in the sport setting.

### **3. ADDITIONAL CONSIDERATIONS AND FUTURE DIRECTIONS**

In the view of Hall and Kerr (2001), the attractiveness of the achievement goal framework is that it enables researchers and practitioners "... to understand considerably more about the direction of behavior and about why athletes make particular motivational choices than is possible from considering Locke and Latham's (1990)

goal-setting theory alone . . . [which] limits us to understanding changes in intensity and persistence once discrete goals are set" (pp. 230–231).

Systematic research that attempts to integrate the constructs and tenets of achievement goal theory with the hypothesized goal-setting process is needed. In essence, the aim of such work would be to better inform us about when and why goal setting would be effective or ineffective, and what might be the best ways to implement goal setting in the sport domain. With respect to the potential mechanisms involved, one research direction would be to test variations in the proposed direct and indirect effects of goals on performance as proffered in Locke's goal-setting theory as a function of the achievement goals manifested. In terms of differences in achievement goal emphases, it would be interesting to consider people's goal orientations and the motivational climate in which they are operating as well as the interaction between these two factors. It would also be conceptually intriguing and practically important to examine the relationship of achievement goals to the nature and motivational implications of the process, performance, and outcomes goals set by athletes in both training and competition. While pursuing such lines of inquiry, the impact of variability in athletes' perceived ability on the interdependence between achievement goals and the goal-setting process should be determined.

In the discussion of achievement goals in this article, the emphasis has been placed on what have recently been termed dichotomous goal models. Such models distinguish competence-based striving in terms of a task (or mastery) and ego (or performance) focus. This approach was adopted for several reasons. First, the two-goal approach to achievement motivation has laid the foundation to date for the large majority of achievement goal research in sport. Second, existing discussions in the literature concerning distinctions between the constructs of achievement goals and discrete goals have been grounded in dichotomous conceptualizations. Finally, space constraints precluded the possibility of fully describing and developing the recently proposed trichotomous goal models and "2 × 2" achievement goal framework in terms of the present discourse. These latter frameworks differentiate approach achievement goals [goals directed toward the attainment of favorable judgments of competence, whether that competence is self-referenced or task-based (i.e., approach task) or normatively referenced (i.e., approach ego)] from avoidance goal perspectives [goals centered on avoiding

unfavorable judgments of ability, whether those perceptions of ability are self-referenced or task-based (i.e., avoidance task) or normatively referenced (i.e., avoidance ego)]. As these multiple goals have been found to differentially link to motivation-related responses such as intrinsic interest, anxiety, and performance, incorporating the approach–avoidance distinction in forthcoming studies of achievement goals and goal setting should prove informative.

Duda made a distinction between the quantity and quality of athletes' motivation. The former refers to an athlete being energized at that point in time and performing adequately. The latter relates to whether (1) the athlete wants to and does invest over time, (2) the activity contributes to the athlete's physical, psychological, emotional, and moral growth and well-being, (3) the athlete usually is able to perform up to her or his ability, and (4) the athlete enjoys the activity and is self-determined. At best, goal-setting theory alone can provide an appreciation of whether motivation is high or low from a quantity perspective. Integrating goal setting within an achievement goal framework, however, promises insight into both aspects. In so doing, we emphasize other achievement behaviors (e.g., perseverance) along with performance. Most critically, though, we consider other motivational outcomes of significance besides achievement (e.g., the athletes' welfare). Finally, as pointed out by Hall and Kerr, we become cognizant of the fact that (and the reasons why) goal setting can help promote quality sport engagement or be maladaptive (i.e., contributing to the quantity of motivation but compromising the quality and/or eventually leading to an erosion of quantitatively high motivation as well). All in all, based on the tenets of AGT and related research, it seems that more task-involving goal setting is what will reap the greatest motivational gain in sport or any other achievement domain.

### See Also the Following Articles

Arousal in Sport ■ Intrinsic and Extrinsic Motivation in Sport ■ Psychological Skills Training in Sport ■ Self-Confidence in Athletes

### Further Reading

Ames, C. (1992). Classrooms: Goal, structures, and student motivation. *Journal of Educational Psychology*, 84, 261–271.  
Burton, D. (1989). Winning isn't everything: Examining the impact of performance goals on collegiate swimmers' cognitions and performance. *The Sport Psychologist*, 3, 105–132.

- Burton, D., & Naylor, S. (2002). The Jekyll/Hyde nature of goals: Revisiting and updating goal-setting in sport. In T. Horn (Ed.), *Handbook of research on sport psychology*, 2nd ed. (pp. 459–500). Champaign, IL: Human Kinetics.
- Burton, D. (1993). Goal setting in sport. In R. N. Singer, M. Murphey, & L. K. Tennant (Eds.), *Handbook of research on sport psychology* (pp. 467–491). New York: Macmillan.
- Burton, D., Naylor, S., & Holliday, B. (2001). Goal setting in sport: Investing the goal effectiveness paradox. In R. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (pp. 497–528). New York: John Wiley.
- Burton, D., Weinberg, R., Yukelson, D., & Weigand, D. (1998). The goal effectiveness paradox in sport: Examining the goal practices of collegiate athletes. *The Sport Psychologist*, 12, 404–418.
- Chidester, T. R., & Grigsby, W. C. (1984). A meta-analysis of the goal-setting performance literature. In J. A. Pearce, & R. B. Robinson (Eds.), *Academy of management proceedings* (pp. 202–206). Ada, OH: Academy of Management.
- Duda, J. L. (1989). The relationship between task and ego orientation and the perceived purpose of sport among male and female high school athletes. *Journal of Sport and Exercise Psychology*, 11, 318–335.
- Duda, J. L. (1992). Sport and exercise motivation: A goal perspective analysis. In G. Roberts (Ed.), *Motivation in sport and exercise* (pp. 57–91). Champaign, IL: Human Kinetics.
- Duda, J. L. (1996). Maximizing motivation in sport and physical education among children and adolescents: The case for greater task involvement. *Quest*, 48, 290–302.
- Duda, J. L. (1997). Perpetuating myths: A response to Hardy's 1996 Coleman Griffith Address. *Journal of Applied Sport Psychology*, 9, 307–313.
- Duda, J. L. (2001). Goal perspectives in sport: Pushing the boundaries and clarifying some misunderstandings. In G. C. Roberts (Ed.), *Advances in motivation in sport and exercise* (pp. 129–182). Champaign, IL: Human Kinetics.
- Duda, J. L. (2004). Verifying the values and sparking the spirit: Research and practice on motivation in sport psychology. In P. Markula (Ed.), *Sharing joy, sharing pain: Feminist narratives on sport studies* (in press). New York: SUNY Press.
- Duda, J. L., & Hall, H. K. (2001). Achievement goal theory in sport: Recent extensions and future directions. In R. Singer, H. Hausenblas, & C. Janelle (Eds.), *Handbook of sport psychology*, 2nd ed. (pp. 417–443). New York: John Wiley & Sons.
- Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology*, 84, 290–299.
- Duda, J. L., & Whitehead, J. (1998). Measurement of goal perspectives in the physical domain. In J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 471–484). Morgantown, WV: Fitness Information Technology.
- Dweck, C. S. (1999). *Self-theories and goals: Their role in motivation, personality, and development*. Philadelphia, PA: Taylor & Francis.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychology*, 34, 169–189.
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72, 1–15.
- Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501–519.
- Epstein, J. (1989). Family structures and student motivation: A developmental perspective. In C. Ames, & R. Ames (Eds.), *Research on motivation in education* (Vol. 3) (pp. 259–295). New York: Academic Press.
- Fry, M. D. (2001). The development of motivation in children. In G. C. Roberts (Ed.), *Advances in motivation in sport and exercise* (pp. 51–78). Champaign, IL: Human Kinetics.
- Hall, H. K. (1990). *A social-cognitive approach to goal setting: The mediating effects of achievement goals and perceived ability*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Hall, H. K., & Byrne, A. T. J. (1988). Goal setting in sport: Clarifying recent anomalies. *Journal of Sport and Exercise Psychology*, 10, 184–198.
- Hall, H. K., & Kerr, A. W. (2001). Achievement goal research in sport: Pushing the boundaries and clarifying some misunderstandings. In G. Roberts (Ed.), *Advances in motivation in sport and exercise* (pp. 183–234). Champaign, IL: Human Kinetics.
- Hardy, L. (1997). Three myths about applied consultancy work. *Journal of Applied Sport Psychology*, 9, 277–294.
- Hardy, L., Jones, G., & Gould, D. (1996). *Understanding psychological preparation for sport: Theory and practice for elite performers*. Chichester, UK: Wiley.
- Harwood, C., Cumming, J., & Fletcher, D. (2004). Motivational profiles and psychological skill use within elite youth sport. *Journal of Applied Sport Psychology*, in press.
- Jones, G., & Hanton, S. (1996). Interpretation of competitive anxiety symptoms and goal attainment expectancies. *Journal of Sport and Exercise Psychology*, 18, 144–157.
- Kim, M-S., & Duda, J. L. (2004). The doping process: Cognitive appraisals of stress, coping strategies, and coping effectiveness. *The Sport Psychologist*, 17.
- Kingston, K. M., & Hardy, L. (1997). Effects of different types of goals on processes that support performance. *The Sport Psychologist*, 11, 277–293.
- Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3, 157–189.
- Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and Preventative Psychology*, 5, 117–124.

- Locke, E. A., & Latham, G. P. (1985). The application of goal setting to sports. *Journal of Sport Psychology*, 7, 205–222.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin*, 90, 125–152.
- Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behaviour and Human Decision Processes*, 39, 52–83.
- Newton, M., & Duda, J. L. (1999). The interaction of the motivational climate, dispositional goal orientations, and perceived ability in predicting indices of motivation. *International Journal of Sport Psychology*, 30, 63–82.
- Newton, M., Duda, J. L., & Yin, Z. (2000). Examination of the psychometric properties of the Perceived Motivational Climate in Sport Questionnaire-2 in a sample of female athletes. *Journal of Sport Sciences*, 18, 275–290.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328–346.
- Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Cambridge, MA: Harvard University Press.
- Ommundsen, Y., & Roberts, G. C. (1999). Effect of motivational climate profiles on motivational indices in team sport. *Scandinavian Journal of Medicine and Science in Sports*, 9, 389–397.
- Roberts, G. C. (2001). Understanding the dynamics of motivation in physical activity: The influence of achievement goals on motivational processes. In G. C. Roberts (Ed.), *Advances in motivation in sport and exercise* (pp. 1–50). Champaign, IL: Human Kinetics.
- Roberts, G. C., Treasure, D. C., & Balague, G. (1998). Achievement goals in sport: The development and validation of the Perception of Success Questionnaire. *Journal of Sport Sciences*, 16, 631–645.
- Swain, A., & Jones, G. (1995). Effects of goal-setting interventions on selected basketball skills: A single-subject design. *Research Quarterly for Exercise and Sport*, 66, 51–63.
- Thill, E., & Brunel, P. (1995). Ego involvement and task involvement: Related conceptions of ability, effort, and learning strategies among soccer players. *International Journal of Sport Psychology*, 30, 101–122.
- Treasure, D. C. (1997). Perceptions of the motivational climate and elementary school children's cognitive and affective responses. *Journal of Sport and Exercise Psychology*, 19, 278–290.
- Treasure, D. C. (2001). Enhancing young people's motivation in youth sport: An achievement goal approach. In G. C. Roberts (Ed.), *Advances in motivation in sport and exercise* (pp. 79–100). Champaign, IL: Human Kinetics.
- Tubbs, M. E. (1986). Goal setting: A meta-analytic examination of the empirical evidence. *Journal of Applied Psychology*, 71, 474–483.
- Weinberg, R. S., Bruya, L. D., & Jackson, A. (1985). The effects of goal proximity and goal specificity on endurance performance. *Journal of Sport Psychology*, 7, 296–305.
- Weinberg, R. S., Bruya, L. D., & Jackson, A. (1990). Goal setting and competition: A reaction to Hall and Byrne. *Journal of Sport and Exercise Psychology*, 12, 92–97.
- Weinberg, R. S., Burke, K. L., & Jackson, A. (1997). Coaches' and players' perceptions of goal setting in junior tennis: An exploratory investigation. *The Sport Psychologist*, 11, 426–439.
- Weinberg, R. S., Burton, D., Yukelson, D., & Weigand, D. (1993). Goal setting in competitive sport: An exploratory investigation of practices of collegiate athletes. *The Sport Psychologist*, 7, 275–289.
- Weinberg, R. S., Burton, D., Yukelson, D., & Weigand, D. (2000). Perceived goal setting practices of Olympic athletes: An exploratory investigation. *The Sport Psychologist*, 14, 279–295.
- Weinberg, R. S., Butt, J., & Knight, B. (2001). High school coaches' perceptions of the process of goal setting. *The Sport Psychologist*, 15, 20–47.
- Weinberg, R. S., Butt, J., Knight, B., & Perritt, N. (2001). Collegiate coaches' perceptions of their goal setting practice: A qualitative investigation. *Journal of Applied Sport Psychology*, 13, 374–398.
- Weinberg, R. S., & Weigand, D. (1993). Goal setting in sport and exercise: A reaction to Locke. *Journal of Sport and Exercise Psychology*, 15, 88–96.
- Weinberg, R. S., & Weigand, D. (1996). Let the discussions continue: A reaction to Locke's comments on Weinberg and Weigand. *Journal of Sport and Exercise Psychology*, 18, 89–93.
- White, S. A. (1996). Goal orientations and perceptions of the motivational climate initiated by parents. *Pediatric Exercise Science*, 8, 122–129.
- Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. *Journal of Applied Psychology*, 72, 416–425.
- Woodman, T., & Hardy, L. (2001). Stress and anxiety. In R. Singer, H. Hausenblas, & C. Janelle (Eds.), *Handbook of sport psychology*, 2nd ed. (pp. 290–318). New York: John Wiley & Sons.



# Group Counseling

Janice L. DeLucia-Waack

University at Buffalo, State University of New York, Buffalo, New York, USA

1. Introduction
  2. Definition of Group Counseling
  3. Therapeutic Factors Theory and Research
  4. Effective Interventions Theory and Research
  5. Considerations
  6. Conclusion
- Further Reading

This article provides a theoretical and empirical rationale for successful group counseling and therapy along with precise definitions. The literature is reviewed with regard to therapeutic factors and effective interventions in group counseling. Considerations for facilitating successful counseling groups are suggested.

## GLOSSARY

**group counseling** A small group intervention that uses therapeutic factors and specific interventions to address personal and interpersonal problems and to promote personal and interpersonal growth and development (e.g., a women's group).

**group therapy** A small group intervention that uses therapeutic factors and specific interventions to address personal and interpersonal problems and to remediate perceptual and cognitive distortions or repetitive patterns of dysfunctional behavior (e.g., an eating disorders group).

**psychoeducational and guidance groups** A small group intervention that uses group dynamics and principles to teach skills in a preventive way (e.g., a friendship skills group).

**task and work groups** A small group activity that uses organizational and community dynamics to facilitate the goals of a group (e.g., a book club).

**therapeutic factors** An element of group therapy that contributes to improvement in a patient's condition and is a function of the actions of the group therapist, the other group members, and the client himself or herself.

## 1. INTRODUCTION

Current group counseling theory and practice has its roots in two divergent events: Sigmund Freud's famous Wednesday night meetings at his house in Vienna and John Pratt's treatment of tuberculosis patients in group "classes" in the United States. Many theoretical approaches and applications have emerged from these early beginnings to create an efficient and empirically validated treatment in psychology today. As research methods became more precise and more interest in studying the effectiveness of groups accrued, the focus moved from showing that groups were more effective than no treatment at all to establishing that groups for certain populations and problems were as effective as, or even more effective than, other active treatments. As research, theory, and practice developed in the field, the pertinent question changed from "Are groups effective?" to "What kinds of groups are effective for what types of people with what problems?"

## 2. DEFINITION OF GROUP COUNSELING

Toward the goal of answering the question of what kinds of groups are most effective for certain types of people with certain types of problems, the Association for Specialists in Group Work (ASGW) in 2000 delineated four types of groups based on their goals and interactional processes—task/work, psychoeducational/guidance, counseling, and therapy groups—to aid in the selection of the appropriate types of groups for different populations (e.g., age groups) with different goals (e.g., combating depression, learning social skills, preventing eating disorders). According to ASGW, counseling groups “address personal and interpersonal problems of living and promote personal and interpersonal growth and development,” whereas therapy groups “address personal and interpersonal problems of living, remediate perceptual and cognitive distortions or repetitive patterns of dysfunctional behavior, and promote personal and interpersonal growth and development.” Examples of counseling and therapy groups include general interpersonal groups, training groups for students learning to be counselors or therapists, and groups directed at the amelioration of specific problems such as depression, eating disorders, and sexual abuse. Both counseling and therapy groups, as defined by ASGW, use “group-based cognitive, affective, behavioral, or systemic intervention strategies.” Although by definition counseling and therapy groups are different in practice, these terms are often used interchangeably and are often difficult to differentiate in practice.

In contrast, task/work groups focus on “efficient and effective accomplishment of group tasks” and psychoeducational/guidance groups focus on skill development to prevent problems, according to ASGW. Both of these groups use “group-based educational and developmental strategies.” Examples of task/work groups include committees, clubs, and classrooms that have common goals to work toward, whereas psychoeducational/guidance groups teach specific skills and coping strategies such as anger management, social skills, self-esteem, assertiveness, and making friends.

In further contrast, counseling and therapy groups tend to be of longer duration and smaller in size than do psychoeducational groups, the latter of which are often only 6 to 10 sessions and may consist of 10 or more members. Whereas task/work and psychoeducational/guidance groups tend to use educational strategies and practice to teach new skills, the emphasis in

counseling and therapy groups is on the use of group process and dynamics to help members identify interpersonal difficulties, identify new ways of thinking and behaving to solve interpersonal problems, practice new behaviors, get feedback from other group members, and learn coping skills from other group members. The major premise inherent in counseling and therapy groups is that group members eventually act in the group in ways similar to how they behave in their relationships with others (e.g., significant others, friends, co-workers). Thus, as behaviors are enacted, group members can evaluate (with the help of other group members) the effectiveness of these behaviors and their impact (both positive and negative) on group members’ relationships. In addition, group members then have the opportunity to practice and try out new behaviors and to receive feedback from others about the perceived impact of the new behaviors.

## 3. THERAPEUTIC FACTORS THEORY AND RESEARCH

Over the years, many group researchers and practitioners have attempted to explain the process of group counseling and therapy by describing 11 therapeutic factors that operate in groups. Although Yalom was not the first researcher or theorist to use the concept of therapeutic factors, his work has had an enormous influence on group theory and research. In 1975, Yalom’s compendium of 11 therapeutic factors was built on earlier writings of group theorists and practitioners and on his research on important change mechanisms in group therapy. In 1985, Bloch and Crouch defined a therapeutic factor as “an element of group therapy that contributes to improvement in a patient’s condition and is a function of the actions of the group therapist, the other group members, and the client himself.” [Table I](#) provides the definitions of these 11 therapeutic factors.

Not only did Yalom define and describe 11 therapeutic factors, he also developed a research paradigm that has had a major influence on how researchers examine therapeutic factors in group settings. In a now classic study, Yalom and colleagues in 1973 developed a 60-item therapeutic factors Q-sort by writing five items that corresponded to each of the 11 factors described by Yalom (12 factors were used because interpersonal learning was split into interpersonal learning [input] and interpersonal learning [output] categories), asking group members to put items in

**TABLE I**  
**Descriptions of Therapeutic Factors**

<i>Therapeutic factor</i>	<i>Definition</i>
Instillation of hope	Member recognizes other members' improvement and that group can be helpful; member develops optimism for his or her own improvement
Universality	Member perceives that other members share similar feelings or problems
Imparting information	Advice giving by therapist or fellow members
Altruism	Member gains a positive view of himself or herself through extending help to others in group
Corrective recapitulation of primary family group	Member experiences the opportunity to reenact some critical familial incident with members of the group in a corrective manner
Development of socializing techniques	Group provides member with an environment that allows member to interact in a more adaptive manner
Interpersonal learning (input)	Member gains personal insight through other members' sharing of their perceptions of the member
Interpersonal learning (output)	Group provides member with an environment that allows member to interact in a more adaptive manner
Cohesiveness	Feeling of togetherness is provided and experienced by the group
Catharsis	Member releases feelings about past or here-and-now experiences; this release leads to member feeling better
Existential factors	Member ultimately accepts that he or she must take responsibility for his or her own life
Imitative behavior	Member learns through the observation of others' learning experiences

categories ranging from "least helpful to me in group" to "most helpful to me in group." From these data, a ranking of the relative importance of the therapeutic factors, as perceived by successfully treated clients, was made from most important to least important as follows:

1. Interpersonal learning (input)
2. Catharsis
3. Cohesiveness
4. Self-understanding
5. Interpersonal learning (output)
6. Existential factors
7. Universality
8. Instillation of hope
9. Altruism
10. Family reenactment
11. Guidance
12. Identification

Yalom's paradigm for examining the importance of the therapeutic factors has spawned a large number of studies of therapeutic factors in groups. Several measures have been developed to assess therapeutic factors in groups (e.g., Critical Incidents Questionnaire, Therapeutic Factors Scale, Curative Factors Scale,

Therapeutic Factors Inventory). Group researchers have now amassed a number of studies that identify clients' ranking of the therapeutic factors across a number of different settings and patient populations. Studies over the years have assessed ratings of therapeutic factors and yielded inconsistent results. These inconsistencies may be due to differences in the groups surveyed or in the methods used to assess therapeutic factors. In an effort to make sense of the inconsistent findings, Kivlighan and Holmes used a cluster analysis of studies that rated the importance of therapeutic factors. Four clusters emerged from the analysis based on the therapeutic factors that were highest and lowest in each cluster. Cluster 1 (affective insight groups) rated acceptance, catharsis, interpersonal learning, and self-understanding as the most valued factors. These groups also rated both guidance and vicarious learning as relatively unimportant. Cluster 2 (affective support groups) rated acceptance, installation of hope, and universality as the most important group factors. Similar to Cluster 1, members of the groups in Cluster 2 rated guidance and vicarious learning as relatively unimportant. Cluster 3 (cognitive support groups) rated vicarious learning and guidance as highly valued while rating self-understanding much

lower. Cluster 4 (cognitive insight groups) rated interpersonal learning, self-understanding, and vicarious learning as the most valued therapeutic factors. The topic of therapeutic factors continues to be intriguing to researchers and practitioners alike, especially as related to the questions of importance in differential stages of group development and for different types of groups.

#### 4. EFFECTIVE INTERVENTIONS THEORY AND RESEARCH

When leaders employ a given intervention during the course of a group session, there is the implicit assumption that their behavior will have a positive impact on individual members and/or on the processes and outcomes of the group. Recent research findings have indicated that group dynamics and outcomes are influenced by factors such as group leaders' interaction styles, personal characteristics, and attitudes. Researchers have found that leaders who are less controlling, exhibit warmth/caring, and set and reinforce clear norms are more likely to have cohesive groups. Increased cohesiveness, in turn, has been found to be positively related to a variety of group treatment outcomes. Morran and colleagues highlighted 10 specific group interventions with research and clinical support.

The first category of interventions, protecting group members and promoting safety, includes protecting, blocking, and supporting. These interventions are typically most useful in the leader's efforts to provide a group climate that is conducive to trust, openness, and cohesion. Research has indicated that members who are dissatisfied with their group experience often implicate the group leader for negligence in providing adequate protection. Sometimes, protecting can be indirect and might include member selection/exclusion procedures, the establishment of appropriate group norms, or the modeling of caring for group members. More direct protecting interventions might take the form of intervening to stop a member who is self-disclosing too much or at a level that is significantly more intimate than the that of the rest of the members. Blocking is a specific type of protection that is used to stop a member from storytelling, rambling, or inappropriately probing, gossiping, or invading the privacy of others. Research suggests that group members frequently attribute their damaging experiences to undue confrontation, criticism, and pressure to self-

disclose by other members. Thus, it is essential that the group leader use blocking to protect potentially vulnerable group members. Supportive interventions help to reassure, reinforce, and encourage members to participate in the group and try out new behaviors. One caution is that premature or excessive supporting interventions may foster dependency on the leader and be detrimental to the individual member and the group. However, when timed appropriately, supporting interventions can provide the encouragement needed for a group member to become involved in the group interaction.

The second category of interventions, energizing and involving group members, includes drawing out, modeling, linking, processing, interpreting, self-disclosing, and providing feedback. These interventions are most often used to stimulate forward progress, increase member participation, and enhance interpersonal learning. Drawing out occurs when the leader directly invites comments or involvement to encourage participation from members who find it hard to share with others or with those who share on a surface level but avoid deeper issues. The leader should encourage participation while still allowing members to choose their own levels of sharing. Modeling occurs when the leader demonstrates the skills, attitudes, and other characteristics he or she hopes to engender within group members, including respect and caring for others, appropriate self-disclosure, feedback (both giving and receiving), and openness. Many group theorists suggest that modeling is a common factor of nearly all interventions because members naturally observe group leaders and other members and tend to imitate what they see demonstrated. Research has found that behaviors displayed by the therapist, including interpersonal behaviors, feedback delivery and acceptance, and "here-and-now" communications; lead to increases in those same behaviors by group members. As a whole, studies suggest that leaders can use modeling interventions to shape positive group dynamics; however, care should be exercised to avoid the unintentional modeling of undesirable behaviors. Linking connects what one group member is saying or doing with the concerns of one or more other members. Linking has been shown to be particularly useful in promoting member interaction, cohesion, trust, and universality among group members.

Processing as an intervention occurs when the group leader and members capitalize on significant happenings to help members reflect on the meanings of their experiences; better understand their own thoughts,

feelings, and actions; and generalize what is learned to situations outside of the group. Simply experiencing events in the group is not sufficient for growth and must be augmented by processing that provides a framework for retaining, integrating, and generalizing the experience. Research findings have identified a positive relationship between higher levels of group processing and greater outcome gains related to both group productivity and individual member achievement and suggest that processing should go beyond emotional insight and also focus on cognitive dynamics or meaning attributions to help members frame and understand their emotional experiences. Effective processing of critical incidents requires that members engage in a process of sharing and exploring experiences among themselves to make sense of what happened, establish what they have learned from the incidents, and begin to plan and practice new behaviors to be used outside of group sessions.

Interpreting involves offering possible explanations for events, behaviors, thoughts, and feelings at a deeper level to be considered as part of the change process. Group members typically experience interactions that provide the opportunity to learn more about themselves. However, such insight may be difficult for members to acquire without some cognitive framework being provided by the leader. In such situations, the leader can assist group members by providing tentative interpretations that can be discussed and considered in the group. Research suggests that leader interpretations help members to integrate complex personal and group-related events, thereby encouraging their investment in the group experience and facilitating generalization from group experiences to personal experiences outside of the group. Interpretations about members' impact on the environment and their patterns of behavior are most associated with member change.

Leader self-disclosure is an intervention whereby the therapist reveals his or her personal feelings, experiences, or here-and-now reactions to group members. As a general rule, the leader should self-disclose only when the information or reaction is directly related to what is happening in the group, models a behavior for members, and/or is expected to result in a direct benefit to the group.

## 5. CONSIDERATIONS

Much research has been conducted about counseling groups, but there is still much more that needs to be

known about how to facilitate groups effectively. Several considerations are important in leading effective counseling groups. First, sufficient training is needed in the type of group to be facilitated. ASGW provides specific guidelines for the training and supervision of task, psychoeducational, counseling, and therapy groups. Research suggests that when group therapists lack supervision, they are not able to identify mistakes and generate new plans of action; instead, they become stuck in a cycle of repeated ineffective interventions. Co-leadership is also recommended as a training tool and generally to enhance group effectiveness. It is suggested that co-therapy is better for a variety of reasons such as efficiency (e.g., the group can still continue if one leader is sick or unavailable), effectiveness (e.g., two leaders can track group members more easily), supervision and training of new group therapists and as a way in which to model healthy communication patterns (e.g., male-female interactions) between co-leaders in front of the group members. It is also theorized that co-leaders who model appropriate interaction between themselves, such as resolving disagreements and compromising, will have group members who demonstrate better skills in these areas.

Second, diversity and multicultural issues groups are inherent in groups and are essential to consider. Group leader competence is critical, specifically multicultural group counseling skills. In addition, knowledge of other cultural worldviews and values, and of how these may inhibit or increase a potential group member's willingness to participate and behave in the group, is essential. Acculturation is noted as an important variable in predicting the behavior of potential group members. Effective groups must help members to understand themselves and others as individuals within the context of their culture and must choose interventions and methods of change based on the interplay between the individuals and their worldviews.

Third, there is support for conducting groups that are homogenous in age and gender due to the impact of cognitive development and different life tasks across age groups on the purpose and content of the groups. The structure of a group, types of exercises, focus on process versus content, and group goals all are determined by the abilities and needs of the group members.

Fourth, preventive and brief interventions can be effective and may ameliorate problems before they become more serious. Thus, psychoeducational groups may be useful, as may counseling and therapy groups.

## 6. CONCLUSION

Group counseling has been shown to be effective with various populations, in different settings, and for specific problem areas. Research clearly shows that group goals must be clearly defined, the leader must have specific leadership training and skills, and the leader must take into consideration group members' expectations about the group, willingness to participate, and cultural expectations and values when designing a group and implementing specific interventions.

### See Also the Following Articles

Groupthink ■ Intergroup Relations and Culture ■ Interpersonal Behavior and Culture

### Further Reading

Association for Specialists in Group Work. (2000). Association for Specialists in Group Work: Professional

- standards for the training of group workers. *Journal for Specialists in Group Work*, 25, 327–354.
- Barlow, S. H., Fuhriman, A. J., & Burlingame, G. (in press). The history of group counseling and psychotherapy. In J. L. DeLucia-Waack, D. A. Gerrity, C. R. Kalodner, & M. T. Riva (Eds.), *Handbook of group counseling and psychotherapy*. Thousand Oaks, CA: Sage.
- Bloch, S., & Crouch, E. (1985). *Therapeutic factors in psychotherapy*. Oxford, UK: Oxford University Press.
- Kivlighan, D. E., Jr., & Holmes, S. E. (in press). The importance of therapeutic factors in counseling groups: A typology of therapeutic factors studies. In J. L. DeLucia-Waack, D. Gerrity, C. R. Kalodner, & M. Riva (Eds.), *Handbook of counseling and group psychotherapy*. Thousand Oaks, CA: Sage.
- Morran, D. K., Stockton, R., & Whittingham, M. (in press). Effective leader interventions for counseling and therapy groups. In J. L. DeLucia-Waack, D. Gerrity, C. R. Kalodner, & M. Riva (Eds.), *Handbook of counseling and group psychotherapy*. Thousand Oaks, CA: Sage.
- Yalom, I. D. (1995). *The theory and practice of group psychotherapy* (4th ed.). New York: Basic Books.



# Group Dynamics in Sport

Packianathan Chelladurai

Ohio State University, Columbus, Ohio, USA

1. Introduction
  2. Definition of a Sport Team
  3. Task Types in Sport
  4. Measures of Group Performance
  5. Cohesion in Sport Teams
  6. The Multidimensional Model of Leadership
  7. The Mediation Model of Leadership
  8. Motivational Climate
- Further Reading

## GLOSSARY

**athletics** Pursuit of excellence characterized by dedication and intensity.

**cohesion** Sticking together to achieve group tasks and satisfy social needs.

**dependence** The extent of interactions among members in the performance of the team task.

**group dynamics** The processes within a group that contribute to its success or failure as well as the scientific study of such processes.

**leadership** The interpersonal process of influencing members toward the achievement of team goals.

**player reaction** The attitudes toward the coach, the team, and the sport cultivated as a function of leader behavior.

**sport** Pursuit of pleasure characterized by moderation and generosity.

**sports team** Two or more individuals sharing a common group identity, facing a common fate, and interacting with each other in interdependent activities in the pursuit of common goals.

**variability** The extent to which the environment changes during the performance of a task.

Group dynamics is a term used to refer to the processes within a group that contribute to its success or failure as well as the scientific study of such processes. In the context of sport, greater emphases have been placed on the processes of cohesion and leadership as they relate to performance and player affective reactions. This article outlines the multidimensional view of both cohesion and leadership and the measurement thereof. In addition, it summarizes research findings on the antecedents and consequences of cohesion and leadership.

## 1. INTRODUCTION

The popularity of sport around the world is evidenced in the economic worth of the sport industry. In the United States alone, the sport industry is estimated to be more than \$160 billion, making it one of the 10 or 12 largest industries in the country. The sport industry is based on the degree to which people are identified with and committed to a sport, a team, and/or an athlete. Such commitment very often translates into heated arguments about the performances of one's team. Most discussions of the successes and failures of a sport team focus on how well one or more athletes performed and on the strategic decisions made by the coach. The debate would also extend to how well members of the

team get along with each other and the rapport between the coach and his or her players. These speculations of the “armchair” experts have been the substance of analysis and critique in group dynamics, a significant scientific endeavor within the broad realm of sport psychology. In this description of the field, the terms “sport groups” and “sport teams” are used interchangeably. It must be noted that group dynamics has been studied in the context of exercise groups as well. However, the treatment of the chosen topics of cohesion and leadership is confined to sport teams as groups.

## 2. DEFINITION OF A SPORT TEAM

In 1998, Carron and Hausenblas defined a sport team as follows:

a collection of two or more individuals who possess a common identity, have common goals and objectives, share a common fate, exhibit structured patterns of interaction and modes of communication, hold common perceptions about group structure, are personally and instrumentally interdependent, reciprocate interpersonal attraction, and consider themselves to be a group.

This definition incorporates the significant attributes of a group highlighted previously by other scholars. One attribute is that the members of the group have a common identity. Although having their unique names, uniforms, coaches, and support groups indicates a separate identity for every team, it is equally important that members accept that identity, share it, and consider themselves to be part of the group. The goals and objectives of sport teams may vary across levels of competition (e.g., winning in pursuit of excellence, learning and having fun in pursuit of pleasure). Another significant attribute is that the group members share a common fate. When a team loses, everyone on the team loses. When a team is disqualified for rule violations committed by some members, all members suffer. Members of a sport team, in sharing a common goal and fate, also have a common understanding of the way in which the team is structured, the manner and mode of communications within the team (including the technical terminology associated with the sport, e.g., what do “screening” in basketball and “stealing” in baseball mean?), and the terms used in the special plays unique to the team. Members also recognize that they are interdependent on each other to achieve the team goals. Given these attributes, members tend to develop a liking for each other (i.e., the interpersonal attraction).

### 2.1. Unique Attributes of Sport Teams as Groups

Sport teams differ from other conventional groups in several respects. One unique attribute of a sport team is the constant roster size as specified by the rules of the game and/or the league (e.g., 12 members on a volleyball team, with 6 players on the court at any one time). It is also possible that a league, or a group of teams in a sport, may jointly decide to reduce the maximum roster size for financial reasons or other exigencies.

Another unique attribute is that most activities of the groups are stipulated and controlled by the rules of the game. For example, only certain positions are permitted to advance the ball in football, use of hands is forbidden in soccer except by the goalkeeper, and holding the ball is not allowed in volleyball. Thus, teams in a given sport are similar in their inputs, their throughput processes are based on shared technology, and their expected outputs are identical.

The league to which a sport team belongs can also regulate the processes to a considerable extent. For example, the National Collegiate Athletic Association (NCAA) stipulates the maximum number of athletic scholarships permissible for an intercollegiate athletic team. It also regulates when a team can start practicing and how many games it can play. In addition, it specifies the scholastic requirements for athletes to be a part of the team. The list goes on, but the point is that sport teams are distinct from other forms of groups due to these extensive external controls over internal processes.

Another important attribute that sets sport teams apart from traditional groups is the presence of clear standards of effort and performance. Members of a sport team recognize that all of them should report to the practice sessions, follow the instructions of the coaches, and work strenuously during every session. It is also clearly understood that during competitions, all players should put forth their best efforts and coordinate their activities in pursuit of victory. In interacting with each other on both task and social spheres, members of a sport team establish and commit themselves to these norms of effort and productivity. In general, these features contribute to the sense of “groupness” among members of a sport team.

Another distinguishing feature of sport teams is the simultaneous presence of competition among the members of the team (particularly in team sports) and the need for cooperation to achieve team success. For instance, because only 5 members of a basketball team can be on the floor at any one time, the team’s 12 members compete

with each other for starting status and/or playing time. At the same time, they cooperate with each other in coordinating their activities for performance success. Successful coaches are those who encourage competition among their players so that each one will try harder to be better than the other players and, at the same time, foster cooperation and coordination among them.

## 2.2. Purposes of Participation in Sport

One criterion for classifying sport teams is the purpose of participation in sport. In 1964, Keating distinguished between “athletics” and “sport.” Sport is derived from the French word *desporter*, meaning a diversionary activity seeking pleasure for both participants. Athletics, derived from the Greek words *athlos* (i.e., a contest), *athlon* (i.e., a prize), and *athlein* (i.e., to contend for a prize), is essentially a competitive activity. In sport, the pleasure and fun can be maximized with a spirit of moderation and generosity. In contrast, athletics is characterized by a spirit of dedication, sacrifice, and intensity. Because the term “athletics” refers to one form of sport, namely track and field in the international context, Keating’s other labels—“pursuit of pleasure” and “pursuit of excellence”—are used to denote these two distinct domains. This distinction parallels the one between recreational sport and competitive sport. Within the preceding broader classification, distinctions among youth sports, adult sports, scholastic and collegiate sports, community sports, and professional sports are also commonly made.

## 3. TASK TYPES IN SPORT

Beyond the distinction between pursuit of excellence and pursuit of pleasure, sport teams do differ in the kinds of tasks in which they are engaged. Two significant attributes have been used to distinguish among sport tasks: dependence and variability.

### 3.1. Dependence

Based on the criterion of dependence, sport tasks are categorized as independent, coactively dependent, proactively–reactively dependent, and interactively dependent. An independent task is one in which an individual can complete the task without interacting with other members of the team (e.g., high jumping,

discuss throwing, individual events in swimming, badminton singles). Individuals in coactively dependent tasks depend on a common but external source for initiation and/or control of their activities. In general, members in a coactively dependent team perform more or less similar tasks, and their collective contributions determine the group’s success (e.g., rowing, synchronized swimming). Reactively–proactively dependent tasks are those in which one team member initiates a task and depends on another team member to complete the task. Conversely, the second person depends on the first person to initiate the task (e.g., football quarterback and receiver, successive runners in a four-person track relay team). Finally, interactively dependent tasks require the team members performing various assignments to interact with each other effectively to complete the group task (e.g., basketball, soccer, hockey).

Sports such as soccer, volleyball, and football are characterized by a high degree of interdependence (i.e., members need to interact with each other during the competitions), whereas sports such as track and field, swimming, and diving are low on this attribute (i.e., performers do not interact with each other in performing their respective tasks). In common parlance, sports from the former set are called “team” sports and those from the latter set are called “individual” sports. Yet coaches, athletes, the public, and the media do conceive of “teams” in individual sports (e.g., the national swimming team, a collegiate track and field team). In general, several athletes in individual sports such as swimming are usually placed under one coach, who tends to organize practice sessions for all athletes at the same specified time and place. During the practice sessions, athletes interact with each other both on a task basis (e.g., competing with each other, setting the pace for each other) and on an interpersonal basis (e.g., encouraging each other, discussing personal matters). From this perspective, athletes on a swimming team are members of a group, and the elements discussed here would be equally applicable to athletes in team sports and those in individual sports.

### 3.2. Variability

Another characteristic distinguishing sports tasks is variability (i.e., the degree of environmental changes requiring participants to adjust to these changes). In high jumping (i.e., a low-variability task), an individual initiates the performance and executes the jump in a relatively stable environment (i.e., the uprights and crossbar are fixed). In basketball rebounding, the

actions are triggered by constantly varying environmental stimuli (e.g., the timing of a shot, the location of the rebound, the location of teammates and opponents).

#### **4. MEASURES OF GROUP PERFORMANCE**

In sport, the performance measures are very often contaminated by factors such as random chance (e.g., the bounce of the ball), an opponent's outstanding performance, strategic choices made by the team or coach, and bad calls made by officials. This section outlines the issues associated with performance measures in sport.

##### **4.1. Subjective Evaluation**

One issue in the measurement of sport performance relates to the subjective evaluation of performance in some sports such as diving, figure skating, and gymnastics. Because of the fallibility of subjective judgments, it is customary to take the averages of the ratings of a specified number of judges to give a score to a performance. Even in sports such as soccer, basketball, hockey, and football, the calls or rulings of referees and officials may be called into question. For example, when two players collide in basketball, a referee may call it a charge (i.e., an offensive foul), whereas a coach may view it as a block (i.e., a defensive foul). The subjective nature of these decisions can provoke heated arguments and catcalls during competitions. The new rules in professional football permit a coach to question a specific call made by an official (e.g., whether a receiver was in-bounds when he caught the ball). The chief referee is bound to look at the official replays and decide whether the official made the correct call or not. It must be noted that the subjective element of a decision may at times be masked by the so-called objective measures. For instance, in football, officials take extraordinary efforts to "bring in the chain" and carefully verify whether the ball crossed the requisite 10 yards for a first down. All of these efforts at objective measurement are based on the subjective judgment of a single referee on where the ball hit the ground when the play ended.

##### **4.2. Zero-Sum Nature of Competitions**

Another issue relates to the zero-sum nature of sports competitions. That is, for every winner, there is a loser.

To alleviate the problem, the win-loss record over an entire season is used as a measure of performance. Another issue is that a basketball team that loses by a score of 100 to 30 is given the same standing as a team that loses by a score of 100 to 99. One way in which to address this issue is to use the point differential (i.e., the difference between points scored for and against the team in a contest) or the ratio of the final score (of the two teams in a contest) as a measure of performance.

##### **4.3. Team Performances in Independent Sports**

Another issue is with the measurement of group performance where members are involved in individual tasks such as archery, badminton, boxing, tennis, golf, swimming, and track and field. For instance, in the international Davis Cup tennis competitions, there are four singles matches and one doubles match. The national team that wins the majority of these five matches is declared the winner. This format of deciding on team performance in individual sports is quite common across all levels of competition.

##### **4.4. Perceptual Measures of Performance**

Given that performance measures may be contaminated by several factors, some scholars have used the perceptions of athlete themselves regarding their personal performance and/or their teams' performance. The argument is that every athlete has set expectations for individual- and team-level performances and is aware of the efforts that he or she, and the team as a whole, had put forth toward achieving those expectations. Therefore, the athletes are the best judges as to whether they had attained their goals. By the same token, they could also be asked about the improvements they had made. These perceptual measures are ideal if the focus of the endeavor is on athletes' development and their affective reactions.

#### **5. COHESION IN SPORT TEAMS**

While the study of group dynamics in sport has gained steam during the past two decades or so, the most notable and sustained efforts have focused on group cohesion. In 1998, Carron and colleagues, incorporating

the thrust of earlier definitions, defined cohesion as “a dynamic process which is reflected in the tendency for a group to stick together and remain united in the pursuit of its instrumental objectives and/or for the satisfaction of member affective needs.” The definition underscores the dual focus of cohesion: task cohesion (i.e., instrumental in achieving group goals) and social cohesion (i.e., satisfying members’ social needs). Groups can be characterized by task cohesion, social cohesion, or both. It is a dynamic process in that the attributes of the members and the group may change over time; thus, the nature and intensity of cohesion may also be in flux. For instance, a newly formed sport team may be marked by task cohesion in the beginning, and continued task interactions among members may lead to interpersonal liking and, thus, to social cohesion.

### 5.1. Multiple Dimensions of Cohesion

A cohesive group is characterized by (a) individual attraction (i.e., an individual values membership in the group) and (b) group integration (i.e., there are perceptions of positive and effective interactions among members). The team would be deemed more cohesive if all members were attracted to the group and if all members perceived the team to be integrated. By crossing the task versus social aspects of cohesion with individual attraction and group integration, Carron and colleagues derived the four-component model of cohesion (Fig. 1), consisting of (a) Attraction to the Group–Task (ATG-T), (b) Attraction to the Group–Social (ATG-S), (c) Group Integration–Task (GI-T), and (d) Group Integration–Social (GI-S).

### 5.2. The Group Environment Questionnaire

From a social cognitive perspective, Carron and colleagues argued that (a) members of a group selectively process the information they gain through experiences of the group’s task and social properties and develop a set of beliefs about the group, (b) these member beliefs about the groups reflect the unity of the group, and (c) these social cognitions of members about the group reflecting its cohesiveness can be measured through a paper-and-pencil questionnaire. Accordingly, they developed a scale to measure the four components of the model. Their theory-driven Group Environment Questionnaire (GEQ) consists of 18 items to measure the four components of cohesion.

	Task Cohesion	Social Cohesion
Attraction to Group	Attraction to Group–Task	Attraction to Group–Social
Integration of Group	Group Integration–Task	Group Integration–Social

**FIGURE 1** Multiple dimensions of cohesion. Adapted, by permission, from A. V. Carron, W. N. Widmeyer, L. R. Brawley, 1985, “The Development of an Instrument to Assess Cohesion in Sports Teams: The Group Environment Questionnaire,” *Journal of Sport Psychology*, 7(3): Figure 2.

### 5.3. Correlates of Cohesion

The following is a summary of Carron and Hausenblas’s 1998 discussion of the research on the relationships of cohesion in sport teams with other relevant variables. Higher perceptions of cohesion were associated with members’ tendency to undertake more of the team activities and to share the responsibility for negative outcomes. When members perceived that the team was cohesive, they tended to attend practices and games more regularly and to put in extra effort during practice sessions. By the same token, higher cohesiveness was related to reduced “social loafing.” Members’ cognitive anxiety was lowered on teams that were perceived to be cohesive.

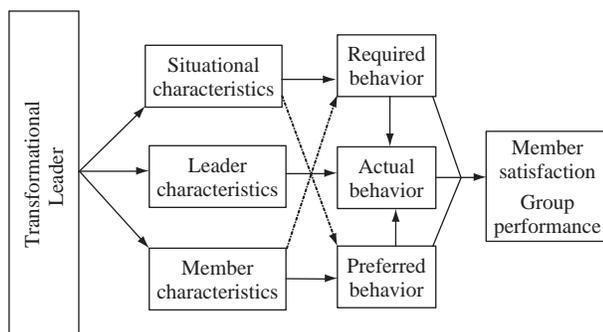
When members perceived the group to be integrated over the task at hand (GI-T), they expressed higher levels of role clarity, role acceptance, and role performance. With increasing task cohesion, members tended to be defensive of the team and to severely self-handicap (i.e., identify external causes for future failures). Perception of cohesion was also related to higher performance by the group. In a 2002 meta-analysis of studies that investigated relationships between cohesion and performance in sport teams, Carron and colleagues found that the significant relationship varied from moderate to large. Contrary to findings relating to groups in general, both task cohesion and social cohesion were significantly related to performance of sport teams. These authors also found that the relationship between cohesion and performance was stronger in women’s teams than in men’s teams. Hence, they suggested

that team-building efforts should focus on cultivating task-oriented attitudes and understandings as well as interpersonal interactions and liking among members. Such a strategy would be particularly effective in the case of women's teams.

A factor that was found to enhance cohesion within a team is size of the group. Small size of the group facilitated higher levels of cohesion. It is noteworthy that sacrifice on the part of members enhanced task and social cohesion, and this in turn led to the conformity to group norms. When leaders emphasized cohesion within the team and executed appropriate actions, cohesion was fostered. Most of the sport teams are characterized by a hierarchical distinction between the leader (i.e., the coach or manager) and the members (i.e., the players). The two major streams of research on leadership in sports are based on the Multidimensional Model of Leadership and the Mediational Model of Leadership.

## 6. THE MULTIDIMENSIONAL MODEL OF LEADERSHIP

Chelladurai's Multidimensional Model of Leadership (Fig. 2) is a synthesis and extension of several theories from mainstream industrial/organizational psychology. In the model, group performance and member satisfaction are considered to be a function of the congruence among three states (or aspects) of leader behavior: required, preferred, and actual. The antecedents of these three states of leader behaviors are the characteristics of the situation, the leader, and the members.



**FIGURE 2** Multidimensional Model of Leadership. From *Handbook of Research on Sport Psychology*, by P. Chelladurai, Macmillan, ©1993 Macmillan. Reprinted by permission of the Gale Group.

### 6.1. States or Aspects of Leader Behavior

Required leader behavior refers to those behaviors shaped by the demands and constraints of the situational characteristics, that is, the parameters of the organization and its environment. The goals and formal organizational structure of the team and the larger system (e.g., professional team vs high school team), the group task and the associated technology (e.g., team sport vs individual sport), social norms, cultural values, and government regulations are some of the situational characteristics that impinge on leader behavior.

Members' preferences for specific leader behaviors, such as coaching and guidance, social support, and feedback, are largely a function of the individual characteristics (e.g., personality variables, age, experience, ability). In addition, the situational characteristics also affect member preferences. If there is an organizational expectation that a leader will behave in a specific manner, this expectation is shared by both the leader and members. Thus, both the leader and members are socialized into the same behavioral expectations and/or preferences in a given context.

A leader's actual behavior is seen as a function of his or her personal characteristics, including personality, ability, and experience. In addition, a leader is considerably influenced by the demands and constraints of the situation as well as by the preferences of the members. As noted previously, the major proposition of the multidimensional model is that the team's performance and member satisfaction are a function of the degree of congruence among the three states of leader behavior.

### 6.2. The Leadership Scale for Sports

In conjunction with proposing the multidimensional model, Chelladurai developed the Leadership Scale for Sports (LSS), which consists of 40 items representing the following five dimensions of leader behavior:

1. Training and instruction (emphasizing and facilitating hard and strenuous training; instructing in the skills, techniques, and tactics of the sport; clarifying the relationship between the members and their tasks; and structuring and coordinating the members' activities)
2. Democratic behavior (allowing greater athlete participation in decisions pertaining to group goals, practice methods, and game tactics and strategies)

3. Autocratic behavior (making decisions independently and stressing personal authority)
4. Social support (being concerned about the welfare of individual athletes, promoting a positive group atmosphere, and inducing warm interpersonal relations with members)
5. Positive feedback (reinforcing athletes by recognizing and rewarding good performance)

### 6.3. Correlates of Leadership

Studies have found that males preferred their coaches to be more training oriented and more autocratic yet more supportive than did female respondents. As for perceived behavior, male athletes perceived their coaches to place greater emphasis on training, instruction, and rewarding behavior and less on democratic behavior, whereas females perceived their coaches to be more democratic and socially supportive. However, there was a tendency at the top levels of competition for the coaches of male and female teams to be similar in their behaviors.

Athletes high on task motivation preferred more training and instruction, whereas those high on affiliation motivation and extrinsic motivation preferred more social support. Athletes high on cognitive structure preferred significantly more training and instruction and less autocratic behavior from the coach, whereas those high on impulsivity preferred more social support behavior from the coach.

With increasing experience, athletes preferred more positive feedback, autocratic behavior, and social support. Preference for social support increased progressively with maturity (i.e., from the high school level to the university level). In general, as athletes gained experience and/or ability, they tended to prefer their coaches to be more autocratic and socially supportive. That leads to the concept of the coach as the benevolent autocrat. Another general trend was for the coaches to rate themselves more positively on these behaviors (i.e., higher on training and instruction, democratic behavior, positive feedback, and social support) than the players did.

As for the situational variables, members of competitive sport teams preferred greater training and instruction, greater social support, less positive feedback, and less democratic behavior from their coaches than did members of recreational sport teams. As for task attributes, athletes involved in interdependent tasks (i.e., team sports) or variable tasks (i.e., open sports such as basketball) preferred more training and instruction and autocratic behavior than did athletes involved in independent

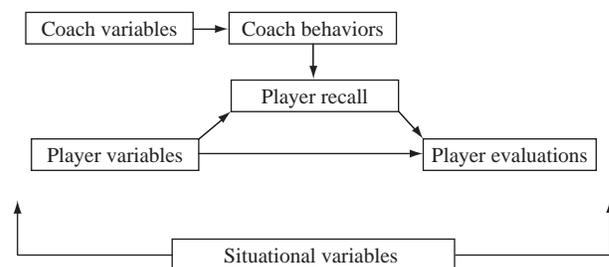
tasks (i.e., individual sports) or nonvariable tasks (closed sports such as swimming). In general, increased task dependence and/or task variability seems to warrant the need for more training and instruction, autocratic behavior, social support, and positive feedback.

With regard to outcomes, athletes were satisfied with leadership to the extent that the coach emphasized (a) training and instruction that enhances the ability and coordinated effort by members that, in turn, contributes to task accomplishment, and (b) positive feedback that recognizes and rewards good performance. Performance as a consequence of leadership has been dealt with inadequately in the literature. The few studies indicate that higher perceptions of training, autocratic behavior, social support, and positive feedback were associated with success in athletics.

The relationships between leadership and performance were quite weak. Some scholars have used player perceptions as a measure of performance, and such perceptions may relate to performance and/or performance improvements. In contrast to performance, satisfaction has been studied more extensively. In 1998, Riemer and Chelladurai developed the Athlete Satisfaction Questionnaire (ASQ), consisting of 56 items to measure 15 facets of athlete satisfaction, including those relating to individual and team performance outcomes, individual and team social outcomes, individual and team task and social processes, and external agents.

## 7. THE MEDIATIONAL MODEL OF LEADERSHIP

The Mediation Model of Leadership (Fig. 3) was developed by Smoll and Smith with a focus on youth sport and



**FIGURE 3** Mediation Model of Leadership. Reprinted with permission from *Journal of Applied Social Psychology*, Vol. 19, No. 18, pp. 1522–1551. ©V. H. Winston & Son, Inc., 360 South Ocean Boulevard, Palm Beach, FL 33480. All rights reserved.

emphasizes players' evaluative reactions as the outcome measures. The three critical elements of the model are coaches' behaviors, players' perceptions and recall, and players' evaluative reactions. Players' attitudes toward their coaches and their sport experience are mediated by their perceptions and recall of the coaches' behaviors. These three elements are affected by three sets of factors: coach individual difference variables, player individual difference variables, and situational factors.

### **7.1. Factors Influencing Leadership Dynamics**

The coach individual difference factors include (a) coaches' goals, (b) coaches' behavioral intentions (i.e., the antecedents of actual behavior), (c) instrumentalities (i.e., a function of the perceived probability of an outcome and the value attached to that outcome), (d) the norms associated with the coaching role, (e) coaches' perceptions of players' motives, (f) self-monitoring (i.e., the extent to which coaches analyzes their own behaviors and their consequences), and (g) coaches' sex. Player individual difference variables include (a) the age that has been shown to affect players' perceptions and evaluations of their coaches, (b) players' sex, (c) players' perceptions of coaching norms, (d) the valence that players attach to various coaching behaviors, (e) players' achievement motive in the sporting context, (f) competitive trait anxiety, (g) general self-esteem (i.e., players' evaluation of general self-worth), and (h) athletic self-esteem (i.e., players' evaluations of themselves as athletes). The situational factors affect coaches' behaviors as well as players' perceptions of those behaviors and reactions to them. These factors include (a) nature of the sport, (b) level of competition (e.g., competitive sport vs recreational sport), (c) practice sessions versus game sessions, (d) previous success/failure record of the team, (e) current game/practice developments (e.g., whether the team is losing or winning), and (f) intrateam attraction.

### **7.2. The Coaching Behavior Assessment System**

Smoll and Smith employed the observational method to assess the actual leader behavior and employed the paper-and-pencil method to assess players' perceptions and recall of that leader behavior as well as coaches' perceptions of their own behaviors. These authors'

Coaching Behavior Assessment System (CBAS) was developed over several years by observing and recording the behaviors of youth soccer basketball and baseball coaches, analyzing these recordings, and categorizing the relevant leader behaviors. The 12 behavioral dimensions measured by the CBAS are broadly classified into reactive and spontaneous behaviors. Reactive behaviors are coaches' responses to (a) desirable performance or effort, (b) mistakes and errors, and (c) misbehaviors. Spontaneous behaviors may be (a) game related or (b) game irrelevant.

The training of the observers has been quite extensive, including a programmed learning module, group instruction in use of the coding system, written tests, practice, and reliability checks. Such training resulted in the coders agreeing with the experts more than 90% of the time. The second coding of the same coaching behaviors after a lapse of 1 week was also consistent to the tune of 87.5% to 100%, with a mean of 96.4%.

To assess players' perceptions and recall of coaches' behaviors, players were provided with a description and an example of each of the 12 behavioral dimensions of the CBAS and were asked to indicate the extent to which their coaches had engaged in each of the behaviors. A similar format was followed in assessing coaches' perceptions of their own behaviors. The relationships among the three measures of coaching behavior—observation, coaches' self-reports, and players' perceptions and recall—were not strong. More specifically, coaches' self-reports of their behaviors did not bear much resemblance to the assessments made by either observers or players. Therefore, the use of coaches' self-reports of their behaviors should be viewed with extreme caution. Player evaluative reactions were scored with a 10-item scale measuring attraction of the coaches and attraction toward teammates.

### **7.3. Correlates of Leadership**

In general, leaders' supportiveness (i.e., reinforcement and mistake-contingent encouragement) and instructiveness (i.e., general technical instruction and mistake-contingent instruction) were significantly and positively related to players' attitudes toward their coaches, the sport, and teammates, whereas punishment and punitive technical instruction had the opposite effect. Also, coaching behaviors affected players' liking for the sport after the season, their perceptions of team solidarity, their evaluation of their coaches, and their

self-esteem. Although general technical instruction was found to be more potent than the other categories, positive reinforcement had minimal effects.

Player self-esteem has been used both as (a) a moderator of the relationship between coaches' behaviors and player attitudes and as (b) an outcome measure. Children with low self-esteem tended to respond most positively to a supportive or an instructive coach-player relationship and tended to react most negatively to coaches who were not supportive or instructive. It was also found that coaching behaviors had far less impact on children with high self-esteem. As for self-esteem as an outcome, children playing for trained coaches exhibited higher levels of self-esteem than they had a year before. Such increases were not found among children playing for nontrained coaches. Coaching behaviors during practices were also shown to be significantly associated with changes in players' perceptions of their physical and cognitive competence as well as success expectancy.

The mediational model has been the foundation for training coaches to improve their behaviors, assess their behaviors, and measure the effects of changes in coaching behaviors on players' evaluative reactions. Research has shown that the trained group of coaches differed from the control group in both observed and perceived behaviors. That is, the former group engaged in behaviors consistent with the training program. Furthermore, the players of the trained coaches evaluated their coaches more positively and expressed higher levels of intrateam attraction than did the players of the control group of coaches. However, the two groups did not differ in performance as measured by win/loss records.

#### **7.4. Commentary on Leadership Studies in Sport**

Some hold that the Multidimensional Model of Leadership is meant for adult sport and that the Mediational Model of Leadership is meant for youth sport because these models were developed in the context of adult sport and youth sport, respectively, and have been employed predominantly in adult sport and youth sport, respectively. But there is nothing in either of the models to prohibit or inhibit its use in either level of sport. Although there are some conceptual differences between the two models, each of them is applicable to either level.

The distinguishing feature of the research carried out by Smoll and Smith is the manner in which leader

behavior is measured, that is, the observational method of the CBAS versus the paper-and-pencil LSS. However, a comparison of leadership behavior categories of the two systems shows that they are, to a large extent, similar to each other in content, although the labels are different. For example, the reinforcement and non-reinforcement categories of the CBAS resemble the LSS dimension of positive feedback; the mistake-contingent technical instruction, general technical instruction, and organization are subsumed by the LSS dimension of training and instruction; and general communication is similar to social support in the LSS. Despite these similarities, the CBAS is distinct in two significant respects. First, the larger number of categories in the CBAS permits a more thorough analysis of leadership in youth sport. Apart from this advantage, the CBAS also measures leader behavior in more specific situations. For example, although "encouragement" in the CBAS is subsumed by the "positive feedback" in the LSS, encouragement is divided into reactive mistake-contingent behavior and spontaneous general encouragement.

In 2002, Horn employed a working model as the framework for her review of coaching effectiveness. While incorporating significant elements of both the multidimensional and mediational models of leadership, she also introduced the construct of "coaches' expectancies, values, beliefs, and goals" as a factor mediating the effects of the "sociocultural context," "organizational climate," and "coaches' personal characteristics" on "coaches' behavior." Furthermore, she referred to self-efficacy, motivational style (i.e., intrinsic or extrinsic motivation), self-reflectiveness, critical thinking aptitudes, locus of control, and trait anxiety as possible predictors of coaches' expectancies, values, beliefs, and goals. Coaches' expectations referred to coaches developing expectations for the performance of their teams and their members that affects coaches' subsequent behaviors. Beliefs were narrowed down to stereotyped beliefs about gender, race and ethnicity, and sexual orientation.

## **8. MOTIVATIONAL CLIMATE**

An emergent thrust in the study of group dynamics pertains to the motivational climate within sport groups created and fostered by coaches. The motivational climate is characterized as (a) task involving (i.e., mastery orientation) or (b) ego involving (i.e., performance orientation). In the climate of mastery orientation, hard work, trying hard, skill improvement, and contributions by all members are emphasized. In

contrast, comparisons with others and performing better than others are promoted in the performance-oriented climate. Research has shown that, in general, the mastery-oriented (task-involving) climate enhances athletes' level of enjoyment, satisfaction, and interest in the sport, whereas the performance-oriented (ego-involving) climate tends to increase tension, pressure, and performance anxiety. Insofar as coaches have considerable influence and control over the climate in which their teams operate, it could be speculated that coaches' own goal orientation would be reflected in their leadership behaviors, which in turn would create either the mastery-oriented climate or the performance-oriented climate. It is also possible that the organization itself may foster the performance orientation that impinges on both coaches and athletes. For example, coaches of Division I schools in the National Collegiate Athletic Association (NCAA) are likely to exhibit those behaviors that would foster a performance-oriented climate, and athletes are likely to endorse that orientation and flourish in it. The relative influences of the organizational goals and coaches' own goal orientation on the climate experienced by team members are yet to be determined. Finally, the idea that pursuit of excellence would be greatly facilitated if the climate reflected both task orientation and ego orientation is also worthy of investigation.

### See Also the Following Articles

Goal Setting and Achievement Motivation in Sport  
 ■ Leadership and Culture ■ Power, Authority, and Leadership

### Further Reading

- Carron, A. V., Brawley, L. R., & Widmeyer, W. N. (1998). The measurement of cohesiveness in sport groups. In J. L. Duda (Ed.), *Advancements in sport and exercise psychology measurement* (pp. 213–226). Morgantown, WV: Fitness Information Technology.
- Carron, A. V., Colman, M. M., Wheeler, J., & Stevens, D. (2002). Cohesion and performance: A meta-analysis. *Journal of Sport and Exercise Psychology*, *24*, 168–188.
- Carron, A. V., & Hausenblas, H. A. (1998). *Group dynamics in sport*. Morgantown, WV: Fitness Information Technology.
- Chelladurai, P. (1993). Leadership. In R. N. Singer, M. Murphy, & K. Tennant (Eds.), *The handbook on research in sport psychology* (pp. 647–671). New York: Macmillan.
- Chelladurai, P., & Saleh, S. D. (1980). Dimensions of leader behavior in sports: Development of a leadership scale. *Journal of Sport Psychology*, *2*(1), 34–45.
- Duda, J. L., & Hall, H. (2001). Achievement goal theory in sport: Recent extensions and future directions. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Hand book of sport psychology* (2nd ed., pp. 417–443). New York: John Wiley.
- Horn, T. S. (2002). Coaching effectiveness in the sport domain. In T. S. Horn (Eds.), *Advances in sport psychology* (2nd ed., pp. 309–354). Champaign, IL: Human Kinetics.
- Riener, H. A., & Chelladurai, P. (1998). Development of the Athlete Satisfaction Questionnaire (ASQ). *Journal of Sport and Exercise Psychology*, *20*, 127–156.
- Smoll, F. L., & Smith, R. E. (1989). Leadership behaviors in sport: A theoretical model and research paradigm. *Journal of Applied Social Psychology*, *19*, 1522–1551.
- Widmeyer, W. N., Brawley, L. R., & Carron, A. V. (2002). Group dynamics in sport. In T. S. Horn (Ed.), *Advances in sport psychology* (2nd ed., pp. 285–308). Champaign, IL: Human Kinetics.



# Groups, Productivity within

Athena Xenikou

Aristotle University of Thessaloniki, Thessaloniki, Greece

Alexandra Hantzi

Panteion University, Athens, Greece

1. Introduction
  2. Social Facilitation/Inhibition
  3. Steiner's Theory of Group Productivity
  4. Social Loafing
  5. Motivation Gains in Groups
  6. Implications for Groups at Work
- Further Reading

Research on group performance mainly examines the conditions under which motivation losses and gains are produced when people work in groups. Motivation losses concern the reduction in individual effort to perform a task due to the fact that one works as part of a group. On the other hand, it has been suggested that motivation gains resulting from group settings need to be more thoroughly investigated.

## GLOSSARY

**free-rider effect** The enjoyment of the benefits resulting from successful group performance without contributing to it.

**production blocking** Production blocking in brainstorming groups arises from the constraint that only one group member can talk at a time.

**social compensation** A group member compensating for the low performance of other group members for the desired collective goal to be accomplished.

**social facilitation/inhibition effects** The facilitation or impairment of individual task performance as a function of the presence of others either as an audience or as coactors.

**social loafing** The decrease in individual effort when one performs as a member of a group in comparison to performing alone.

**sucker effect** The expectation that others withhold efforts in performance groups, which can lead an individual reducing his or her contribution to prevent being a victim of free-riding.

**transactive memory** A knowledge shared by group members about who is good at what.

## 1. INTRODUCTION

Most people have at some point wondered if it is more productive to work as part of a team or alone. This is a simple but challenging question, which social psychology has faced since it appeared as a discipline. Researchers have worked hard for decades to answer whether group performance is superior to individual performance. In general, the findings have shown that groups are more effective than separate individuals because groups learn faster, produce a greater quantity and quality of work, and make fewer errors.

However, groups do not accomplish their full potential since their performance is not as high as the sum-total of the maximum individual performance of each group member. It has been suggested that coordination losses and motivation losses are responsible for groups performing less efficiently than expected. Researchers have also identified conditions under which motivation gains result out of working as part of a group. Social compensation refers to the increased effort of a group

member in order to compensate for the insufficient contribution of another member when the group product is perceived as valuable. Moreover, there is evidence that complex and self-involving tasks carried out by groups that are psychologically meaningful to their members do not lead to productivity losses but gains.

## **2. SOCIAL FACILITATION/ INHIBITION**

The most elementary issue concerning the impact of social factors on individual task performance is whether the mere presence of others influences individual performance. Social facilitation and inhibition (SFI) refer to the positive and negative effects that audiences (as well as coactors) may have on individual task performance. Interest in how the presence of others affects individual performance dates back to 1898 when Tripplett conducted one of the first experiments in the field of social psychology. Tripplett had children spin fishing reels either alone or in coactive pairs and found that participants performed better coactively. Following Tripplett's work, many researchers replicated the finding that the presence of others facilitates individual performance, whereas others demonstrated the social inhibition of individual performance under the same conditions.

In 1965, these apparently inconsistent results were explained by Robert Zajonc, who proposed that the presence of others facilitates individual performance in the case of simple tasks, whereas it impairs performance at complex tasks. Zajonc put forward a drive theory of social facilitation suggesting that the mere presence of others increases generalized drive or physiological arousal. Arousal, in turn, leads to the generation of dominant responses—that is, responses that are well learned, habitual, or innate. Simple tasks are successfully performed when dominant responses are generated, whereas complex tasks being more demanding in nature cannot be accomplished by the emission of dominant responses.

A number of other theoretical accounts of how task complexity affects individual performance at social settings have been provided. Zajonc had such a great impact that some of these theoretical accounts did not challenge the drive approach of his theory and presented factors considered as determinants of drive/arousal. For example, (i) people's concern with evaluation and (ii) distraction/conflict caused by attending both to the task at hand and to other people were factors proposed by different theoretical approaches to account for elevated

arousal at social settings. Recent theories have challenged the biologically based explanation offered by Zajonc's theory and have focused on cognitive and attentional factors as determinants of SFI effects.

There is a strong empirical basis suggesting that task complexity is related to the manifestation of SFI effects as proposed by Zajonc. However, it has been found that physiological arousal is not increased by the presence of others irrespective of task complexity. Complex tasks do lead to an elevated drive when others are present, whereas simple tasks do not seem to provoke increased levels of physiological arousal. Finally, recent research, in accordance with an attentional view of social facilitation, has shown that the presence of unpredictable others facilitated performance at a complex task by the emission of a nondominant response. In summary, Zajonc's proposition that task complexity plays an important role in explaining SFI effects has been strongly supported by the findings, but the exact mechanism(s) leading to facilitation and impairment at simple and complex tasks, respectively, has not been identified.

## **3. STEINER'S THEORY OF GROUP PRODUCTIVITY**

Although the literature on social facilitation suggests that under certain conditions one feature of group settings (i.e., the presence of others) enhances individual performance, most of the work on group performance argues that when people work together in groups they do not perform at the maximum level of their potential. In 1972, Steiner proposed an influential analysis of group productivity, which focuses on the notion of a group's potential productivity, the maximum possible performance level that the group can achieve. Potential productivity is determined by group members' resources and task demands. Member resources concern the capabilities, knowledge, and skills of group members that can be useful at completing the task at hand. Task demands mainly refer to the criterion that is used to estimate a group's performance level. The criterion of performance at a weight-lifting task is the total weight lifted, whereas in the case of a brainstorming task it is the number of novel ideas generated by the group.

Steiner's model assumes that a group is not in a position to achieve its potential productivity due to process loss (faulty group process), and that a group's actual productivity is potential productivity minus losses due to faulty process. Process loss refers to coordination as

well as motivation losses. When performing in groups it is necessary for group members to coordinate their actions to achieve the desired group goal. For example, in a rope-pulling task group members have to synchronize their actions to achieve their full potential. This need for coordination inevitably leads to loss in productivity since group members cannot pull at exactly the same time or direction. Regarding motivation losses, individuals are considered to exert less effort when working as part of a team than when working alone.

Steiner also provided a task classification or taxonomy (Table 1), which aimed at further assisting researchers in the investigation of group productivity. According to Steiner's classification, there are three task characteristics that can be used to categorize different tasks. The first task feature is whether a task can be subdivided to separate parts performed by different people. The tasks that cannot be subdivided are unitary, whereas the tasks that can be subdivided are divisible. An example of a unitary task is finding a solution to a mathematical problem. On the other

hand, a task that certainly (and hopefully) can be divisible among roommates is cleaning their apartment. The second task feature that Steiner used to classify tasks is the performance criterion mentioned earlier, leading to a distinction between maximizing and optimizing tasks. When a task performance criterion concerns the quantity or speed of production, the task is considered to be a maximizing task; the rope-pulling task is a good example of a maximizing task. On the other hand, optimizing tasks put a demand for a correct or optimal solution (e.g., a mathematical problem). Finally, the way that group members' inputs are related to the group product leads to four different types of tasks: disjunctive, conjunctive, additive, and discretionary tasks. In a disjunctive task, members make separate contributions and the best contribution must be chosen (e.g., a reasoning problem). In conjunctive tasks, group performance is determined by the contribution of the least capable member (e.g., climbing a mountain while members are tied to one another). In additive tasks, the individual contributions are simply added together (e.g., a rope-pulling task). Finally, in discretionary tasks, individual inputs are combined in any way the group chooses (e.g., roommates cleaning their apartment).

Steiner's model has been criticized on the basis that it explicitly makes the assumption that groups cannot reach their full potential. It has been argued that groups might be in a position not only to achieve performance at the highest level of their potential but also to overperform it. Steiner's theory has put emphasis on coordination and motivation losses that influence group performance, while discounting the possibility that groups may exhibit motivation gains.

#### 4. SOCIAL LOAFING

Productivity loss in group settings has been defined in terms of both coordination and motivation losses. The first experiments on group productivity were conducted during the last decades of the 19th century by Ringelmann. In these experiments, individuals were asked to pull a rope either alone or in groups of various sizes. The results showed that when individuals worked in groups, the collective output was less than the sum total of the individual performance measured when individuals pulled the rope alone. This finding can be explained as the result of coordination loss. However, subsequent research showed that this finding

TABLE 1  
Steiner's Task Classification<sup>a</sup>

<i>Task characteristics</i>	<i>Task taxonomy</i>
Can a task be subdivided to separate parts performed by different people?	Yes → <i>Divisible tasks</i> No → <i>Unitary tasks</i>
Does the task performance criterion concern quantity/speed of production or a correct/optimal solution?	The performance criterion concerns Quantity/speed of production → <i>Maximizing tasks</i> Correct/optimal solution → <i>Optimizing tasks</i>
How are group members' inputs related to group product?	Members make separate contributions and the best contribution must be chosen → <i>Disjunctive tasks</i> Group performance is determined by the contribution of the least capable member → <i>Conjunctive tasks</i> The individual contributions are simply added together → <i>Additive tasks</i>

<sup>a</sup>Adapted from Steiner (1972).

is caused by the joint effect of coordination problems and reduced effort exerted by group members.

Regarding motivation losses, there is a difference between the effort that individuals exert when working collectively and working coactively (or individually). When people work collectively, their individual contributions are combined to form a group product, whereas when they work coactively their individual inputs are not combined to a single product. Individuals' tendency to produce lower effort when working collectively compared to working coactively (or individually) was called social loafing by Latané and colleagues in 1979.

Social loafing has been observed in a wide variety of tasks, including physical tasks (e.g., rope pulling, shouting, and clapping), cognitive tasks (e.g., idea-generation tasks), perceptual tasks (e.g., maze performance), and evaluative tasks (e.g., evaluation of poems and editorials), suggesting that social loafing is a robust phenomenon. As a general phenomenon not restricted to physical tasks, it poses the important practical question of why and when less effort is exerted in group settings.

A number of theoretical explanations have been proposed to explain why social loafing takes place in groups. It has been suggested that social loafing occurs because (i) the pressure people feel as group members to accomplish the task at hand is less intense than the pressure one feels when working alone, (ii) the presence of other group members is drive reducing because others share the responsibility for carrying out the task, (iii) there is reduced identifiability or evaluation potential in collective action compared to individual action, and (iv) group members feel their contributions to the group product not to be essential but dispensable.

Apart from social loafing, two other effects have been reported, namely, the free-rider effect and the sucker effect. The free-rider effect concerns the enjoyment of the benefits resulting from the successful group performance without contributing to it. The free-rider effect might be seen as the result of a person's conscious decision to indulge in social loafing. On the other hand, members of a group who work hard to accomplish the group goal run the risk that others may free ride on their efforts and become suckers. The expectation that others withhold efforts in performance groups can lead to the individual reducing his or her contribution to the group in order to prevent being a victim of free-riding (the sucker effect).

The literature that examines group conditions leading to social loafing has also attempted to propose various solutions to remedy motivation loss in group settings.

On a practical level, the following can reduce or overcome social loafing in real-world contexts: (i) identification and evaluation of individual inputs, (ii) meaningful and self-involving tasks, (iii) close ties among group members, (iv) small rather than large groups, (v) uniqueness of individual contributions, and (vi) feedback on individual or group performance.

## 5. MOTIVATION GAINS IN GROUPS

Even though research on group performance has concentrated on productivity loss in group settings due to reduced motivation, there are findings showing that motivation gains can also result from performing within groups. In 1991, Williams and Karau demonstrated conditions under which people may actually increase their efforts when working collectively compared to when working individually (coactively). They call this effect social compensation. Social compensation refers to a group member compensating for the low performance of other members in order to accomplish the desired group goal.

One factor that might produce social compensation is the expectation that other group members are performing insufficiently. The perception of inadequate coworker inputs may be derived from a general lack of trust in the reliability of others to perform well when their contributions are pooled with those of others or/and direct knowledge of coworkers' insufficient efforts or inabilities. The second factor that may be necessary to produce social compensation is that the group product is in some way important to the individual. If the group output is not meaningful and important to group members, then one does not expect social compensation to occur.

Social compensation is not expected to occur in all situations in which individuals believe their coworkers are contributing less than their fair share. Besides group product meaningfulness previously mentioned, other factors are important for compensation to take place. Social compensation is more likely to occur (i) when the individual has to remain in the group working collectively, (ii) at earlier stages of the collective effort, and (iii) when the group size is relatively small. Finally, people may be motivated to compensate in some cases for their own prior underperformance rather than the underperformance of other group members.

According to social compensation, when group members of unequal strength work on an important task, the most competent member will compensate for the insufficient contribution of the other group

member(s). There are, however, findings showing that under specific conditions the weaker member of the group tries hard to match the performance level of the stronger member (the Köhler effect). One necessary condition for this effect to occur is that the group product must be meaningful to its members. The second condition is that the discrepancy in ability between the group members is of a moderate degree. A high discrepancy in individual performance is not expected to lead to more effort exerted by the weakest members because people believe that there is a limit to the discrepancy in ability that can be compensated for through increased effort.

The social loafing paradigm has been criticized on the basis that little effort is expended in making the participants feel that they are part of a group. In the paradigm, a number of individuals combine their efforts for a group output without interacting with each other, competing with other groups, or expecting any form of future interaction. Even though the findings of research on social loafing are discussed in terms of group membership, it seems that participants mainly act on the basis of their personal identity rather than a group-based sense of self. Motivation losses described by research on social loafing might be the result not of faulty group processes but of individuals acting as persons rather than group members. It has been demonstrated that when social identity is made salient there are motivation gains leading to the accomplishment of the desired goal as defined by the group rather than some external source of influence.

## 6. IMPLICATIONS FOR GROUPS AT WORK

Despite the fact that most research on group performance has emphasized the inefficiency of groups compared to individuals, groups feature in most essential aspects of everyday life. Their central role in real-world contexts might have to do with the fact that many tasks cannot be accomplished by individuals. If one considers how many people with different skills are involved in making a car or building a house, then it becomes clear that getting rid of groups cannot be a practical solution to increase performance.

Regarding groups at work, how can we use the accumulated knowledge on group performance to understand how groups function and to propose ways on a practical level for groups to become more effective?

In organizational settings, brainstorming groups are used in order for the organization to stimulate employees to think creatively. In brainstorming groups, members are asked to generate as many ideas as possible, not to criticize others' ideas, and to build on ideas presented by other group members. Research by Diehl and Stroebe in the late 1980s showed that brainstorming groups produce half the ideas produced by the same number of individuals who brainstorm alone (nominal groups). This productivity loss in brainstorming groups has been discussed as a coordination loss resulting from mutual production blocking. Mutual production blocking arises from the constraint that only one group member can talk at a time in a brainstorming group.

Although there is ample evidence showing brainstorming groups to be less effective than nominal groups, organizations of all kinds make wide use of brainstorming groups. A reason that might explain organizational members' confidence in brainstorming groups is that members of such groups are more satisfied with their performance in comparison to individuals who brainstorm alone. This discrepancy between satisfaction and productivity can be accounted for by the greater ease of idea generation in brainstorming groups compared to nominal groups. Since brainstorming is widely used at work, some advice needs to be given on the conditions that make brainstorming groups more effective. First, the size of the group should be kept as small as possible, not exceeding four members. Second, because time pressure negatively affects the productivity of brainstorming groups, groups should decide when it is the best time to stop. Finally, it is advisable that group members are heterogeneous in their knowledge about the brainstorming topic so that members can have a stimulating impact on each other.

Moreover, our knowledge on group productivity can be applied to training at work in order to improve organizational effectiveness. Interventions designed to increase productivity through learning, such as training programs, were found to have a strong positive effect. Regarding group training programs, it has been suggested that group learning is associated with a transactive memory system—that is, a knowledge shared by group members on who is good at what. When members of a work group are trained together, they develop complementary specializations that enable the group to share knowledge and skills. In this way, the group as a whole makes use of social compensation—that is, individual members compensate for each other's inadequacy. However, a practical problem related to group training at work involves the

disruption that employee turnover has on the smooth operation of the work group.

The business world has acknowledged the importance of one's ability to work within groups as a key factor to success. In this sense, groups are seen as capable of accomplishing more than the sum of their individual parts. However, the bulk of research on group performance suggests that, more often than not, groups are not as efficient as individuals.

### See Also the Following Articles

Conflict within Organizations ■ Groupthink ■ Intergroup Relations and Culture ■ Interpersonal Conflict

### Further Reading

- Baron, R. S., & Kerr, N. L. (2003). *Group process, group decision, group action*. Buckingham, UK: Open University Press.
- Brown, R. (2000). *Group processes: Dynamics within and between groups*. Oxford, UK: Blackwell.
- Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53, 497–509.
- Haslam, A. (2001). *Psychology in organizations: The social identity approach*. London: Sage.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37, 822–832.
- Ringelmann, M. (1913). Recherches sur les moteurs animés: Travail de l'homme [Research on animate sources of power: The work of man]. *Annales de l'Institut National Agronomique*, 12, 1–40.
- Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.
- Stroebe, W., Diehl, M., & Abakoumkin, G. (1996). Social compensation and the Kihler effect: Toward a theoretical explanation of motivation gains in group productivity. In E. Witte, & J. H. Davis (Eds.), *Understanding group behavior: Small group processes and interpersonal relations* (pp. 37–65). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tripplett, N. (1898). The dynamogenetic factors in pacemaking and competition. *American Journal of Psychology*, 9, 507–533.
- Williams, K. D., & Karau, S. J. (1991). Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology*, 61, 570–581.
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149, 269–274.



# Groupthink

Ramon J. Aldag

University of Wisconsin, Madison, Wisconsin, USA

Sally Riggs Fuller

University of South Florida, Tampa, Florida, USA

1. The Groupthink Model Defined and Explained
  2. History and Impact of the Groupthink Phenomenon
  3. Evidence Regarding the Groupthink Phenomenon
  4. Bases for Groupthink's Appeal
  5. Implications and Suggestions for Future Research and Theory
  6. Prescriptions for Effective Group Problem Solving
- Further Reading

## GLOSSARY

**availability heuristic** Assessment of the frequency of a class or the probability of an event based on the ease with which instances or occurrences can be brought to mind.

**defensive avoidance** A condition in which an individual fails to engage in vigilant search and appraisal and instead adopts a dominant mode of coping that involves avoidance of cues that stimulate anxiety or other painful feelings.

**groupthink** A condition in which highly cohesive groups in "hot" decision situations display excessive levels of concurrence seeking that suppress critical inquiry and result in faulty decision making.

**"hot" cognitions** Thought processes regarding vital, affect-laden issues (i.e., in "hot" decision situations).

**illusory correlation** The tendency to see relations between variables where none exists.

**mindguard** In groupthink, a self-appointed individual who acts to shield a group from adverse information.

**Pygmalion effect** Self-fulfilling prophecy; it is the tendency to create something in the image that one has of it.

**strong version of groupthink** The view that groupthink characteristics cluster due to their common ties to specific

antecedent conditions and are linked in a causal chain from those antecedent conditions, to symptoms, to defects, to outcomes.

**weak version of groupthink** The view that groupthink may be evidenced by the presence of some subset of the characteristics proposed by Janis and that the causal ordering that he posited may be suggestive rather than necessary.

Groupthink is a condition in which highly cohesive groups in "hot" decision situations display excessive levels of concurrence seeking that suppress critical inquiry and result in faulty decision making. In the groupthink process, antecedent conditions are seen as resulting in a variety of symptoms of groupthink that lead to defective decision making and, in turn, to flawed decision outcomes. This article addresses the nature and history of the phenomenon, offers evidence of its validity and explanations for its widespread acceptance, and provides implications and suggestions.

## 1. THE GROUPTHINK MODEL DEFINED AND EXPLAINED

Janis proposed that highly cohesive groups are likely to suffer from groupthink, a strong concurrence-seeking tendency that suppresses critical inquiry and results in faulty decision-making processes and flawed outcomes. Janis discussed major historical fiascoes such as the lack of preparedness for the Japanese attack on Pearl Harbor,

the escalation of war in Korea, the failed U.S.-sponsored landing of anti-Castro rebels in the Bay of Pigs, and escalation of U.S. involvement in the war in Vietnam. First presented in a 1971 issue of *Psychology Today*, this groupthink phenomenon quickly gained remarkably broad and firm acceptance, dominating the literature on group decision making for decades.

At the individual level, Janis and Mann argued that thinking about vital, affect-laden issues generally results in “hot” cognitions, in contrast to the “cold” cognitions of routine problem solving, and that such hot cognitions can result in errors in scanning of alternatives. Such situations induce stress that results in defensive avoidance, characterized by lack of vigilant search, distortion of the meanings of warning messages, selective inattention and forgetting, and rationalizing. Groupthink is seen as the group analog of defensive avoidance or “a collective pattern of defensive avoidance.” In 1972, Janis wrote that he used the term “groupthink” to refer to a mode of thinking engaged in by people in cohesive in-groups, with members’ strivings for unanimity overriding their motivation to submit alternative courses of action to realistic appraisal. He further stated, “Groupthink refers to a deterioration of mental efficiency, reality testing, and moral judgment that results from in-group pressures.”

Janis chose the term “groupthink” due to its frankly Orwellian connotation, similar to “doublethink” and “crimethink.” In 1982, Janis wrote that “the invidiousness is intentional.” This view of groupthink as an undesirable phenomenon continued to be evidenced in Janis’s more recent work.

Janis explicitly defined the context in which groupthink is thought to occur and divided these antecedent conditions into three categories. First, a moderate to high level of group cohesion is a necessary, but not sufficient, condition for groupthink. In addition, structural faults and a provocative situational context are secondary antecedents that define situations in which groupthink is likely to occur. In the structural fault category are insulation of the group, lack of impartial leadership, lack of norms requiring methodical procedures, and homogeneity of group members’ social backgrounds and ideologies. The provocative situational context antecedents focus on the role of stress as a situational factor. The first stress factor is characterized as external threats of losses combined with a low hope of finding a better solution than that of the leader. The internal stress antecedent stems from temporary low self-esteem attributable to members’ recent failures, perceptions that the task is too difficult to

accomplish, and the perception that there is no alternative that is morally correct.

According to Janis, the antecedents lead to several proposed symptoms of groupthink, including an illusion of invulnerability, rationalization to discount warnings and other negative feedback, a belief in the inherent morality of the group, stereotyped views of members of opposing groups, pressure on dissenters, self-censorship, illusion of unanimity, and self-appointed “mindguards” acting to shield the group from adverse information.

According to Janis, groupthink results in a number of consequences that interfere with effective group decision making. For instance, the group limits its discussion to only a few alternatives. After a course of action is initially selected, group members ignore new information concerning its risks and drawbacks. At the same time, they avoid information concerning the benefits of rejected alternatives. Members make little attempt to use experts to obtain more precise information. In addition, because they are so confident that things will turn out well, members fail to consider what may go wrong and, as such, do not develop contingency plans. These “defects” are seen as leading to impaired performance and other undesirable outcomes.

Janis suggested methods to prevent or minimize the supposedly dysfunctional consequences of groupthink. These “remedies” included the following:

- Group leaders should encourage all group members to air their doubts and objections.
- Group leaders should adopt an impartial stance rather than initially stating their preferences.
- Group members should be encouraged to discuss the group’s deliberations with trusted associates and report their reactions back to the group.
- Outside experts should be invited to meetings and encouraged to challenge group members’ views.
- When a competitor is involved, time should be devoted to assessment of warning signals from the competitor and of alternative scenarios of the competitor’s intentions.
- When considering alternatives, the group should split into subgroups to meet separately from time to time.
- The group should hold a “second chance” meeting after a preliminary consensus is reached on a preferred alternative.
- The group should consider using dissonance-inducing group processes.

## 2. HISTORY AND IMPACT OF THE GROUPTHINK PHENOMENON

More than 30 years after its conception, the groupthink phenomenon retains a remarkably strong intuitive appeal and acceptance. A recent Google search revealed more than 42,700 groupthink "hits." Groupthink is presented as received doctrine in sources ranging from *Educational Gerontology* to the *Utne Reader*, from the *New Criterion* to *Vogue*, and is offered as the cause for everything from flawed jury deliberations, to problems of the Washington Redskins, to the U.S. decision to invade Iraq. The *New York Times* recently reported that groupthink is three decades old and still going strong, invoking the phenomenon as a potential explanation for the space shuttle *Columbia* disaster. In addition to its firmly entrenched status with practitioners in a wide variety of areas, groupthink continues to be featured in popular textbooks and to be the subject of theory and research.

## 3. EVIDENCE REGARDING THE GROUPTHINK PHENOMENON

Most support for groupthink has come from retrospective case studies that have focused on decision fiascoes rather than comparing the decision-making processes associated with good decisions with those associated with bad decisions.

Experimental studies of groupthink have considered only a small portion of the model, often without a cohesive group and in situations inconsistent with Janis's proposed antecedents. Furthermore, they have relied nearly exclusively on student samples dealing with hypothetical or simulated decisions, with potential resultant problems for external validity. In the laboratory, many real-world group characteristics, including ongoing power relationships and political maneuverings, have been necessarily ignored. Empirical studies of groupthink have tended to use short-term decision quality as the sole outcome measure and have contained serious threats to internal validity. Interestingly, Janis apparently discounted these experiments. For instance, in his 1989 book, he cited none of the experimental studies of groupthink, with his discussion of groupthink referring only to case studies and anecdotal pieces.

Another important aspect of prior research concerns the situations chosen for study. The groupthink

conceptualization is restrictive, primarily due to the model's antecedents and the international policy context for which groupthink was developed. The narrow focus of the model's antecedents is not surprising given the fact that Janis extracted them directly from major policy decision fiascoes. Early case evidence provided in support of groupthink was drawn largely from these and similar hot decision situations. The ability to generalize from such situations is questionable. For instance, the extremely high group cohesiveness and clearly unified goals in many major military policy decisions, or in groups facing natural disasters, are unlikely to be encountered often in other settings. Nevertheless, laboratory and survey research on groupthink has typically used settings considerably different from Janis's conceptualization.

Strong and weak interpretations of groupthink have emerged. The strong interpretation, commonly cited, suggests that groupthink is an integrated set of characteristics with deterministic linkages. That is, groupthink characteristics cluster due to their common ties to specific antecedent conditions and are linked in a causal chain from those antecedent conditions, to symptoms, to defects, to outcomes. The weak interpretation is sometimes presented in response to the failure of empirical examinations to provide results that are wholly consistent with groupthink. It suggests that groupthink may be evidenced by the presence of some subset of these characteristics and that the causal ordering posited by Janis may be suggestive rather than necessary.

When the weak interpretation is presented, it is crucial to ask which characteristics, if any, are unique to groupthink. That is, some groupthink characteristics, such as belief in the correctness of a decision and pressures to achieve consensus, are common to many group processes and, in fact, may be beneficial. For instance, convergence to the majority viewpoint may be desirable if the majority viewpoint is correct. Other characteristics, such as mindguards and self-censorship, better discriminate between groupthink and other group processes. It is difficult to accept the argument that the presence of characteristics common to most problem-solving groups is supportive of the existence of groupthink.

Janis did not accept this weak form of groupthink. To the contrary, he wrote the following in 1989:

It does not suffice merely to see if a few of the eight telltale symptoms of groupthink can be detected. Rather, it is necessary to see if practically all the symptoms were manifested and also to see if the antecedent

conditions and the expected immediate consequences—the symptoms of defective decision making—are also present.

Recent research has continued to show little empirical support for the groupthink phenomenon. For example, in 2000, Park conducted a comprehensive empirical investigation of Janis's model, including all 24 variables (antecedent conditions, symptoms of groupthink, signs of defective decision making, and outcomes). Of 23 predictions drawn from the groupthink model, only 2 were confirmed. Conversely, 7 of the 23 relationships were significant opposite the direction predicted.

As another example, in 1999, Choi and Kim examined groupthink and team activities in 30 organizational teams faced with impending crises. They found that groupthink symptoms consisted of two factors. Contrary to groupthink predictions, one of those factors (termed "group identity") was significantly positively related to team performance, whereas the other factor (termed "concurrence seeking") showed an insignificant negative relationship to performance.

Addressing what is probably the most-cited example of groupthink, the space shuttle *Challenger* disaster, Maier (who developed a popular documentary on the topic) noted in 2002 that new evidence regarding the disaster and further analysis of past evidence demonstrates convincingly that the disaster "emphatically is not an example of groupthink." He noted that two of groupthink's defining features, the conviction of invulnerability and the illusion of unanimity, were conspicuously absent. Maier discussed evidence that the decision to launch was driven by uncertainties rather than perceived infallibility and that certain actions were taken only when it was clear that opinions would not be unanimous.

In short, research continues to find very little empirical support for the groupthink phenomenon. Coupled with several theoretical critiques, such evidence calls into question the validity of the phenomenon. Elsewhere, it has been noted that others have attached a variety of possible meanings to groupthink, as indicated in what follows.

- Groupthink is overreliance on concurrence seeking.

This meaning, which seems to have gained considerable popularity, would grant no value added to groupthink. That is, overemphasis on concurrence seeking was a widely recognized phenomenon decades before Janis coined the term "groupthink" (e.g., by Schanck in 1932), and the groupthink phenomenon simply

adopted this element. If groupthink is to have any unique explanatory value, it must have additional meaning beyond this sole concept.

- Groupthink is a complete sequence of clusters of dysfunctional characteristics that result in flawed decision making.

As noted earlier, Janis accepted this view of groupthink, which is labeled a "strong view" in this article. Quite simply, there is no support for this interpretation of groupthink. No study conducted has ever supported—or come close to supporting—this strong view. Instead, researchers have adopted weak versions in which any partial support, regardless of the number of disconfirming findings and whether or not that partial support relates to elements unique to the groupthink model, is hailed as further evidence of groupthink's validity. Furthermore, support for the posited groupings of groupthink characteristics derives from anecdotes, casual observations, and intuitive appeal rather than from rigorous research. There has been no full factor analysis of groupthink variables. Incomplete factor analyses (in which exploratory factor analysis was applied to variables within sets rather than to all variables in the model) support a simpler model than that presented by Janis. Path analysis of the complete groupthink model has revealed only a scattering of the relationships proposed by Janis. In addition, in 1992, Tetlock and colleagues used a quantitative method, the group dynamics q-sort, for analyzing historical cases of group decision making to compare Janis's case interpretations with other historical interpretations of the same cases and to test the causal linkages predicted by the groupthink model. A test of the causal sequence of the groupthink model yielded only limited support. For example, group cohesiveness and provocative situational context proved to be only weakly related to concurrence seeking, with structural and procedural faults playing larger roles. In sum, there is no solid support for the clusterings of characteristics posited by the groupthink phenomenon or for their proposed causal sequence.

- Groupthink is an undesirable constellation of characteristics resulting from highly cohesive groups.

Again, there is no empirical support for this meaning. Cohesiveness, supposedly the critical trigger in the groupthink phenomenon, has simply not been found to play a consistent role. Conceptually, there are many reasons to question this view of groupthink. First, the

benefits of group cohesion, including enhanced communication among members, heightened member satisfaction, decreased member tension and anxiety, and higher levels of task accomplishment, have long been recognized. Furthermore, any general statement about the dangers of cohesiveness ignores critical contingency variables. As just one example, the stage of group development may be important because members of a fairly cohesive group in the mature stage of group development are likely to be secure enough in their roles to challenge one another. Given all of this, the findings relating to cohesiveness in the groupthink literature are unsurprising in that they demonstrate convincingly that cohesiveness does not regularly lead to negative outcomes. More than 25 years ago, Flowers stated that a revision of Janis's theory, one that would eliminate cohesiveness as a critical variable, may be justified. Any such elimination would, of course, largely eviscerate the groupthink phenomenon. Janis viewed cohesiveness and an accompanying concurrence-seeking tendency that interferes with critical thinking as the central features of groupthink.

- Groupthink is any set of group processes that are antecedent to poor decision outcomes.

This view is tautological, suggesting that groupthink is all of those bad things that precede poor outcomes. It is not surprising, then, when poor outcomes are found to follow these bad things. Furthermore, because many group processes and characteristics leading to poor outcomes had been identified long ago, this meaning, in which groupthink most clearly adopts the form of a metaphor for dysfunction, essentially grants groupthink no status—and Janis no contribution—beyond that of providing a memorable label. Nevertheless, much of the popularity of groupthink may be attributable to adoption of this perspective. That is, there is apparently a reluctance to abandon a compelling metaphor that encourages focus on flawed group processes.

#### 4. BASES FOR GROUPTHINK'S APPEAL

In the current authors' view, groupthink has gained the status of received doctrine. Groupthink has a seductive intuitive appeal, and vivid examples that are superficially consistent with the phenomenon can readily be recounted. But how did groupthink come to be so widely and enthusiastically accepted in the absence of

compelling evidence? This section, based in part on Aldag and Fuller's earlier review in *Psychological Bulletin*, explores the bases for groupthink's appeal.

##### 4.1. Implicit Assumptions

Groupthink's appeal appears to rest in part on a number of implicit assumptions. First, although groupthink assumes that groups are used primarily to enhance the quality of decisions, group decision making is often used specifically to increase the acceptance of a decision. Also, organizational members may choose to use groups for other reasons, including socioemotional benefits and accomplishment of members' own secondary goals (e.g., "hidden agendas").

Second, groupthink implicitly assumes that groups rationally pursue a unified goal. However, group processes are characterized by interplays of covert and overt motivations, concern not just about the current problem but also about residues of past problems and the anticipation of future problem-related interactions, and many other subtleties outside the typical focus of the rational model. Processes and outcomes that appear irrational to "objective" observers may in fact be functional from the point of view of individual members or even of the entire group. For instance, such processes and outcomes may serve to maintain the motivation of the group leader, help ensure the continued use of group processes, prevent the defection of certain group members, serve as a means for members to show allegiance to coalitions, and minimize the likelihood of the use of still other political acts.

A third implicit assumption is that actions taken to prevent or minimize groupthink will produce a more rational and effective group process. However, these remedies may have unintended consequences. That is, attempts to create a microrational decision process in an inherently political environment may prove to be misguided; other literatures have suggested the dangers of presenting an oasis of change in a nonreinforcing desert. The groupthink remedies are decision specific, whereas the causes for supposed group dysfunctions may be systemic. That is, the application of specific techniques to prevent or minimize group difficulties may amount to treating the symptom of a deeper problem.

Fourth, fundamental to groupthink is the assumption that a group's illusion of well-being is dysfunctional. However, in 1988, Taylor and Brown argued that such illusions actually promote many criteria of mental health, including the ability to care for others, be happy or

content, and engage in productive and creative work. They added that such positive illusions are especially useful when an individual receives negative feedback or is otherwise threatened and may be especially adaptive under these circumstances. Thus, this illusion of well-being may be most useful specifically in the hot decision situations described by Janis. Although Taylor and Brown focused on the individual, the processes that they detailed seem equally applicable to groups.

Such contentions are reinforced by consideration of the related literature of the self-fulfilling prophecy, referred to in some settings as the Pygmalion effect. The Pygmalion and similar expectation effects appear to be common in organizational settings. Taken as a whole, evidence concerning these related effects suggests that nonveridical perceptions may often yield positive consequences. The literature on optimism–pessimism leads to similar conclusions. That is, although pessimism is often found to be associated with more realistic worldviews, optimism typically yields superior outcomes. These arguments are not meant to imply that overly positive perceptions are necessarily desirable; they are only meant to imply that the opposite should not be automatically assumed.

A final groupthink assumption concerns causes of characteristic groupthink attitudes and behaviors. Rather than, or in addition to, stemming from individual psychological factors, conformity behaviors may be politically motivated. For instance, they may reflect compliance rather than internalization. Thus, many groupthink characteristics could be due to conscious actions by members to subscribe to political norms.

What the preceding implicit assumptions have in common is a beguiling, albeit unrealistic, simplicity.

The reliance on anecdotes in groupthink scenarios and case studies contributes to the availability heuristic. It does this in at least two ways. First, groupthink examples tend to be vivid and “catchy” and, thus, are salient and readily available in individuals’ memories. Second, such examples play to individuals’ preference for case, as opposed to base, data. A concrete instance of the appearance of groupthink symptoms in a fiasco may be seen as compelling evidence, especially in the absence of base data.

## 4.2. Implicit Theories of Group

The groupthink phenomenon is consistent with implicit theories of groups. The roles of implicit theories in influencing responses have been widely documented. One important characteristic of implicit theories is

their “gap-filling” function. In the context of groupthink, individuals observing a situation in which some groupthink characteristics are present may assume the existence of others. Furthermore, research shows that feedback given to group members or observers about group performance affects the characteristics ascribed to those groups. For example, feedback about group process has been shown to affect evaluations of outcomes, and feedback about outcomes has been shown to affect evaluations of process. Individuals who are told that a group has performed poorly are more likely to report instances of “poor” interaction processes such as lack of willingness to hear other members’ views. Thus, the focus on poor decision outcomes in groupthink research may lead to reports of poor group functioning.

## 4.3. Illusory Correlation

A focus on only the conjunction of groupthink characteristics and negative outcomes invites illusory correlation. In this sense, the groupthink phenomenon is similar to the “Friday the 13th” phenomenon; only the yes–yes (i.e., groupthink/Friday the 13th–poor outcomes/bad luck) cell is considered. If the yes–yes cell is not empty, support for the phenomenon is inferred. In fact, of course, support for the phenomenon requires examination of all cells. Furthermore, when research is used to explain a poor decision, the process used to “validate” groupthink nearly guarantees support, however partial. That is, in the aftermath of a fiasco, transcripts and other documents, videos, and casual observation may be sifted for “evidence” of groupthink. If that search unearths comments that deal with conformity, caution, or closed processes, a “hit” is registered in the yes–yes cell and the presence of groupthink is announced. However, some reference to such terms is nearly inevitable in the mass of evidence surrounding a fiasco.

## 4.4. Framing

The negative language of groupthink (e.g., “victims of groupthink,” “defects of groupthink”) and the focus on error can invite distortions in responses caused by scale use tendencies and related psychometric difficulties and can also result in framing effects. Individuals (whether research participants or researchers) presented with negatively framed terminology may adopt the readily available negative frame and respond accordingly. Therefore, even simple attempts by the participants to give responses that are consistent with the tone of the

questions would result in negatively oriented responses. On a more fundamental level, this framing has resulted in a focus on error rather than on decision quality *per se*. Janis noted in 1982 that he began studying fiascoes “for the purposes of studying sources of error in foreign policy decision making.” However, just as there is more to the performance of a football team than the absence of fumbles and interceptions, there is more to group decision quality than the absence of error. A focus on negative outcomes of group processes may divert attention from group synergies. One example is the assembly effect bonus, which is productivity that exceeds the potential of the most capable member as well as the sum of the efforts of group members working separately. Thus, one may learn little about superior group performance by focusing solely on fiascoes. Instead, a focus on decisions with a broad range of outcomes, including superior performance, is necessary.

In the current authors’ view, groupthink owes much of its popularity to a characteristic it shares with other popular concepts such as total quality management. A core concept with some validity (e.g., statistical quality control) is incorporated as an element of a broader renamed concept (e.g., total quality management). Support for the core concept is subsequently treated as confirmation of the broader concept and, by association, for its various elements. In the case of groupthink, a core concept with some validity (i.e., overemphasis on concurrence seeking may be harmful) is subsumed in a complex, essentially deterministic model. Subsequent instances of that core concept are then presented as evidence for the validity of the broader groupthink phenomenon.

## 5. IMPLICATIONS AND SUGGESTIONS FOR FUTURE RESEARCH AND THEORY

Groupthink has stimulated research on group dysfunctions, provided links to other literatures such as stress and vigilance, emphasized potentially important variables in group decision making, and encouraged policymakers to take remedies for excessive concurrence seeking seriously. Indeed, Janis’s recommendations for “remedies” for groupthink are themselves an excellent compilation of approaches to help preclude group dysfunction.

However, group problem solving is a complex and challenging phenomenon. Although the groupthink

model has played an important role in generating interest in group problem-solving processes and in acting as a catalyst for associated theory and research, it has not incorporated three decades of theory and research, has received limited empirical support, and is restrictive in scope. Recent theory and research, as well as critical evaluation of the model, suggest that more comprehensive models are necessary to guide researchers and practitioners in dealing with the group decision phenomenon.

It is not unusual for theories to generate initial widespread interest and enthusiasm and then to meet with subsequent revision, rejection, or reaffirmation. However, groupthink has generally resisted dispassionate reevaluation. This is perhaps due in large part to the raw intuitive appeal of the phenomenon as well as to the fact that studies of groupthink have often been searches for confirmation. Rigorous evaluation of the phenomenon is further rendered difficult by the fact that a variety of views of groupthink have emerged, as have contrasting positions on what level of evidence is needed to indicate support.

Similarly, popular acceptance of groupthink has been extraordinary. Perhaps this is understandable: Groupthink has served as a vivid bogeyman that can be readily summoned to illustrate the dangers of overemphasis on concurrence seeking, and it continues to serve its purpose. This article has provided a variety of potential reasons for this enthusiastic acceptance and has shown that they are consistent in yielding illusory support.

Although attention to groupthink has yielded some clear benefits over the past three decades or so, there is growing consensus that the phenomenon lacks conceptual and empirical support, is restrictive in scope, and is inappropriately deterministic. As such, future research should seek to build on decades of groupthink and other research and theory on group processes and outcomes to yield more encompassing, less causally constrained perspectives. New theories and models should generalize to a broad range of group types and situations (e.g., including self-managing groups and virtual groups), incorporate additional relevant variables (e.g., leader motivations, stage of group development, power relationships), and recognize that groups, as well as group leaders, may have multiple goals (e.g., sustained motivation of the group leader, continued use of group processes, prevention of defection of group members). Finally, the research should avoid selection on the dependent variable; instead, it should consider instances of both favorable and poor group

outcomes and should seek to be theory building rather than theory confirming. In this way, the energy, enthusiasm, and intellectual and other resources that have been directed toward the groupthink phenomenon for more than three decades may yield broader, more inclusive, and more valid models and theories of group decision processes.

## 6. PRESCRIPTIONS FOR EFFECTIVE GROUP PROBLEM SOLVING

As this analysis has indicated, the current authors believe that the topic of group problem solving should be viewed more comprehensively than within the confines of the restrictive groupthink phenomenon. Taking into account the broader literature on group problem solving, it is clear that there are legitimate threats to group processes beyond—and sometimes inconsistent with—those suggested by the groupthink model. With this in mind, the authors offer the following advice to increase the effectiveness of group problem solving.

The first recommendation is to pay careful attention to group process and provide structure and guidance to ensure that the full process proceeds effectively. There are five stages to the problem-solving process: problem identification, generation of alternatives, evaluation of alternatives, decision making, and decision implementation and control. It is important that a group proceed systematically through these stages, allotting appropriate time and attention to each stage. In addition to proper structuring of the process by the group leader, some tools (e.g., the nominal group technique) and computerized decision aids (e.g., decision insight systems) may also prove to be valuable for this purpose.

A second recommendation is to consider the level of conflict in the group. The leader must be able to assess the degree of conflict and strive to bring the group toward an optimal level, where different viewpoints are encouraged and accepted but where the differences are viewed as adding information and contributing to the discussion. In any group, or at any time, conflict might have to be reduced, stimulated, or maintained. If conflict levels are very low, productive dissent might be needed. That is, as suggested by Schanck and others, group decisions may suffer if there is a strong pressure to conform and remain silent against the dominant view. In such cases, techniques suggested by Janis (e.g., the devil's advocate, "second chance meetings") might be

useful. In addition, leaders can form subgroups that take different sides of an issue and can encourage norms that all employees, regardless of their positions, should fully express their concerns, doubts, and ideas, emphasizing that it is essential to examine issues critically and to find flaws and problems before competitors or customers do. Another approach to encourage productive dissent is to form task forces, made up of employees with various perspectives, outlooks, organizational positions, and backgrounds, to examine major problems and offer recommendations. Conversely, too much conflict will also result in poor group process and decisions. In such cases, it may be useful to encourage the group to focus on larger goals, directly intervene to alleviate conflict, separate conflicting parties, work to improve communication among members, train members in finding integrative solutions, or bring in third-party mediators.

Third, the discussion suggests that it is important to consider many outcomes of group problem solving rather than focusing just on short-term decision quality; decision implementation and commitment to decisions may be as important as the decision itself. Future motivation of the group and leader, future use of the group, and member affective responses may be important outcomes for the group and/or its members. The leader must recognize trade-offs among multiple goals, sometimes subordinating short-term decision quality to goals such as maintaining member satisfaction and enhancing the likelihood of future use of the group.

Fourth, caution must be exercised in the use of any individual technique. In an organization, a group functions within a broader environment, that is, the culture of the organization. As such, a specific remedy applied to a group may be ineffective if it contradicts the dominant cultural norms in the organization. Therefore, attempts to improve group functioning must be viewed within the broader context and should explicitly consider systemic issues. It may be more appropriate to assess and perhaps alter organizational culture or contingencies of reinforcement than to implement specific techniques in the hopes of overcoming the consequences of the group's decision environment.

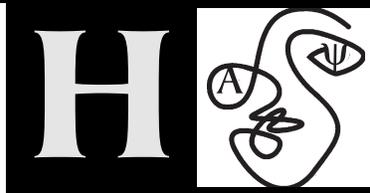
Finally, it is important for the group leader to recognize and manage the complexity and challenges of group problem solving. The leader must strive to keep an open mind, avoiding limiting assumptions and models—no matter how alluring—and maintaining broad behavioral repertoires to permit flexible responses to the rich contingencies influencing group processes and outcomes.

### See Also the Following Articles

- Conflict within Organizations ■ Groups, Productivity within  
 ■ Intergroup Relations and Culture ■ Interpersonal Conflict

### Further Reading

- Aldag, R. J., & Fuller, S. R. (1993). Beyond fiasco: A reappraisal of the groupthink phenomenon and a new model of group decision processes. *Psychological Bulletin*, *113*, 533–552.
- Callaway, M. R., & Esser, J. K. (1984). Groupthink: Effects of cohesiveness and problem-solving procedures on group decision-making. *Social Behavior and Personality*, *12*, 157–164.
- Choi, J. N., & Kim, M. U. (1999). The organizational application of groupthink and its limitations in organizations. *Journal of Applied Psychology*, *84*, 297–306.
- Flowers, M. L. (1977). A laboratory test of some implications of Janis's groupthink hypothesis. *Journal of Personality and Social Psychology*, *33*, 888–895.
- Fuller, S. R., & Aldag, R. J. (1998). Organizational Tonypanody: Lessons from a quarter century of the groupthink phenomenon. *Organizational Behavior and Human Decision Processes*, *73*, 163–184.
- Janis, I. L. (1971, November). Groupthink. *Psychology Today*, pp. 43–46, 74–76.
- Janis, I. L. (1972). *Victims of groupthink*. Boston: Houghton Mifflin.
- Janis, I. L. (1982). *Groupthink* (2nd ed.). Boston: Houghton Mifflin.
- Janis, I. L. (1989). *Crucial decisions: Leadership in policymaking and crisis management*. New York: Free Press.
- Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.
- Longley, J., & Pruitt, D. G. (1980). Groupthink: A critique of Janis's theory. In L. Wheeler (Ed.), *Review of personality and social psychology* (pp. 507–513). Beverly Hills, CA: Sage.
- Maier, M. (2002). Ten years after *A Major Malfunction . . .*: Reflections on "The Challenger syndrome." *Journal of Management Inquiry*, *11*, 282–292.
- Park, W. (1990). A review of research on groupthink. *Journal of Behavioral Decision Making*, *3*, 229–245.
- Park, W. -W. (2000). A comprehensive empirical investigation of the relationships among variables of the groupthink model. *Journal of Organizational Behavior*, *21*, 873–887.
- Paulus, P. B. (1998). Developing consensus about groupthink after all these years. *Organizational Behavior and Human Decision Processes*, *73*, 362–374.
- Schanck, R. L. (1932). A study of a community and its groups and institutions conceived as behaviors of individuals. *Psychological Monographs*, *43*(2, whole No. 195).
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social-psychological perspective on mental health. *Psychological Bulletin*, *103*, 193–210.
- Tetlock, P. E., Peterson, R. S., McGuire, C., Chang, S., & Feld, P. (1992). Assessing political group dynamics: A test of the groupthink model. *Journal of Personality and Social Psychology*, *63*, 403–425.
- 't Hart, P., Stern, E. K., & Sundelius, B. (1997). *Beyond groupthink: Political group dynamics and foreign policymaking*. Ann Arbor: University of Michigan Press.
- Whyte, G. (1989). Groupthink reconsidered. *Academy of Management Review*, *14*, 40–56.



# Health and Culture

George D. Bishop

National University of Singapore, Singapore

1. The Concept of Culture
  2. Culture and Health-Related Cognition
  3. Culture and Health-Related Behavior
  4. Culture and Physiological Responses
  5. Integrating Culture into Health Psychology
  6. Conclusion
- Further Reading

## GLOSSARY

**balance model** A model that describes health in terms of a harmonious balance between different elements or forces in the body.

**culture** Guidelines that are inherited from the social environment that guide how people in a particular cultural group view the world, respond to it emotionally, and behave in it.

**machine model** A model of the body as being an internal-combustion or battery-driven machine.

**“plumbing” model** A model that views the body as being a series of cavities or changes connected to each other through pipes or tubes.

This article examines the interrelationship of culture and health with emphasis on the ways in which culture influences health outcomes through its effects on health-related thoughts and behaviors as well as physiological responses. Examples are given of these influences and the article concludes with a discussion of how cultural influences can best be integrated into research and theory in health psychology.

Considerations of culture are becoming increasingly prevalent in psychology, and health psychology is no exception. This article explores the relationship between culture and health and ways in which considerations of culture are important in the development of a theoretically and empirically sound health psychology. Specifically, culture plays a key role in how people understand states of health and illness as well as the types of health-related behavior in which they engage. In addition, there is growing evidence that culture influences physiological responses. As such, considerations of culture are essential to understanding various aspects of health.

## 1. THE CONCEPT OF CULTURE

Many definitions have been given for culture. Most, however, view culture as a set of guidelines that people inherit from their social environments that guide how they view the world, respond to it emotionally, and behave in it relative to the natural environment, other people, and supernatural forces or gods. As such, culture provides a set of lenses that shape experience. These “lenses” are transmitted from one generation to the next in various ways and provide the basis for shared experience within the cultural group. One of the ways in which culture shapes experience is through the provision of categories for dividing and organizing experience. In the case of health, such categories are key to understanding how people experience health

and illness and how they report that experience to others, such as health professionals.

It is important to note that culture is not something that is uniform within a society; rather, any society is certain to have multiple cultures within it. These cultures may be differentiated on the basis of ethnicity, social class, gender, occupation, and any number of other attributes that are associated with shared experience within a group that differentiates it from other groups. Furthermore, cultures change over time as different cultural groups interact, conditions change, technology develops, and so on. Hence, culture is best seen as dynamic rather than static.

## 2. CULTURE AND HEALTH-RELATED COGNITION

One of the key ways in which culture influences health is through health-related cognition.

### 2.1. Variation

Variation in health and illness concepts between cultures is obvious even to the most casual observer. These variations are exhibited in a wide variety of ways, but four of these are particularly striking.

#### 2.1.1. *The Basic Nature of Health and Illness*

First, there is substantial variation between cultures in conceptions of the basic nature of health and illness. For example, in Western societies placing a strong emphasis on the scientific understanding of health and illness, these concepts are understood in natural terms with disease viewed as pathology due to natural causes such as germs, genetic aberrations, or injury. In some other societies, however, as well as among some minorities in Western societies, illness may be understood more in interpersonal or supernatural terms. For example, illnesses may be seen as being the result of violations of social norms or standards of personal behavior, particular emotions (e.g., anger or jealousy), or transgressions of moral or religious taboos.

#### 2.1.2. *Beliefs about Anatomy and Physiology*

Second, beliefs about anatomy and physiology often show marked cultural variation, with cultures often

differing in their basic model of the body and what constitutes health and illness. For example, a number of forms of traditional medicine, including Chinese traditional medicine as well as traditional medical beliefs in India, Latin America, and the Caribbean, hold to the balance model. This model describes healthy functioning of the body in terms of a harmonious balance between different elements or forces within the body. Illness is believed to be caused by imbalance, which must be rectified for the person to return to health. In contrast, in Western societies the body is often conceptualized in terms of the “plumbing” model and/or machine model. The former views the body as being a series of cavities or chambers (e.g., stomach and chest) connected to one another by pipes or tubes. In this model, health is maintained by an uninterrupted flow of various substances (e.g., blood, air, food, and urine) between cavities or between cavities and the exterior of the body. By the same token, disease is the result of blockage. The machine model, by contrast, views the body as a machine that operates according to the principles of physics and chemistry and that may run down or break down, requiring a doctor to repair it.

#### 2.1.3. *Concepts and Terminology*

Third, cultures often differ markedly in the concepts and terminology used to describe various conditions. For example, symptoms often diagnosed as depression in the West may be interpreted as neurasthenia (nervous exhaustion) in China. Also, conditions defined in one culture may be undefined in another. For example, traditional Mexican American medical beliefs define conditions such as *susto* (fright), *empacho* (surfeit), *caida de mollero* (fallen fontanelle), and *mal de ojo* (evil eye), which have no direct referents among conditions known to medical professionals or Anglo-Americans.

#### 2.1.4. *Underlying Structure of Illness Representations*

Finally, the underlying structure of illness representations often differs between cultures. Research has shown that the mental representations that people have of different diseases show a characteristic organization in memory that differs from culture to culture. For example, research has consistently found that disease representations among Anglo-Americans are organized along dimensions of contagiousness and severity. In contrast,

disease representations among Mexican Americans appear to be organized according to concepts of “hot” and “cold” and whether a disease is associated with children or old people, and those of Chinese Singaporeans are organized according to a combination of dimensions deriving from Chinese and Western medicine.

## 2.2. Implications

These cultural differences in health-related cognition have a number of important implications. First, the cognitive representations that people have of diseases may well have a strong influence on their interpretation of symptoms and bodily symptoms. The existence of specific disease concepts makes it possible and perhaps even likely that bodily experience will be interpreted in line with that concept, whereas the absence of that concept would foreclose that possibility. For example, the fact that Mexican Americans have available to them the concept of *empacho* makes it possible for them to experience this condition, whereas members of other groups not possessing this concept would not be able to experience it. Similarly, the prominence of the concept of *neurasthenia* in China and of depression in North America makes it more likely that the similar symptoms associated with these two conditions would be interpreted, and quite likely experienced, differently.

In addition, health-related cognitions have implications for what people do when experiencing distress. Illness representations provide a definition for when bodily distress represents illness and requires assistance from a medical professional or other healer. It has been noted, for example, that one of the reasons why heart attack victims delay seeking treatment is that they misinterpret the symptoms and underestimate the seriousness of the situation. Furthermore, once the person has determined that help is needed, mental representations of illness direct the person's help-seeking behavior. Research with Chinese Singaporeans has shown that whether a person seeks help from a practitioner of Western scientific medicine or Chinese traditional medicine depends on the extent to which the person endorses illness beliefs associated with Chinese traditional medicine. Finally, once help is sought, people's mental representations of their condition are likely to strongly influence whether they stay in treatment. Research with hypertensives, for example, has shown that the cognitive “model” they have of hypertension and its relationship to specific symptoms is a key determinant of whether they stay in treatment.

## 3. CULTURE AND HEALTH-RELATED BEHAVIOR

A second important way in which culture influences health is through behavior.

### 3.1. Behavior and Disease Risk

The relationship between behavior and health is well established and there is strong evidence for the importance of behavioral risk factors such as smoking, diet, exercise, and alcohol use. It is also obvious to even the most casual observer that these behaviors often differ dramatically between cultural groups. For example, alcohol use is very prevalent in some countries, such as Russia, Ireland, France, and Italy, but much lower in other countries, such as predominantly Muslim countries in which alcohol use faces religious prohibition. Also, dietary habits differ substantially between cultures with regard to what is defined as food as well as which foods are allowed or prohibited, how foodstuffs are prepared, and overall patterns of eating. For example, although dog meat is eaten in some societies, most Americans and many others are appalled at the idea of considering dogs to be a source of food. Also, religion often dictates what foods are allowed. Orthodox Hindus, for example, are forbidden to eat any kind of animal, whereas Jews and Muslims are prohibited from eating pork, and any meat they do eat must be slaughtered in specific ways.

Cultural differences in behavior have an important impact on diseases prevalent in different countries and among different cultural groups. For example, research on cultural differences in diet has shown that the “Western” diet, with its high saturated fat content, is associated with high levels of coronary heart disease (CHD), whereas the low-fat Japanese diet is associated with low levels of CHD. Furthermore, the “Mediterranean” diet, which is high in monounsaturated fats and antioxidants, has been found to be protective against CHD. Similarly, elements of the Japanese diet have been associated with the development of stomach cancer and the habit of chewing a mixture of tobacco, betel nuts, and other substances in countries such as India and Afghanistan has been associated with high levels of oral cancer. Also, studies of alcohol consumption have shown that cross-cultural differences in alcohol consumption are strongly related to the incidence of diseases such as liver disease, CHD, stroke, hypertension, and diabetes. Finally, the

prevalence of HIV infection and AIDS as well as other sexually transmitted diseases is strongly related to cultural attitudes toward sexuality and the acceptance of promiscuity, commercial sex work, and alternative sexual practices.

### 3.2. Help-Seeking

In addition to culture's role in the etiology of disease, it is also important to note cultural differences in help-seeking and the role of culture in treatment. As noted earlier, cultural differences in cognitive representations of disease are likely to influence whether a person seeks treatment for specific symptoms. In addition, there are also cultural differences in reasons for seeking treatment and the kind of treatment that is considered to be appropriate. For example, it has been noted that for Italian Americans, seeking medical care tends to be triggered by an interpersonal crisis or when the person perceives that the symptoms interfere with personal relationships. In contrast, Irish and Anglo-Americans tend to seek help when symptoms are perceived to interfere with work or physical functioning. Furthermore, when seeking help different cultural groups may present their symptoms differently and adopt a different "language of distress." Research has found that Irish Americans tend to underplay their symptoms, taking a stoical attitude toward them, whereas Italian Americans tend to present their symptoms in a more voluble, emotional, and dramatic manner. These differences, in turn, may have an influence on diagnosis and the treatment received.

After the person has sought treatment, culture may play an important role in the interaction with the health care practitioner. Due to migration, most societies are becoming increasingly multicultural, with the result that it is commonly the case that patients and health care practitioners may come from different cultural backgrounds. When this is the case, differing cultural assumptions may prove a barrier in communication as well as in the acceptance of medical advice. For example, it has been noted that Native Americans tend to underuse medical services that are available to them and often do not follow medical recommendations even when they do utilize services. Among the reasons for this are conflicts between scientific and traditional Native American understandings of disease as well as the history of oppression of Native Americans and differences in communication styles between nonnative health care practitioners and Native clients. The end result may be distrust of the nonnative health care practitioner unless the practitioner makes specific efforts to

understand Native beliefs and also approaches Native clients in a manner that employs both verbal and non-verbal behavior that is culturally appropriate.

## 4. CULTURE AND PHYSIOLOGICAL RESPONSES

In addition to the often very evident effects of culture on beliefs and behavior, there is also a growing body of evidence suggesting cultural differences in physiological responses that may have important health implications. There is often a tendency to assume that physiological functioning and disease processes are the same across cultural groups. However, this may not always be the case. For example, numerous studies have shown significant relationships between personality traits such as trait anger and hostility and cardiovascular responses to certain stressors. However, there is also evidence that this relationship may vary according to ethnicity. In one study, correlations between hostility and cardiovascular responses to stress that were negative among Caucasian participants were positive among African American participants and vice versa. Also, there is evidence that Asian Indians show a different pattern of cardiovascular responses to stress than do Chinese or Malays. Other research has shown that the relationship of socioeconomic variables and hostility to cardiovascular reactivity to stress and left ventricular mass is different between Caucasians and African Americans.

The reasons for these differences are not clear and they may be the result of either cultural differences between groups or genetic differences. There is evidence, however, that even genetic influences may vary by cultural group. For example, studies of the serotonin transporter gene have shown that although polymorphisms of this gene are related to both personality and cardiovascular responses to stress, this relationship differs between African Americans and Caucasians. This suggests an interaction between culture and genetics in physiological responses to stress that may have important implications for understanding health differences between groups.

## 5. INTEGRATING CULTURE INTO HEALTH PSYCHOLOGY

It is clear from the previous discussion that culture and health are closely intertwined, with cultural factors

having a strong influence on health through cognitive, behavioral, and physiological pathways. However, the study of cultural factors in health psychology has only recently begun, and there is a great deal more that needs to be done to fully integrate cultural considerations into health psychology.

Research on cultural differences has most commonly involved making comparisons between specified cultural groups, an approach sometimes referred to as the “packaged” approach. The assumption is that differences between cultural groups are the result of differing cultural characteristics. However, this approach suffers from the fact that it ignores potential genetic and other physical differences between groups and does not reveal which aspects of culture are responsible for the differences observed.

Because of the limitations of the packaged approach, cross-cultural psychologists strongly advocate going beyond the superficial aspects of existing cultural groupings by examining the underlying dimensions of culture. Any number of dimensions might be identified, but attention has focused primarily on dimensions identified in classic research by Hofstede (individualism–collectivism, masculinity–femininity, certainty avoidance, and power distance) as well as those based on cultural values. Only a few studies have examined the relationship between dimensions of culture and health-related variables, but results have been promising. For example, studies of Chinese values and their relationship to health have shown that the values of social integration and reputation are significantly related to both alcohol and tobacco consumption as well as to diseases such as cerebrovascular disease, heart disease, and cancer.

## 6. CONCLUSION

Culture plays a significant role in health through several different mechanisms. Cultures differ in their basic understanding of the phenomena of health and illness as well as in their concepts of specific health conditions.

These differences have important implications for the interpretation of symptoms and bodily symptoms and also for the kinds of help that people seek when experiencing distress and whether they stay in treatment once they seek help. Culture also plays a significant role in health-related behavior. Substantial cultural differences exist with regard to behaviors such as dietary practices, smoking, alcohol use, sexual practices, and exercise, which are known to have a significant impact on risk for various diseases. Also, cultural groups differ in their propensity to seek medical attention and cultural factors figure prominently in the type of care that people seek and in their relationship to health care providers. Finally, there is increasing evidence of cultural differences in physiological mechanisms related to disease. Although the evidence for the importance of culture in health is strong, there is a very real need to go beyond the study of differences between cultural groups to the examination of how underlying dimensions of culture are predictive of health-related cognition, behavior, and physiological responses.

### See Also the Following Articles

Health Promotion in Schools ■ Health Psychology, Cross-Cultural ■ International Classification of Diseases (WHO) ■ Women's Health

### Further Reading

- Angel, R., & Thoits, P. (1987). The impact of culture on the cognitive structure of illness. *Culture, Medicine, and Psychiatry, 11*, 465–494.
- Bishop, G. D. (1991). Understanding the understanding of illness: Lay disease representations. In J. A. Skelton, & R. T. Croyle (Eds.), *Mental representations in health and illness*. New York: Springer-Verlag.
- Helman, C. G. (1995). *Culture, health and illness: An introduction for health professionals* (3rd ed.). Oxford, UK: Butterworth Heinemann.
- Kazarian, S. S., & Evans, D. R. (Eds.). (2001). *Handbook of cultural health psychology*. San Diego: Academic Press.



# Health Promotion in Schools

Ellis P. Copeland and Franci Crepeau-Hobson

University of Northern Colorado, Greeley, Colorado, USA

1. Introduction
  2. Child and Adolescent Health Concerns
  3. Health Promotion
  4. Coordinated School Health Programs
  5. Summary
- Further Reading

## GLOSSARY

**health** A state of complete physical, mental, spiritual, and social well-being, not merely the absence of disease or infirmity.

**health-promoting school** A place where all members of the school community work collaboratively to provide students with integrated and positive experiences and structures that promote their health.

**health promotion** A combination of behavioral, educational, social, spiritual, economic, and environmental efforts that support the development, maintenance, and enhancement of behaviors and lifestyles conducive to emotional and physical well-being.

**indicated prevention programming** Interventions designed for children or families demonstrating danger signs (e.g., multiple risk factors, substance abuse, and neglect) and that are more intensive and intrusive than other prevention programs.

**initiative** The dedication of cumulative effort over time to achieve any goal composed of the elements of intrinsic motivation, concentration, and challenge.

**mental health** The successful performance of mental function, resulting in generative activities, fulfilling relationships, and the ability to cope with adversity.

**selective prevention programming** Interventions targeted at high-risk populations. The target can be children, families, or a high-risk school. The focus is most often on reducing risk factors or targeted behavior changes.

**social emotional learning** Knowledge, habits, skills, and values that provide a foundation for a child's academic, personal, social, and civic development and are necessary for success in school and life.

**universal prevention programming** Interventions designed for the entire population, such as a districtwide program designed to prevent substance abuse and/or violence.

The most serious and expensive health and social problems in the United States today are caused by negative behavioral patterns. Unhealthy behavioral patterns have adversely affected the physical, emotional, and mental health of youth, resulting in problems with alcohol, drugs, diet, anger management, and poor impulse control. These health problems and unhealthy behaviors are largely preventable and studies of health education and promotion in schools have demonstrated that health education programming has effectively reduced the prevalence of health risk behaviors among youth. Individual and family interventions have shown some success, but programs that champion a seamless, comprehensive (kindergarten–12) system of health promotion demonstrate great promise.

## 1. INTRODUCTION

Despite advances in medical sciences, today's youth face complex health challenges that are especially daunting. A significant number of youth enter school with "pre-existing" health conditions (e.g., asthma and diabetes) that, without appropriate intervention, can affect both academic and social progress. By adding communicable diseases, malnutrition, behavior-related conditions (e.g., exposure to smoking, eating disorders, substance abuse, and sexually transmitted diseases including HIV), violence, suicide, and accidents and injuries to the mix, one quickly realizes that there is a formidable problem. In the past, schools and other health and social services played a major role in impacting the health of young people through their involvement in immunization, screening, and educational programming. However, schools' involvement in health promotion has been sporadic and not as coordinated as previous immunization efforts.

Added to the health factors are the multiple environmental risk factors that are often antecedents of a variety of disorders. An increase in health and behavior problems can be linked to significant family, school, and neighborhood changes that have occurred during the past few decades. These changes include a breakdown in the traditional family and neighborhood; a devaluing of parenting, teachers, and religious life; damaging television and media messages (e.g., sex, violence, marketing of unhealthy dietary behaviors, and promotion of immediate gratification); lack of positive role models and a reduction in meaningful contact with positive adult role models; and an increase in inadequate housing and unsafe environments (often inclusive of the school environment).

Problems with regard to living conditions and the environment have taken their toll and, in large part, led the Centers for Disease Control and Prevention (CDC) to use the Youth Risk Behavior Surveillance System (YRBSS) to monitor the following six categories of priority health risk behaviors among youth and young adults every 2 years since 1991:

- Behaviors that lead to intentional (e.g., violence) and unintentional injuries
- Smoking/tobacco use
- Alcohol consumption and other drug use
- Sexual behaviors leading to disease, HIV infection, or unintended pregnancy
- Poor nutrition/unhealthy dietary behaviors
- Lack of/inadequate physical activity

These are all behaviors that can increase during childhood. The YRBSS monitoring system enables states to compare their data to national data and trends.

Relatedly, vital statistics indicate that 70.6% of all deaths among youth and young adults aged 10–24 years result from motor vehicle accidents, unintentional injuries, homicide, and suicide. Analysis of this national school-based survey indicates that many high school students engage in behaviors that increase their risk for death from one of these four causes. For example, 14.1% of the youth surveyed had rarely or never worn a seat belt during the 30 days preceding the survey, and 8.8% had attempted suicide during the 12 months prior to the survey.

The leading causes of mortality and morbidity among all individuals in the United States are related to health risk behaviors that are frequently established during youth: tobacco use, alcohol and drug use, unhealthy dietary behaviors, physical inactivity, etc. Thus, a realistic analysis of the physical, behavioral, and mental health of youth appears to be in order.

## 2. CHILD AND ADOLESCENT HEALTH CONCERNS

### 2.1. Physical Health

In 2001, almost 14% of high school students surveyed were at risk for becoming overweight, whereas 10.5% already fit this category. The problem is greater for males (14% overweight) than females (7% overweight). Fewer than 22% consumed the recommended number of servings per day of fruits and vegetables and milk, and an increasing number of schools include fast-food choices on their lunch menus. Furthermore, eating habits that begin in childhood contribute to numerous chronic diseases in adulthood (e.g., heart disease, cancer, and osteoporosis).

In addition to the unhealthy trends in dietary habits of youth in the United States, there has been a decrease in opportunities for physical activity, with slightly more than half (51.7%) of high school students enrolled in physical education (PE) classes. Less than one-third of these students attended PE classes daily—an increase from 25.4% in 1995 but far below the 1991 level of 41.6% attending daily. The percentage of students reporting no vigorous or moderate physical activity in the past 7 days is 9.5%. Combining lack of physical activity with increasingly unhealthy eating patterns, in addition to 2 hours or more of television and/or video

and computer games per day, results in multiple risk factors. Not surprisingly, children who participate in regular physical activity have more favorable cardiovascular disease risk factors than less active children.

The vulnerable ages for tobacco use are 10–18 years, when most users start smoking, chewing, or dipping and become addicted to the substance. Very few individuals begin to use tobacco as adults; almost all first use occurs prior to graduation from high school. Slightly more than one in three persons use tobacco by age 18 (33.9% report using tobacco in the past 30 days). The earlier a young person begins to use tobacco, the more likely the person is to abuse the substance as an adult. A surgeon general's 1994 report directly linked the deaths of 434,000 million Americans each year to tobacco usage. Because both the duration and the amount of tobacco use are related to eventual chronic health problems, early interventions with youth concerning tobacco use must be stressed as a physical health issue.

## 2.2. Drugs and Alcohol

Underage drinking is a serious problem in the United States. Alcohol is the drug most commonly used and abused by children aged 12–17 years. Youth in the United States are reportedly drinking alcohol at an earlier age than ever before, typically beginning in early adolescence (approximately age 13). A recent survey found that more than half of eighth graders and 80% of high school seniors had used alcohol sometime in their lives, and that 25% of eighth graders and more than 60% of high school seniors had been drunk at some time. The percentage of middle and high school students who had five or more drinks of alcohol at one time (within a few hours) on one or more of the past 30 days reached a high in 1997 at 33.4% and was 29.9% in 2001.

Health-related risks of teenage alcohol use are significant. Children who begin using alcohol before the age of 15 are four times more likely to become an alcoholic as an adult than those who start drinking at age 21. Research indicates that more than 40% of teens that start drinking at age 14 or younger become dependent on alcohol as adults. The risk of death from car accidents and drowning increases with alcohol use. Drinking large amounts of alcohol can result in alcohol poisoning, which can lead to coma or even death. Adolescents who drink are much more likely to attempt suicide, engage in violence against others, and become victims of violent crime.

Young people who drink alcohol are also more likely to engage in other risky behaviors, including illicit

drug use: Youth who drink are 7.5 times more likely to use some type of illicit drug. The most widely used illegal drug is marijuana, despite declining usage. In 2000, almost half of all 7th–12th graders reported having tried marijuana sometime during their lives, which was down from 53% in 1997. The number of these kids who regularly use marijuana declined 13% during the same time period (approximately 21% reported being regular users in 2000).

Unfortunately, the downward trends observed in marijuana use have not been noted for other types of drugs, especially so-called club drugs. For example, since 1999, teen ecstasy (MDMA) use has increased 71%. More than 8% of high school seniors reported having used ecstasy at least once. Other club drugs, including rohypnol, methamphetamine, and LSD, are also gaining popularity. These drugs have sometimes been added to the beverages of unsuspecting females in order to sedate and sexually assault them. In addition, club drugs can produce a range of negative effects, including hallucinations, paranoia, brain damage and amnesia, and, in some cases, death. Numbers of emergency room visits by young persons under the influence of club drugs, especially ecstasy, increased approximately 60% from 1999 to 2000.

Drug and alcohol use increases risks in other areas as well. Adolescents who use drugs and/or alcohol are much more likely to be sexually active than youths who do not drink or get high. Furthermore, substance-abusing teens are more likely to start having sex at an earlier age and to have multiple partners. These youth are less likely to use protection/contraception. Despite these well-documented trends, very few prevention, treatment, and counseling programs deal with the combined problem of substance abuse and risky sexual behavior.

## 2.3. Sexual Health

Research indicates that slightly more than half of females and nearly two-thirds of males have had intercourse by their 18th birthday. Reported rates of sexual activity have declined among male teens in recent years. However, this trend has not been observed for females. In fact, increases in sexual activity for female teens have been noted for each year of age during the past several decades. The greatest increases have been noted for 14- to 16-year-old females, with twice as many of these girls engaging in sexual behavior as there were 15 years ago. This is concerning for several reasons. First, the younger females are when they first have sex, the more likely they

are to have had unwanted or involuntary first sex. Furthermore, the younger the age of first sexual intercourse, the greater the risk of unwanted pregnancy and sexually transmitted diseases and infections.

It is estimated that 40% of teen sex is unprotected. Relatedly, 7 in 10 teen pregnancies occur in teens not using contraception, resulting in approximately 500,000 babies born to teen mothers each year in the United States. The majority of sexually active youth say they used some form of contraception the first time they had sex, and 57.9% of youth report they used a condom during their most recent sexual encounter.

Substantial morbidity and social problems also result from the unprotected sexual behaviors of youth. Every year, more than 3 million teens—approximately 25% of those who are sexually active—contract a sexually transmitted disease. Teens appear to be especially at risk for contracting chlamydia and gonorrhea, with reported rates for adolescent females being much higher than those for sexually active adults. These infections are the leading causes of pelvic inflammatory disease, and chlamydia is a common cause of infertility, ectopic pregnancy, and perinatal complications.

Human papillomavirus, which is associated with most cases of cervical dysplasia (a precancerous condition of the cervix), and hepatitis B virus are also prevalent among sexually active teens. Some studies suggest that up to 15% of sexually active adolescent females are infected with at least one of these conditions. Rates of teen HIV are also high. It is estimated that approximately 25% of new HIV cases each year occur in young people ages 13–21. Most of these new cases are sexually transmitted. Further complicating these issues is the fact that teens are less likely to have health insurance or a primary health care provider; thus, screening, diagnosis, and treatment of sex-related health problems are often delayed. These delays may lead to medical complications that would otherwise be avoidable.

## 2.4. Mental Health

Today, children are faced with more challenges than ever before, but it is clear that the current system of providing services is keeping the prevalence of mental disorders in check but doing little to reduce the number (incidence) of youth affected. Professionals have estimated that more than 20% of the 70.4 million children aged 17 or younger in the United States experience the signs and symptoms of a *Diagnostic and Statistical Manual of Mental Disorders* disorder during the course of a year. However, 11% experience significant

impairment and 5% experience extreme functional impairment. The surgeon general's 1999 report estimated that 20.9% of youth met diagnosable criteria ( $\leq 70$ ) on the Child Global Assessment Scale. Anxiety disorders (13.0) followed by disruptive disorders (10.3) and mood disorders (6.2) were the most common impairments (the rate is greater than 20% due to comorbidity). There is a link between child and adult disorders, but a significant number of youth “grow out” of their disorders and a significant number of adults develop a mental disorder later in life.

By examining self-report data, an argument can be made that these national estimates are low. For example, a 2001 CDC report found that the number of middle and high school students that reported an “attempted suicide during the past 12 months” was 8.8%. The percentage of students who, during the past 12 months, stated that they felt so sad or hopeless almost every day for 2 weeks or more in a row that they stopped doing some usual activities was 28.3%. The percentage of youth who reported that they had carried a weapon (gun, knife, or club) in the past 30 days was 17.4%. These behaviors and self-reports are consistent with mood disorders and disruptive disorders and appear to indicate that national estimates of incidence and prevalence are low.

On average, only approximately one-fourth of children in need of mental health care receive the help they need. The use of psychotropic medication for children and youth has increased sharply, with more than \$1 billion spent in 1998 for these medications, approximately two-thirds of which was spent on stimulants and antidepressants.

Schools are important settings in which children's mental disorders can be recognized and addressed; research suggests that schools may function as the de facto mental health system for children and adolescents. Of those receiving mental health care, 70–80% receive that care in the school setting. Furthermore, 83% of schools report providing case management for students with behavioral or social problems.

Approximately 60% of the 1500 school-based health centers in the United States have mental health professionals on staff. With support from primary care providers, nearly 80 of these clinics provide crisis intervention services. Zins and Wagner recognized the full spectrum of health services while providing numerous examples of mental health activities, such as helping with transitions from elementary to middle school and providing instruction in stress or time management.

Although there has been extraordinary progress in improving public health (including mental health)

through medical science, more is known about how to treat mental illness than about how to prevent mental illness and promote health. There are numerous research-based programs aimed at curbing at-risk behaviors, such as violence and drug use, but a rigorous applied psychology of how to promote positive youth development is lacking. However, the psychologist's role is obvious given the link between learning and health.

### 3. HEALTH PROMOTION

Today's changing world, in which political and market forces and monetary gain often dictate human behavior, is not the most favorable environment for children to develop moral and ethical values. Many children find at the end of their school years that they may be equipped with an academic qualification but not with the coping skills to face a competitive and often dangerous world. It is in this context that "health" in schools has come to occupy a place as important as learning the three R's. Zins and Wagner defined health promotion as a combination of behavioral, educational, social, spiritual, economic, and environmental efforts that support the establishment, maintenance, and enhancement of behaviors and lifestyles conducive to overall emotional and physical well-being. The definition is inclusive and systemic, and it directly encourages the mobilization of community resources and the development of environments that will lead to healthy behavioral patterns during the formative years of childhood and adolescence.

The health problems and unhealthy behaviors observed in young people are largely preventable and studies of health education in schools have demonstrated that programming has effectively reduced the prevalence of health risk behaviors among youth. Given these data, it appears that school health programs are becoming one of the most effective means available to prevent major mental and physical problems as well as significantly reduce the adverse social problems that affect children and youth. School-based clinics first began to appear in the 1980s. These early programs were problem focused (e.g., alcohol and drug use and teen pregnancy) and reported some success. For example, interventions in middle schools reported significant reductions in smoking, obesity, alcohol, and marijuana use.

Because the problem behaviors targeted by these individual programs often occur together, share risk and protective factors, and can be addressed by similar interventions, there is support for more comprehensive and coordinated models of risk prevention and

promotion of protective behaviors and environments. These comprehensive initiatives typically target multiple outcomes; span a number of years; and coordinate community, family, and school-based efforts.

Many schools have implemented programs targeting one or more categories of problem behaviors (e.g., violence and drug use). Unfortunately, adoption of multiple programs to address social and health issues can result in a variety of implementation problems. Often, multiple programs are poorly coordinated and compete with each other as well as core academics for limited instructional time and resources. Furthermore, these programs are often of short duration and are rarely sustained from year to year. They frequently lack the environmental supports at home and at school that allow for the maintenance of newly learned knowledge and skills.

Health promotion requires psychologists to be a part of a team of professionals from many fields working cooperatively with teachers, students, community members, religious leaders, and parents to provide comprehensive school health programs. These integrated efforts include such disciplines as health education, nursing, physical education, nutrition (food) services, counseling, and school psychology. The integrated approach better mobilizes community resources to enhance health-promotion efforts in the schools and society. Also, the comprehensive effort reaches virtually all children, creating an environment that builds on health in a more integrated manner.

Psychologists and other mental health professionals interested in introducing and gaining support for implementing a health-promotion program in their schools would be best served by beginning with a formal needs assessment. If such an assessment is not feasible, facilitating open discussion and active listening among staff is another strategy that can be utilized to delineate a hierarchy of concerns. After the needs have been identified, a system analysis is necessary to obtain an understanding of such things as the school board's philosophy, goals, and budget. Next, stakeholders, including school personnel, who are invested in health education should be identified. In 1995, Kelly suggested that internal support should be gathered before moving to external audiences. After input and approval of all involved school staff have been obtained, direct communication with external audiences can be initiated. Communication with community agencies, parents, churches and synagogues, hospitals, alternative health providers, and the business community is necessary to manage issues that may arise, including political groups that may perceive

health promotion as “too radical.” Furthermore, direct communication allows for the services provided to be designed around community needs.

School-based health programs are costly; therefore, those who govern funding seldom mandate such programs until serious problems arise. Federal and state health care policy is then tapped to address the remedial/tertiary concerns. A critical issue is the degree to which programs should be targeted to specific at-risk groups using selective or indicated prevention approaches or spread across all groups with no differentiation using universal prevention approaches. Research has shown universal approaches to be far more likely to be successful, produce lasting change, and reduce long-term health costs. However, universal prevention programs may not be of sufficient dosage or targeted enough to have a discernible impact on higher risk children.

Most school-based centers must seek multiple financial supports. Obvious initial sources include the local school board and the local private sector. Several states have adopted laws or officially supported the development of school-based clinics. Federal or state block grants are often used to fund such programs. Medicaid may also be a source of funds if medically necessary services are provided to children and families who are Medicaid eligible. Managed care is another funding option. School-based clinics or school-linked health programs may become certified as providers for students whose parents are enrolled in a managed care plan. Finally, the CDC provides resources to ensure that the identified curriculum to reduce risk behaviors is available for state and local agencies interested in using it. The CDC directly funds coordinated school health programs in a number of states.

#### **4. COORDINATED SCHOOL HEALTH PROGRAMS**

A health-promoting school is a school in which all members of the school community work together to provide students with integrated and positive experiences and structures that promote and protect their health. This includes both formal and informal curricula in health, the creation of a safe and healthy school environment, the provision of appropriate health services, and the involvement of the family and the wider community in efforts to promote health.

Full-service schools (schools in which health, mental health, social, and/or family services may be

located) or comprehensive health education are considered effective formats to promote healthy behaviors and prevent dysfunctional choices. However, without a radical change in school missions and society values (i.e., society choosing to directly address major issues such as poverty and unsafe environments), most of the 100,000 schools in the United States will not implement such a program in the near future. Selective or indicated prevention approaches are more consistent with current federal funding (need-based block grants) and perhaps offer the best option for the near future.

The School Health Programs has developed a comprehensive school health program model. According to the model, a comprehensive school plan provides

- A curriculum that addresses and integrates education about a range of categorical health issues at developmentally appropriate ages
- Programming for grades K–12
- Instruction for a prescribed amount of time at each grade level
- Integrative activities to help young people develop the skills they need to avoid the use of tobacco, dysfunctional behaviors, and engaging in a sedentary lifestyle
- Appropriate instruction from qualified teachers, health professionals, and concerned community members

Ideally, a school health-promotion model would be one that is holistic and that includes the interrelationships between the physical, social, emotional, and environmental (e.g., shaded play area) aspects of health. Health-promoting schools provide opportunities for families to take part in the development of health skills and knowledge of their children. Furthermore, an effective and comprehensive health-promotion model addresses the importance of the social ethos of the school in supporting a positive learning environment—an environment in which both healthy relationships and the emotional well-being of children are strengthened.

Schools, by themselves, should not be expected to address the nation’s most serious health and social problems. However, school health programs can be one of the most efficient means of preventing major health and social problems. By reducing the incidence of mental and physical health problems that afflict young people and often continue into adulthood, school health programs may reduce the spiraling cost of health care, improve educational outcomes, and increase productivity and quality of life. Schools are widely recognized as the major setting in which activities should be undertaken to promote competence and prevent the

development of unhealthy behaviors. Furthermore, from a primary and secondary prevention perspective, an economic perspective, and a demographic standpoint, it appears essential that social initiatives such as health promotion be linked to schools.

#### **4.1. Positive Youth Development**

According to Larson, high rates of boredom, alienation, and disconnection from meaningful challenge are typically not signs of psychopathology but, rather, signs of deficiency of positive youth development. Furthermore, many cases of problem behavior, such as drug use, aggression, premature sexual activity, and minor delinquency, may also be related to an absence of a positive life trajectory. Although involvement in extracurricular activities, such as sports, art, and drama clubs, has been clearly demonstrated as beneficial in reducing high-risk behaviors and increasing academic success, these advantages are not being fully promoted by schools or engaged by a significant number of today's youth. Due to a variety of factors, many schools are not fully addressing the mental and other health needs of children and youth. Unfortunately, most efforts have been focused on when and how things go wrong for youth, while neglecting the positive aspects. Indeed, today's schools are so focused on the development of academic skills and raising test scores that health promotion and the benefits of extracurricular activities are largely ignored.

After-school programs offer an alternative that can aid in the promotion of positive youth development and provide a diversion between 3:00 and 8:00 PM, the peak hours for violent juvenile crimes. The development of "assets" has been key to a number of successful after-school programs. Survey data collected by the Search Institute identified 40 developmental assets that young people need to build in order to be healthy and successful. The survey data revealed the protective power of assets: The more assets a child has, the more likely it is that he or she will do well in school, engage in community service, and value diversity, and the less likely it is that the young person will engage in risk-taking behaviors such as drug or alcohol abuse, drop out of school, or engage in violent behavior. These assets are commonsense values and experiences that help youth to grow up caring, healthy, and responsible. Applying the assets as a framework for positive youth development provides schools and communities with a set of factors that are associated with fewer risk behaviors and increased health. Furthermore, the assets provide a common language for talking about what young

people need and for pulling people together across different social groups.

Structured voluntary activities have promoted positive youth development and may be a key to successful in-school social and emotional learning. An adolescent's time is too often spent in two opposite experiential domains. In school, youth experience concentration and challenge without being intrinsically motivated. Conversely, in most leisure activities, they report experiencing intrinsic motivation but a lack of challenge and concentration. An exception to this is structured voluntary activities. Voluntary activities are those with rules, constraints, and goals as well as those in which the adolescent is highly motivated to be involved. They include sports, arts, hobbies, and participation in youth organizations. Based on a review of the literature, Larson concluded that involvement in structured voluntary activities is associated with multiple positive outcomes, including the development of initiative, greater achievement, and diminished delinquency. Thus, schools promote positive youth development when they provide opportunities for youth to engage in service learning and/or structured activities that contain the elements of intrinsic motivation, concentration, and challenge.

In recent years, a number of national reviews of school-based prevention programs have been performed to determine "best practices" in positive youth development. Using core competencies of social and emotional learning (SEL) as a basis for evaluation of programming and implementation, the Collaborative for Academic, Social, and Emotional Learning (CASEL) reviewed 81 programs that met its rigorous review standards. Following a 3-year review process, CASEL identified 22 programs that met its final criteria of outstanding SEL instruction, evidence of effectiveness, and outstanding professional development. Many of the programs selected were for all grade levels (K–12) because evidence-based/effective SEL programming typically begins at an early age and continues through high school.

## **5. SUMMARY**

Today's youth face a plethora of risk factors and complex health challenges. In addition, the most serious and expensive health and social problems, as well as the leading causes of death for both youth and adults, result from negative behavioral patterns. Schools can be a dynamic setting for promoting health and an optimal entry point for enabling children to grow into healthy adults. Furthermore, health and education are

closely related: Good health provides children with better opportunities to pursue learning. Conversely, better education empowers children to engage in healthy practices and behaviors and act as change agents to promote the health of their families and communities. Thus, there are considerable gains to be made from partnerships between health and education.

A health-promoting school provides a setting in which education and health programs come together to foster health as an essential component of life. A health-promoting school responds to the health, educational, spiritual, and psychological needs of schoolchildren in an integrated and holistic way, in which interdisciplinary mechanisms are maximized. It is within this framework that schools can better assist students in becoming knowledgeable, responsible, and healthy citizens dedicated to the common good.

### See Also the Following Articles

Alcohol Dependence ■ Bullying and Abuse on School Campuses ■ Emotional and Behavioral Problems, Students with ■ Forensic Mental Health Assessment ■ Health and Culture ■ Mental Measurement and Culture ■ Risk-Taking and Sexual Behavior ■ School Violence Prevention ■ Sexual Behavior ■ Sexual Behavior and Culture

### Further Reading

- Anderson, R. N. (2001). Deathleading causes for 1999. *National Vital Statistics Report*, 49(11), 1–88.
- Annie E. Casey Foundation. (2003). *When teens have sex: Issues and trends*. Retrieved May 30, 2003, from <http://www.aecf.org>
- Beginning too soon: Adolescent sexual behavior, pregnancy and parenthood. A review of research and interventions.* (n.d.). Retrieved June 1, 2003, from <http://aspe.os.dhhs.gov>
- Brenner, N. D., Martindale, J., & Weist, M. D. (2000). Mental health and social services: Results from the school health policies and programs study 2000. *Journal of School Health*, 7(7), 305–312.
- Center for Health and Healthcare in School. (n.d.). *Children's mental health needs and school-based services: A fact sheet*. Retrieved December 5, 2002, from <http://www.healthinschools.org>
- Centers for Disease Control and Prevention. (2001). *Youth risk behavior surveillance: United States 2001*. Retrieved December 7, 2002, from <http://www.cdc.gov>
- Collaborative for Academic, Social, and Emotional Learning. (2003). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs*. Chicago: Collaborative for Academic, Social, and Emotional Learning.
- Focus Adolescent Services. (2001). *Ecstasy (MDMA) and club drugs*. Retrieved May 30, 2003, from <http://www.focusas.com>
- Larson, R. W. (2000). Toward a psychology of positive youth development. *American Psychologist*, 52, 256–265.
- McQueen, A. (1998). *Study links teen promiscuity with drugs, alcohol*. Retrieved June 1, 2003, from <http://cyi-stars.org>
- National Youth Violence Prevention Resource Center. (2003). *Teens and alcohol*. Retrieved May 30, 2003, from <http://www.safeyouth.org>
- Partnership for a Drug-Free America. (2003). *Partnership attitude tracking study*. Retrieved June 1, 2003, from <http://www.drugfreeamerica.org>
- Payton, J. W., Wardlaw, D. M., Graczyk, P. A., Bloodworth, M. R., Tompsett, C. J., & Weissberg, R. P. (2000). Social and emotional learning: A framework for promoting mental health and reducing risk behaviors in children and youth. *Journal of School Health*, 70(5), 179–185.
- RAND. (2001). *Health research highlights: Mental health care for youth*. Retrieved December 4, 2002, from <http://www.rand.org>
- Rones, M., & Hoagwood, K. (2000). School-based mental health services: A research review. *Clinical Child & Family Psychology Review*, 3(4), 223–241.
- Satcher, D. (1999). Getting physical: Exercising our demons: Sedentary lifestyles and fatty foods leave Americans overweight and unhealthy. *Atlanta Journal Constitution*, p. A17.
- Saunders, R., & Pate, R. (2000). Promoting physical health. In K. M. Minke, & G. G. Bear (Eds.), *Preventing school problems—Promoting school success*. Bethesda, MD: National Association of School Psychologists.
- School Health Programs. (2000). *Center for Disease Control and Prevention at-a-glance 2000*. Retrieved December 7, 2002, from <http://www.cdc.gov>
- Search Institute. (2000). *Take it to heart: An asset-based guide to nurturing children, youth, and families in faith community* [Brochure]. Minneapolis, MN: Author.
- U.S. Department of Health and Human Services. (2001). *America's children: Key national indicators of well-being, 2001*. Retrieved December 12, 2002, from <http://www.childstats.gov>
- Zins, J. E., & Wagner, D. L. (1997). Health promotion. In G. G. Bear, K. M. Minke, & A. Thomas (Eds.), *Children's needs II: Development, problems, and alternatives* (pp. 945–954). Bethesda, MD: National Association of School Psychologists.



# Health Psychology, Cross-Cultural

Malcolm MacLachlan

Trinity College Dublin, Dublin, Ireland

Mesfin Samuel Mulatu

National Institute of Mental Health, Bethesda, Maryland,  
USA

1. Concepts and Definitions
  2. Theoretical Foundations
  3. Cross-Cultural Health Psychology in Practice
  4. Current Issues and Future Directions
- Further Reading

## GLOSSARY

**acculturation** A process that occurs when people from one culture encounter and react to people from another culture; it may occur at an individual (psychological) level or at a group (social) level.

**cross-cultural psychology** A specialized field in psychology with an approach that compares and contrasts individual psychological functioning in various cultural contexts.

**cultural psychology** A specialized field in psychology with an approach that roots behavior in its cultural context; it views culture and psychology as “making each other up.”

**ecocultural framework** An approach that seeks to understand cultural similarities and differences as adaptations to various human and environmental contexts.

**global health** An attempt to address health problems that transcend national boundaries, may be influenced by circumstances and experiences in other countries, and are best addressed by cooperative actions and solutions; the underlying assumption is that the world's health problems are shared problems and, therefore, are best tackled by shared solutions.

**health inequality** A term that refers to disparities or variations in health status achieved by individuals, groups, or communities; in its frequent form of use, it refers to differences in the levels of health between men and women,

between majority and minority communities, and between the rich and poor within societies.

**health psychology** A specialized field in psychology that studies the psychosocial aspects of mental and physical health and illness as well as the contexts in which these occur.

**HIV/AIDS** Abbreviation for human immunodeficiency virus/acquired immune deficiency syndrome; HIV is a retrovirus that infects cells of the immune system (primarily the CD4 cells and macrophages) and then destroys or impairs their functions, leading to immune deficiency and resulting in a disease complex known as AIDS.

**primary prevention** Health activities that are designed to reduce or eliminate the risk of acquiring a health problem; for example, interventions that encourage sexual abstinence or avoidance of intravenous drug use represent primary prevention efforts against HIV infection.

**secondary prevention** Health activities that focus on minimizing or eliminating the impact of an already acquired health problem on the affected individuals, families, and communities or on slowing or stopping the progress of a disease during its early stages; an example among AIDS patients may include efforts aimed at preventing opportunistic infections and at managing pain and psychosocial distress.

Cross-cultural health psychology is defined by examining what each of the components of this term means in its own right and by acknowledging the challenges that arise by combining them. Cross-cultural health psychology incorporates aspects of health psychology, anthropology, behavioral medicine, and public health. It is primarily concerned with

understanding health and illness from a psychological perspective, but within the context of its cultural constructions. This article reviews some relevant models that combine these individual and community perspectives. For illustrative purposes, from the vast array of health problems that are addressed by cross-cultural health psychology, the article focuses its review on HIV/AIDS due to its pandemic nature. The article considers risk factors, determinants, and prevention; disease management and impact alleviation; and social inequalities and, more broadly, impacts on population health. It concludes by arguing for a broadening of the conceptualization of what “cross-cultural” means and notes seven themes that relate culture to health. The effects of globalization and the increasing complexities of what constitutes cross-cultural health psychology present the latter with an opportunity to make a valuable contribution to the emerging field of global health.

## 1. CONCEPTS AND DEFINITIONS

In defining cross-cultural health psychology, one needs to do so reflexively by unpacking each of the words in the term. If one takes “psychology” as the study of human behavior, one must qualify this further by acknowledging that “psychology” has mostly been understood to be synonymous with a particular perspective on human behavior. This perspective is characteristic of a broader rationalist, reductionist, and individualist approach to advancing knowledge that predominates in Northern and Western conceptions of what are valid means of truth seeking. This is problematic for psychology because such an approach is, in itself, a cultural construction. However, other psychologies, curiously referred to as “indigenous,” relate different conceptions of how one ought to account for human behavior. Thus, one first needs to acknowledge that there are different—culturally constructed—conceptions of just what psychology is and how it should study human behavior.

In some ways, “health” is easier to define, if only because the World Health Organization’s (WHO) definition is rather broad and inclusive in emphasizing biological, psychological, and social aspects of health and in acknowledging that health and illness are not “all or none” categories. All people are healthy or ill to some extent, and even those who suffer greatly through physically disabling diseases may be quite healthy in

other respects, whereas some icons of physical excellence and beauty may be quite ill in some ways. Of course, the way in which distress is expressed may differ among cultural contexts (e.g., by psychologizing or somatizing), as may the understanding of what it means to be healthy or ill and the ways in which to respond to illness.

Therefore, “health psychology” may be taken to be the study of how human behavior influences health, broadly defined. “Health” here refers to well-being in general and is not confined to either physical complaints or mental complaints. Health psychology seeks to bridge the Cartesian divide by acknowledging that human experience is mediated through the body and, thus, is embodied. The biopsychosocial model, so often identified with health psychology, should be seen not as multiple but rather as integrative, that is, as being of one—body and mind together. Furthermore, the contextual aspects of health, in terms of its social settings and conditions, should also be to the fore in health psychology.

“Cross-cultural” psychology must always first be “cultural” psychology for it to be meaningful. Cultural psychology seeks to understand how behavior is influenced by the social context in which it occurs. It further acknowledges that this context is woven through particular customs, rituals, beliefs, ways of understanding and communicating, and so on, so that distinctive patterns of behavior are cultivated. Only by understanding how a culture patterns meaning can we be sure to know that the sort of things we might want to compare among different cultures actually have some similarity. Having established a meaningful similarity in the structure or function of aspects of human behavior, it may then be enlightening to compare such behavior in different cultural contexts, that is, across cultures. Cross-cultural psychology may incorporate both the localized cultural perspective and the comparative cultural perspective.

In sum, cross-cultural health psychology seeks to describe the diversity of health (and illness) behavior and to link this to the cultural context in which it occurs. It recognizes that culture and psyche “make each other up” and that the ways in which they do this have implications for health. The value of cross-cultural health psychology is becoming increasingly recognized not only in the treatment of health problems that present in specific cultural contexts but also in how social and cultural factors must be understood for the prevention of illness and the promotion of health.

## 2. THEORETICAL FOUNDATIONS

Cross-cultural health psychology borrows theoretical and practical approaches from a wide variety of related disciplines, including the broader fields of psychology, anthropology, behavioral medicine, and public health. Health psychology and behavioral medicine contribute the biopsychosocial model of health and disease. A recently revised version of this model argues for the reorientation of the fields of health and medical sciences such that research and clinical practice encompass both health and illness. The model aims to decipher the interactive and reciprocal relationships among biophysical, psychological, and social processes in health and disease. It takes into account the social networks of the individual to achieve, maintain, and maximize health and well-being for individuals, their significant others, and society. The major challenge for this model is for its biological, psychological, and social aspects to truly act in synthesis rather than as simply parallel and complementary aspects of health and illness.

Cross-cultural psychologists, anthropologists, and others argue that current theoretical models and practice in medicine and health psychology are insufficient to capture the centrality of culture to the human experiences of health and illnesses. They argue for theoretical frameworks that recognize the ubiquitous effect of culture on health and illness; that treat culture as a system of meaning that links body, mind, and world; and that promotes the significance of cultural context as a resource. Frameworks that give greater weight to cultural and ecological contexts and process as they relate to health and health promotion have been suggested. For instance, Berry's ecocultural framework provides a scheme with which factors at both the individual (psychological) and community (cultural) levels are considered in cross-cultural application of research in cognitive, affective, behavioral, and social aspects of health phenomena. The community level of work typically involves ethnographic methods and yields a general characterization of shared health conceptions, values, practices, and institutions in a society. The individual level of work involves the psychological study of a sample of individuals from the society and yields information about individual differences (and similarities) that can lead to inferences about the psychological underpinnings of individual health beliefs, attitudes, behaviors, and relationships. There are important reasons why cultural- and individual-level factors need to be taken into account when dealing with health phenomena.

Cultural-level health phenomena are taken into account to establish the broad context for the development and display of individual health phenomena. Without an understanding of this background context, attempts to deal with individuals and their health problems may well be useless or even harmful; individuals are not "culture free." The reason for considering individual-level health phenomena is that not all persons in a cultural group hold the same beliefs or attitudes, nor do they engage in the same behaviors and relationships. Without an understanding of people's individual variations from the general community situation, harm may well be inflicted again; individuals are not "cultural dopes."

Health models that take into account broader cultural- and ecological-level factors have also been proposed in the field of public health. For instance, Airhihenbuwa proposed a theoretical model that places cultural and environmental factors at the core of the development, implementation, and evaluation of health programs. His PEN-3 model consists of three dimensions of health beliefs and behaviors: (a) health education, (b) educational diagnosis of behavior, and (c) cultural appropriateness of health behavior. Each of these three interdependent dimensions is influenced by three categories of factors (representing the acronym "PEN"). Accordingly, health education needs to be designed to address the health needs of all, involving the person, the extended family, and the neighborhood or community in which the person lives. Airhihenbuwa noted that educational diagnosis of health behavior helps researchers and program developers to determine the factors that influence individual, family, and/or neighborhood and community health actions. Critical elements that need to be assessed as part of such a diagnosis include perceptions, enablers, and nurturers. Health-related perceptions include attitudes, values, and beliefs within a cultural context that facilitate or hinder health-promoting or disease-preventing actions. Enablers are distal cultural, societal, systemic, or structural factors that may enhance or be barriers to health actions, whereas nurturers are relatively proximal influences on health behaviors originating from extended family, kin, friends, and the neighborhood. Finally, the most critical dimension of developing effective health programs is the determination of the cultural appropriateness of the health behavior. Within a given cultural context, some health beliefs and behaviors can be positive, endowing the person, the family, and/or the community with known health benefits. Other health beliefs and behaviors may be existential, having no harmful health consequences

but bearing important indigenous meaning to the community. Still others may be negative health beliefs and behaviors, contributing to the origin or exacerbation of health problems. Such an assessment of the cultural appropriateness of health beliefs and behaviors empowers health program developers to consider the cultural perspectives, knowledge base, and resources of the people for whom the programs are designed.

A relatively more extensive framework for planning health promotion programs is the PRECEDE/PROCEED model developed by Green and Kreuter. This model takes into account the myriad of factors that determine the health and quality of life of individuals and communities in planning, implementing, and evaluating health programs. The PRECEDE part of the model (the acronym stands for predisposing, reinforcing, and enabling constructs in educational/ecological diagnosis and evaluation) involves various phases during which the determinants of health are diagnosed. It begins with a social assessment to identify quality of life and concerns of the target population, followed by assessment of health problems and behavioral and environmental factors linked to these health problems. Then, the predisposing, reinforcing, and enabling factors that influence health behaviors are assessed during the educational and ecological assessment phase, followed by an assessment of the organizational and administrative capabilities and resources for the development and implementation of health programs to address the problems identified. The PROCEED part of the model (the acronym stands for policy, regulatory, and organizational constructs in educational and environmental development) offers additional steps for developing policy and initiating the implementation and evaluation process appropriate to the health outcomes, intervention types, and target populations. The PRECEDE/PROCEED model, like PEN-3, recognizes that health is determined by multiple factors and that interventions need to take multilevel and multidimensional actions to be effective.

More specific individual-level health behavior theoretical models (e.g., Health Belief Model, theory of planned behavior, social cognitive theory) are often used in health psychology and public health to explain and predict risk and protective behaviors. Embedded in these models is the belief that risk behaviors, cognitions, and their determinants are identifiable, modifiable, and causally linked to health outcomes. In general, these health behavior models propose that one or a combination of the following factors predict the adoption and maintenance of changes in health behaviors: (a) knowledge and information about the health threat; (b) perceived severity of

and personal vulnerability to the threat; (c) perceived control or self-efficacy to adopt the behavior change; (d) perceived efficacy of the behavior change to prevent the threat; (e) perceived benefits compared with costs associated with behavioral change; (f) intentions, attitudes, and social norms supporting the adoption of the behavioral change; (g) actual skills to bring about the behavioral change; and (h) availability of cues for behavioral change. Recognizing the limitation of individual-focused factors to behavior change, other theoretical models in public health and related disciplines have expanded this list by adding variables at the interpersonal (e.g., couple communication, power in relationships), community (e.g., supportive norms, social networks), and environmental and structural (e.g., accessibility services, policies) levels that have strong bearings on individual-level risk behaviors and cognitions.

In sum, combining applied individual- and community-level methodologies from clinical psychology and public health and the conceptual underpinnings from interactive biopsychosocial and ecocultural frameworks, cross-cultural health psychology focuses on factors that determine and maintain health and health outcomes in cross-cultural and multinational contexts. This approach presupposes consideration of multiple causal pathways and levels of analysis in research and practice of health psychology. Psychosocial, cultural, and environmental factors are believed to be causally linked to health directly through exposure to noxious and disease-causing conditions such as carcinogens and pathogens. Indirectly, psychosocial stress associated with adverse psychosocial, cultural, and environmental conditions influences health by affecting the performance of health-related behaviors as well as the functioning of the central nervous system and the endocrine system, primarily immune and cardiovascular functioning. Overall, it is believed that biological, psychological, sociocultural, and environmental processes interactively affect, and are affected by, health and illness.

### **3. CROSS-CULTURAL HEALTH PSYCHOLOGY IN PRACTICE**

Applied health psychology consists of several research and practice areas that overlap with the broader fields of clinical psychology, behavioral medicine, and public health. In cross-cultural contexts, three areas of research and practice have been particularly predominant: (a) identifying health risk factors and their determinants

and taking preventive actions, (b) alleviating and managing the impacts of diseases, and (c) addressing social inequalities and population health. Although a vast range of health problems could be considered here, this section discusses each of these three areas by giving pertinent examples from the literature on human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS) due to its pandemic nature. Extensive coverage of cross-cultural applications of health psychology in understanding and tackling other major global health problems, including malaria, tuberculosis, nutritional deficiencies, and noncommunicable diseases, can be found elsewhere.

### **3.1. Risk Factors, Determinants, and Prevention**

Before focusing exclusively on HIV/AIDS-related issues, the first two paragraphs of this section briefly consider the spectrum of disease and mortality globally. According to *The World Health Report 2002* published by the WHO, the leading causes of death in high-mortality “developing” societies (e.g., Sub-Saharan Africa, Latin America, Asia) are communicable diseases and maternal and child conditions, including HIV/AIDS, tuberculosis, malaria, respiratory infections, and childhood and maternal undernutrition. In highly “developed” countries (e.g., North America, Western Europe, Pacific Rim), the leading causes of death are noncommunicable diseases, including various forms of cardiovascular diseases, cancers, and psychiatric disorders. In low-mortality developing countries (e.g., Middle East, Eastern Europe, Southeast Asia), a mix of communicable and noncommunicable diseases and maternal and childhood conditions characterize the leading causes of death.

These differences in leading causes of death among countries and regions have been attributed to differences in risk factors to which individuals in these societies are exposed. This is a subject of interest in cross-cultural health psychology research and practice. According to the WHO report, in the high-mortality developing societies, childhood and maternal undernutrition, unsafe sexual activities, indoor smoke from solid fuels, and unsafe water, sanitation, and hygiene contribute the most to the total burden of ill health. In contrast, in highly developed countries, tobacco, high blood pressure, alcohol, high cholesterol, and high body mass index play major roles in the loss of health and functioning. In low-mortality developing countries, a mixed pattern of risk factors found in both high-mortality developing and

low-mortality developed countries seem to contribute significantly to disability and death. Two important themes appear to characterize this report. First, cross-cultural and cross-national differences in health and disease are associated with socioeconomic development and geopolitical situations of the countries and regions. Second, the single most important causes of burden of disease and disability in both developing and developed countries are modifiable health behaviors or conditions that are responsive to appropriate health measures.

Focusing on HIV/AIDS, the most extensively studied risk factor is unprotected sexual intercourse. With the rapid spread of HIV/AIDS, other sexually transmitted infections, and other reproductive health problems, increased attention has been given to understanding the patterns and determinants of sexual risk behaviors and to designing and implementing effective interventions. Although HIV as the cause of AIDS and risk factors for its transmission (e.g., unprotected sex, contaminated blood transfusion, mother-to-child transmission, unsterilized injections) have been known since the early 1980s, controlling its spread has been a formidable challenge to public health across the world, particularly in Sub-Saharan African countries where the epidemic has hit the hardest.

Health psychologists working in a cross-cultural context have applied health behavior and social psychological theories in both research and applied settings. In research, health psychologists are involved in the design and implementation of rapid assessment techniques to measure the level of knowledge, attitudes, and practices of individuals in general or of those who are considered to be particularly at risk for infection. These efforts have become expanded and standardized and are being used in multiple national surveys on a regular basis with the intent to collect comparable risk and protective data. An important example in these efforts is the Demographic and Health Surveys, a worldwide initiative supported by the U.S. Agency for International Development (USAID) and governments of several developing countries with the aim of providing comparable data and analysis on population health, including HIV/AIDS. These studies, and similar other studies, help to determine the levels of awareness and attitudes about HIV/AIDS and the extent of risk behaviors for HIV infection. Based on these data, it is known that awareness about HIV/AIDS and its primary routes of infection is high in nearly every country but that high-risk behaviors continue to be practiced to varying degrees in both developed and developing countries. For example, unprotected heterosexual intercourse is a significant means of HIV spread in

Sub-Saharan Africa, whereas unprotected homosexual intercourse and injecting drug use are major sources of infection in Western countries.

Research has also focused on the determinants of sexual risk (e.g., sex with multiple partners) or protective behaviors (e.g., condom use) using general health behavior theories (e.g., theory of planned behavior) or those that are developed specifically to address HIV/AIDS-related risk (e.g., AIDS Risk Reduction Model). According to a recent report on the epidemic by the UN Program on HIV/AIDS (UNAIDS), sexual risk taking or inconsistent condom use is closely intertwined with a variety of personal and interpersonal psychosocial factors, including poor knowledge, a low sense of threat to infection, high sensation seeking, and a strong need to comply with social norms, as well as with group- or community-level risk factors, including poverty, migration, and pervasive norms that perpetuate women's subordination within sexual relationships. In the context of Sub-Saharan Africa, negative attitudes toward sex with condoms, female sexual compliance to ensure economic security, a lack of negotiation skills to implement condom use, a desire for children, and an external locus of control have been identified as potential obstacles to condom acceptance and use. The relative strength of individual-, interpersonal-, community-, and structural-level factors varies from society to society, depending on the level of socioeconomic development, the spread of the disease, and the sociopolitical climate of the country. However, it is now recognized that sexual risk and protective behaviors, like other health behaviors, are determined by multiple factors and imply that interventions need to address many factors to be successful.

Interventions against risky sexual behaviors often teach recipients about facts related to HIV and reproductive biology, encourage positive attitudes toward those affected and infected by HIV/AIDS, instill skills in sexual negotiation and in the acquisition and use of condoms, and make condoms and voluntary testing and counseling more accessible. Given that sexual beliefs, attitudes, and activities are strongly influenced by cultural and broader structural factors, some interventions have addressed these factors appropriately. Most interventions include a combination of behavioral outcomes, theoretical approaches, and formats and methods of delivery. In some countries and regions, some formats are more acceptable and amenable than others. For instance, in Sub-Saharan Africa, radio and stage dramas are often used for communicating preventive messages. In more literate societies, print and other mass communication approaches are used.

Evidence from both more and less industrialized countries indicates that behavioral and psychosocial interventions are effective in enabling people to make and maintain necessary preventive sexual behavior changes. Reviews of these interventions generally conclude that psychosocial and behavioral interventions are effective in increasing awareness, skills in sexual negotiation, frequency of communications with sexual partners about safer sex, and skills in the acquisition and use of condoms as well as in reducing frequency of sexual intercourse, early sexual initiation among youth, and the incidence of sexually transmitted infections. Although intervention effectiveness varies by the type, length, and other characteristics of the intervention, the most efficacious HIV prevention programs have been directed specifically at groups at risk for infection, focused on relationship and negotiation skills, involved multiple sustained contacts, and used a combination of culturally appropriate media of delivery. Although most intervention studies have not explicitly studied the role that cultural variables play within cross-cultural contexts, those that have integrated cultural factors into their HIV prevention themes and contents have confirmed the pervasive influence of culture on risk perception, risk-taking behaviors, and adoption of protective behaviors.

### **3.2. Disease Management and Impact Alleviation**

Advances in health sciences, care, and technology have created important new challenges and roles for cross-cultural health psychologists in the area of secondary prevention. Many of the acute and chronic medical problems facing the world, and the curative or palliative treatments for these problems, have profound impacts on numerous areas of patient functioning and quality of life. These impacts may involve physical and psychosocial disabilities and may originate directly from the health problems themselves, from treatment side effects, or indirectly as a consequence of the disabilities or limitations they impose on patients and families. In general, medical conditions present psychosocial challenges for patients and their families in the areas of (a) coping with and managing pain, (b) coping with psychosocial impacts of disease, (c) adhering to demanding medical regimens, and (d) coping with social stigma and discrimination.

Chronic health problems, such as HIV/AIDS, are accompanied by persistent pain and other psychosocial difficulties in the form of emotional distress, including

anxiety, depression, low self-esteem, and irritability, as well as disruptions in interpersonal relationships, including marital difficulties and family stress. Pain and other psychosocial consequences of diseases must be addressed not only because they contribute to diminishing quality of life of the patients and their families but also because studies have documented that psychosocial distress is related to length of survival, immunological functioning, and/or some measures of disease progression. Sources of persistent pain and psychosocial distress among people with chronic diseases are multiple. Among people living with HIV/AIDS, psychosocial problems occur as a consequence of the physical illness, disability, and wastage and also as side effects of the multitudes of medication taken. Neurological and cognitive impacts of HIV infection may also affect the day-to-day functioning of patients. Negative attitudes and reactions of others toward patients contribute to patients' sense of isolation and psychosocial distress. Furthermore, socioeconomic strains caused by the loss of jobs and incomes add to the psychosocial distress that the patients and their families experience. The actual level of psychosocial distress experienced by people living with HIV/AIDS depends on several factors, including the resources available to patients and their families, socioeconomic development of their countries, and the strength of national policies protecting the rights of HIV-infected people. For instance, the potential loss of jobs, housing, family support, and even life due to admitting to HIV infection is a major concern and source of distress in many developing societies, although such concerns are significantly reduced in Western societies.

Interventions against persistent pain and psychosocial distress due to HIV/AIDS involve multidisciplinary teams and a variety of components and strategies, including individual psychotherapy, social support in a group setting, cognitive-behavioral skills training, and pharmacological treatment to reduce and manage pain, anxiety, and depressive symptoms. Self-management interventions are also used when the aim is to help patients monitor and cope with their pain and illness and to maintain satisfactory quality of life. These interventions include providing information about treatment conditions, management of symptoms and psychological consequences, and communication and coping strategies such as relaxation and meditation.

Reviews of studies from various countries and using various outcome measures have found strong support for interventions against persistent pain and psychosocial consequences of HIV/AIDS. Although the positive effects of these interventions are not always consistent,

and it is difficult to discern the relative efficacy of their components, there is sufficient evidence that interventions against psychosocial consequences of diseases can affect quality of life as well as disease outcomes. Consistent with the biopsychosocial and ecocultural perspectives of both pain and psychosocial distress, there is strong evidence that culturally appropriate interventions that take into account sociocultural resources, explanatory models, and contextual factors are more likely to effectively reduce persistent pain, psychosocial distress, and negative disease outcomes. For instance, the use of traditional healers, midwives, and indigenous social support systems in developing countries and the use of "alternative and complementary therapies" in the West have been found to be potentially useful interventions for managing pain and for mitigating the psychosocial impacts of HIV/AIDS.

Low adherence to prescribed treatments is a ubiquitous problem among patients with chronic conditions, thereby contributing to undermining the potential benefits that may accrue from treatment. Adherence to highly active antiretroviral therapy (HAART) and other anti-HIV medication has been studied in various cultural settings. In these studies, health psychologists are involved in examining the reasons for nonadherence and in designing interventions to decrease its likelihood. Studies from several countries and cultures have identified several factors that are correlated with nonadherence. Poor adherence tends to occur in treatment regimens that are complex, have strong side effects, and place too many cognitive and scheduling demands. Psychosocial factors, including emotional distress, low self-efficacy beliefs, substance abuse, pessimism about prognosis, lack of social or provider support, and limited access to health service resources, have also predicted lower levels of adherence. Thus, both individual-level variables and cultural constructions regarding the meaning of AIDS are important regarding adherence to treatment regimens.

Interventions against nonadherence with anti-HIV medication involve a combination of cognitive, behavioral, and psychosocial components. Health psychologists, along with other health care professionals, provide structured dosing instructions, self-monitoring and problem-solving skills, behavioral reminders and strategies to increase compliance with medical advice, counseling and social support in either individual or group settings, and information about drug options, interactions, and side effects. Studies that evaluated these interventions showed that patients who participated in interventions had higher levels of adherence and lower levels of viral

load than did patients who took standard care, although these findings are not always statistically significant and long-lasting. In addition, it has been noted that comprehensive interventions that combine educational, behavioral, and affective components have led to better intervention outcomes than have single-focused interventions. Effective strategies are also characterized by effective communications with patients that are tailored to individual patients' needs, capacities, and resources.

Stigma refers to undesirable or discrediting socially constructed attributes that individuals possess and that affect their quality of life as well as their social standing. Perceived or felt stigma, as opposed to enacted stigma or discrimination, involves real or imagined fear of negative social attitudes and discrimination due to one's undesirable personal attributes or association with others who have such attributes. Stigma is a powerful force that often causes prejudicial thoughts, behaviors, and actions by individuals, groups, communities, and institutions. Although nearly all health conditions can lead to stigmatizing experiences, diseases that elicit fear of illness, contagion, and death, as well as those whose origins are uncertain and whose treatments are limited, tend to cause a relatively greater amount of stigmatization. Patients with diseases such as leprosy and tuberculosis have historically evoked fear and revulsion in others. Although advances in knowledge and treatment have increased acceptance of such patients, other more recent health conditions, such as HIV/AIDS, are currently among the most stigmatizing health conditions around the world.

Among HIV-infected individuals, stigma manifests itself in many ways. HIV-infected individuals may refuse to disclose their status to others due to fear of negative reactions from their partners, family members, employers, and community members. This has been found to be a major obstacle to preventive efforts in many Sub-Saharan African countries. These individuals may also refuse to receive the necessary health services due to potential negative reactions of health care providers or to fear of being discriminated against by others. HIV-infected people who belong to certain sectors of society, such as commercial sex workers in Sub-Saharan Africa and injecting drug users in Western countries, may face an extra layer of stigmatization due to their association with "deviant" or "immoral" behaviors. These and other stigmatizing experiences lead to negative effects on test-seeking behavior, willingness to disclose HIV status, health service-seeking behavior, quality of health care received, and social support solicited and received.

Reviews of intervention studies from the United States, Europe, and Sub-Saharan Africa indicate that antistigma efforts aim at increasing the tolerance of individuals and communities for people living with HIV/AIDS, the willingness of health care providers to treat people with HIV/AIDS, and the abilities of people living with HIV/AIDS to cope with stigma and discrimination. Interventions have used a combination of strategies, including providing information about HIV/AIDS and impacts of stigma, building skills to diffuse and cope effectively with conflict situations, counseling to address individual concerns and provide social support, and providing opportunities for public contact with people infected with, and affected by, HIV/AIDS. The review indicates that interventions that combine information, skill building, and contact with people living with HIV/AIDS reduce stigma significantly more than do interventions that include any one of these separately. Not all studies have brought appreciable reductions in stigma, due partly to limitations in the conceptualization, measurement, and methodological problems associated with the interventions.

### **3.3. Addressing Social Inequalities and Population Health**

Recently, research and practice in social inequality in health has gained ground in health psychology and public health across nations. There is a growing awareness that socioeconomic hierarchies, defined by a wide variety of indexes such as social class, accumulated wealth, occupational prestige, income, and education, and negative social attitudes and actions associated with those hierarchies, including racism, discrimination, and stigmatization, have been linked to differences in levels of health outcomes. Major national and multinational studies indicate that unfavorable outcomes in many areas of health are relatively more prevalent among people occupying the lowest socioeconomic status and/or among historically marginalized ethnic groups and immigrants. For instance, African and Hispanic Americans in the United States, Bangladeshi and Pakistanis in the United Kingdom, and Africans and Asians in South Africa tend to have unfavorable socioeconomic standing as well as poorer overall health than do White people in their respective countries. Health psychologists, as well as other professionals, are involved in efforts directed at understanding the mechanisms through which these effects occur and the manner in which these social inequalities in health can be reversed.

Research on ethnic densities has found that the larger the number of people in a minority cultural group living within a given area, the lower the psychiatric admission rate for members of that group, suggesting that living in a neighborhood that encourages an unfamiliar lifestyle, culture, and language is a risk factor for the development of mental disorder. Such segregation may help to protect group members from direct prejudice and provide them with social support that operates through culturally familiar customs. Research into the broader mechanisms of socioeconomic differences in health attributes their origins to two broad factors: material deprivation and psychosocial disadvantage. Included in the material explanation of social inequalities in health are access to basic necessities of life (e.g., food, shelter), access to amenities that contribute to a comfortable and productive life (e.g., home ownership, means of transportation), and access to curative, preventive, and/or health-promoting services (e.g., health care, exercise facilities). The psychosocial interpretation of health inequalities proposes direct and indirect effects of psychosocial stress associated with occupying lower socioeconomic status. Stress and emotional distress associated with racism, discrimination, and living in an unfavorable social position may relate to health directly by affecting physiological systems or indirectly by affecting health behaviors.

Efforts to address social inequalities in health and to promote population health involve a wide array of interventions to increase disease-preventing/health-promoting resources of individuals, strengthen capacities of communities, improve access to essential facilities and services, and encourage macroeconomic and sociocultural changes by way of conducive policies and government actions. A strong argument in favor of bringing changes in social policies and government programs that directly address the fundamental social causes of social inequalities in health is critical. For instance, structural-level interventions focusing on increasing accessibility of HIV/AIDS preventive and treatment services and on enacting favorable national policies have shown considerable potential in addressing the disproportionate burden of HIV/AIDS among ethnic and racial minorities and vulnerable groups in both resource-poor and resource-rich countries. Cross-cultural health psychology contributes to knowledge about behavior change approaches that incorporate culturally appropriate strategies. It can also bring forth knowledge about the social world and community-level variables (e.g., supportive social networks, social cohesion) that contribute to resilience, wellness, and health. Social and

health policies that are evidence based and socioecologically appropriate are more likely to be effective in reducing social inequalities and improving individual and population health.

#### 4. CURRENT ISSUES AND FUTURE DIRECTIONS

Traditionally, cross-cultural health psychology has sought to compare health in various cultural contexts, where those contexts either were geographically distant or involved migration from one geographical area into the culture of another geographical area. Although these fields of study are still of interest and relevance, perhaps a broader conceptualization of what “cross-cultural” means is required. Furthermore, it is important to recognize that cultural comparisons are naive if they fail to take account of the broader sociopolitical and economic context. This is particularly relevant to health when cultural differences are compounded by the disempowerment of one group relative to another. For instance, in thinking of the practical implementation of health promotion initiatives in Sub-Saharan Africa, it is clear that traditional indigenous approaches to health are not operating on an equal power base with more Western-oriented biomedical interventions. Aid agencies that provide resources for health care tend to support those perspectives that reflect their own cultural assumptions. Although this is not necessarily wrong, cross-cultural health psychology must be aware of this bias and power imbalance to be truly informative for effective action.

MacLachlan identified seven realms where culture and health are related to each other and where these relationships have implications for people’s sense of identity, empowerment, and well-being (Table I). These different realms reflect a range of new challenges for cross-cultural health psychology, where “cross-cultural” may be taken to mean between different realms of culture. Hermans and Kempen argued that cross-cultural psychology’s traditional emphasis on geographical location is becoming increasingly irrelevant. Instead, they emphasized the influence of interconnecting cultural and economic systems, of cultural hybridization creating multiple identities, and of increasingly permeable cultural boundaries. Most contemporary societies are patterned by a multiplicity of meanings and practices that likely contribute to greater uncertainty. As Hermans and Kempen stated, “Uncertainty is not primarily in a culture’s core but in its contact zones.”

**Table 1**  
**Realms within Which Culture and Health Interact**

---

<b>Cultural colonialism</b>
Rooted in the 19th century, when Europeans sought to compare a God-given superior “us” with an inferior “them” and to determine the most advantageous way of managing “them” in order to further European elites.
<b>Cultural sensitivity</b>
Being aware of the minorities among “us” and seeking to make the benefits enjoyed by mainstream society more accessible and modifiable for “them.”
<b>Cultural migration</b>
Taking account of how the difficulties of adapting to a new culture influence the opportunities and well-being of geographical migrants.
<b>Cultural alternativism</b>
Different approaches to healthcare offer people alternative ways of being understood and of understanding their own experiences.
<b>Cultural empowerment</b>
As many problems are associated with the marginalization and oppression of minority groups, a process of cultural reawakening offers a form of increasing self and community respect.
<b>Cultural globalization</b>
Increasing (primarily) North American political, economic, and corporate power reduces local uniqueness and reinforces and creates systems of exploitation and dependency among the poor, throughout the world.
<b>Cultural evolution</b>
As social values change within cultures, adaptation and identity can become problematic, with familiar support systems diminishing and cherished goals being replaced by alternatives.

---

After MacLachlan (2004b).

Conventionally, cross-cultural health psychology has recognized the interplay between culture and health in colonialism, through sensitivity to minority groups, and in the acculturation issues of migrants. The recent enthusiasm for alternative medicine has also attracted the interest of some cross-cultural health psychologists because it can be seen as introducing greater pluralism into health care. However, less noticed by cross-cultural health psychologists are the distinctive (and often more complex) relationships between culture and health that are evident in the renewed enthusiasm for cultural empowerment, especially among marginalized indigenous groups. Another complex issue is the health effects of globalization, where a singular politicoeconomic system based on North American values, with

accompanying ontological assumptions about the causes of and cures for health problems, is in conflict with local practices and beliefs. Yet another challenge for cross-cultural health psychology is how rapid social and cultural changes within societies may bring in their wake what might be termed “disorders of acculturation.” For instance, cultural evolution and social change may require a process of temporal acculturation that has significant implication for health.

Although the realms described here may require cross-cultural health psychologists to rethink some of their discipline’s traditional assumptions, some aspects of traditional health psychology will be applicable in the “newer” realms where culture and health interplay. For instance, Berry’s framework for conceptualizing the acculturation strategies of (geographical) migrants has also been found to have predictive value for the well-being of people experiencing temporal acculturation through a period of rapid social change. There is no doubt that cross-cultural health psychology offers a valuable comparative perspective on the human experience of health across cultures. As an orientation, it has much to offer. As a methodology, it is challenged by a greater need for cultural authenticity than for universal applicability. However, a comparison of culturally sensitive efforts can identify generally applicable principles, although just how those are enacted must again be meaningfully translated into the matrix of local living.

Kazarian and Evans concluded that although health psychologists have made important contributions to developing a more holistic approach to health, they have not effectively integrated culture into their approach. Their emphasis on the role of cultural health psychology in the creation of culturally competent health professionals is timely given the recent enthusiasm for a broader conceptualization of health, referred to as “global health.” Global health is an attempt to address health problems that transcend national boundaries, may be influenced by circumstances and experiences in other countries, and are best addressed by cooperative actions and solutions. The underlying assumption is that the world’s health problems are shared problems and, therefore, are best tackled by shared solutions. As Lee wrote, “Health care systems are traditionally structured, regulated, and delivered at the national level. Health research is correspondingly focused on the state, sub-state and inter-state levels. . . . There is a clear desire for a fuller understanding of the links between globalisation and health, beginning with the development of relevant conceptual frameworks and rigorous empirical analysis.”

The health impact of globalization with regard to HIV/AIDS, coronary heart disease, and many other public health problems can be all the better appreciated by a comparative analysis—both conceptual and empirical—of how these difficulties affect people, and are affected by people, in a cultural context around the world. Beyond the needs for research, analysis, and policy development to which cross-cultural health psychology should contribute, there is also enthusiasm for new curricula that place global health center stage in the education of health professionals. With the increasing complexities of the contexts that may now be considered to constitute “cross-cultural” health psychology, and with greater recognition of the persuasive influence of global forces on human health, the value of cross-cultural health psychology may be increasingly realized in its contribution to addressing the challenges of global health.

### See Also the Following Articles

Acculturation ■ Cross-Cultural Psychology, Overview  
 ■ Health and Culture ■ Homosexuality

### Further Reading

- Aboud, F. E. (1998). *Health psychology in global perspective*. Thousand Oaks, CA: Sage.
- Adler, N. E., & Ostrove, J. M. (1999). Socioeconomic status and health: What we know and what we don't. *Annals of the New York Academy of Sciences*, 896, 3–15.
- Airhihenbuwa, C. O. (1995). *Health and culture: Beyond the Western paradigm*. Thousand Oaks, CA: Sage.
- Barlow, J., Wright, C., Sheasby, J., Turner, A., & Hainsworth, J. (2002). Self-management approaches for people with chronic conditions: A review. *Patient Education & Counseling*, 48(2), 177–187.
- Berry, J. W. (1997a). Cultural and ethnic factors in health. In A. Baum, S. Newman, J. Weinman, R. West, & C. McManus (Eds.), *Cambridge handbook of psychology, health, and medicine* (pp. 98–103). Cambridge, UK: Cambridge University Press.
- Berry, J. W. (1997b). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review*, 46, 5–68.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology*. Cambridge, UK: Cambridge University Press.
- Blankenship, K. M., Bray, S. J., & Merson, M. H. (2000). Structural interventions in public health. *AIDS*, 14(Suppl. 1), S11–S21.
- Brown, L., Macintyre, K., & Trujillo, L. (2003). Interventions to reduce HIV/AIDS stigma: What have we learned? *AIDS Education & Prevention*, 15(1), 49–69.
- Campbell, C. A. (2002). Prostitution, AIDS, and preventive health behavior. *Social Science & Medicine*, 32, 1367–1378.
- Fogarty, L., Roter, D., Larson, S., Burke, J., Gillespie, J., & Levy, R. (2002). Patient adherence to HIV medication regimens: A review of published and abstract reports. *Patient Education & Counseling*, 46(2), 93–108.
- Glanz, K., Lewis, F. M., & Rimer, B. K. (Eds.). (1997). *Health behavior and health education* (2nd ed.) San Francisco: Jossey-Bass.
- Green, L. W., & Kreuter, M. W. (1999). *Health promotion planning: An educational and ecological approach* (3rd ed.). Mountain View, CA: Mayfield.
- Halpern, D., & Nazroo, J. (2000). The ethnic density effect: Results of a national community survey of England and Wales. *International Journal of Social Psychiatry*, 46, 34–46.
- Hawe, P., & Shiell, A. (2000). Social capital and health promotion: A review. *Social Science & Medicine*, 51, 871–885.
- Hermans, H. J. M., & Kempen, H. J. G. (1998). Moving cultures: The perilous problems of cultural dichotomies in a globalizing society. *American Psychologist*, 53, 1111–1120.
- Kazarian, S. S., & Evan, D. R. (Eds.). (2001). *Handbook of cultural health psychology*. San Diego: Academic Press.
- Lee, K. (2003). *Health impacts of globalisation*. Basingstoke, UK: Palgrave Macmillan.
- Lindau, S. T., Laumann, E. O., Levinson, W., & Waite, L. J. (2003). Synthesis of scientific disciplines in pursuit of health: The Interactive Biopsychosocial Model. *Perspectives in Biology & Medicine*, 46(Suppl. 3), S74–S86.
- MacLachlan, M. (Ed.). (2001). *Cultivating health: Cultural perspectives on promoting health*. Chichester, UK: Wiley.
- MacLachlan, M. (2004a). *Culture and health: A critical perspective* (2nd ed.). Chichester: Wiley.
- MacLachlan, M. (2004b). Culture, empowerment, and health. In M. Murray (Ed.), *Critical health psychology*. London: Sage.
- MacLachlan, M., Smyth, C. A., Breen, F., & Madden, T. (in press). Acculturation and mental health in modern Ireland. *International Journal of Social Psychiatry*.
- McDonald, H. P., Garg, A. X., & Haynes, R. B. (2002). Interventions to enhance patient adherence to medication prescriptions: Scientific review. *Journal of the American Medical Association*, 288, 2868–2879.
- Merson, M. H., Dayton, J. M., & O'Reilly, K. (2000). Effectiveness of HIV prevention interventions in developing countries. *AIDS*, 14(Suppl. 2), S68–S84.
- Mulatu, M. S., & Berry, J. W. (2001). Cultivating health through multiculturalism. In M. MacLachlan (Ed.), *Cultivating health: Cultural perspectives on promoting health* (pp. 15–35). Chichester, UK: Wiley.
- National Research Council. (2001). *New horizons in health: An integrative approach*. Washington, DC: National Academy Press.
- Nazroo, J. Y. (2003). The structuring of ethnic inequalities in health: Economic position, racial discrimination, and racism. *American Journal of Public Health*, 93, 277–284.

- Parker, R. G., Easton, D., & Klein, C. H. (2000). Structural barriers and facilitators in HIV prevention: A review of international research. *AIDS, 14*(Suppl. 1), S22–S32.
- Rivers, K., & Aggleton, P. (2000). HIV prevention in industrialized countries. In J. L. Peterson, & R. J. DiClemente (Eds.), *Handbook of HIV prevention* (pp. 245–263). New York: Kluwer Academic/Plenum.
- Schneiderman, N., Antoni, M. H., Saab, P. G., & Ironson, G. (2001). Health psychology: Psychosocial and biobehavioral aspects of chronic disease management. *Annual Review of Psychology, 52*, 555–580.
- Swanson, B., Keithley, J. K., Zeller, J. M., & Cronin-Stubbs, D. (2000). Complementary and alternative therapies to manage HIV-related symptoms. *Journal of the Association of Nurses in AIDS Care, 11*(5), 40–60.
- Taylor, B. (2001). HIV, stigma, and health: Integration of theoretical concepts and the lived experiences of individuals. *Journal of Advanced Nursing, 35*, 792–798.
- Tuldra, A., & Wu, A. W. (2002). Interventions to improve adherence to antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes: JAIDS, 31*(Suppl. 3), S154–S157.
- UN Program on HIV/AIDS. (1999). *Sexual behavioural change for HIV: Where have theories taken us?* Geneva, Switzerland: Author.
- UN Program on HIV/AIDS. (2002). *Report on the global HIV/AIDS epidemic*. Geneva, Switzerland: Author.
- Wilkinson, R. G., & Marmot, M. G. (Eds.). (1998). *Social determinants of health: The solid facts*. Geneva, Switzerland: World Health Organization.
- Wilson, B. D., & Miller, R. L. (2003). Examining strategies for culturally grounded HIV prevention: A review. *AIDS Education & Prevention, 15*(2), 184–202.
- World Health Organization. (2002). *The World Health Report 2002: Reducing risks, promoting healthy life*. Geneva, Switzerland: Author.
- Yudkin, J. S., Bayley, O., Elnour, S., Willott, C., & Miranda, J. J. (2003). Introducing medical students to global health issues: A bachelor of science degree in international health. *Lancet, 362*, 822–824.



# History of Applied Psychology, Overview

Helio Carpintero

Complutense University, Madrid, Spain

1. Introduction
  2. Early Landmarks in Applied Research
  3. The Institutionalization of the Field: Emerging Societies of Applied Psychology
  4. Significant Problems in Early Applied Research
  5. The Influence of World War II (1939–1945)
  6. Theory and Practice: The Influence of the Era of Schools
  7. Applied Psychology in the Second Half of the 20th Century
  8. General Trends
- Further Reading

## GLOSSARY

**evolutionism** A doctrine that maintains the formation of complex real entities from simpler ones, through an evolving natural process that brings in differentiation and heterogeneity in more homogeneous situations.

**institutionalization** The process of creating certain social structures that will be in charge of some definite roles originally solved by people's personal and original behavior.

**intervention** A concrete action that tries to introduce some changes in a given situation, usually planned and devised according to some previous theory, and adapted to the here-and-now peculiarities.

**positivism** A philosophical doctrine, first formulated by French A. Comte (19th century), that maintains the primacy of scientific knowledge and the need for building society upon a scientific worldview.

**school** A group of persons working on a certain field, according to some idiosyncratic views, and usually organized in a master–disciples structure.

**theory** A conceptual construction that tries to represent a certain group of phenomena in order to explain and predict them.

## 1. INTRODUCTION

Theoretical and applied psychology have frequently been considered as two completely different ways of approaching and conceiving of mental and/or behavioral phenomena. Their lines of development have met one another from time to time. But most frequently they have remained unrelated, as different questions, instruments, and methodologies have dominated in each of them, creating two different worlds.

Interest in psychological questions goes back to ancient times, as proven by philosophical reflections and literary creations. The diversity of characters and the description of typical reactions to a variety of situations form the core of countless literary masterpieces all over the world. Nonetheless, neither philosophical reflections nor literary fantasies enter our tableau. Psychology conceived as an empirical science is little more than a century-old enterprise. For didactical purposes, Wilhelm Wundt's laboratory at Leipzig

University (1879) is credited with the honor of representing the first event in its history.

This new science tried to reformulate old questions about the nature of the human mind and its creations—knowledge, volition, choice, consciousness—and its connection with the biological organism that owned such a mind. Two main ideas dominated the second half of the 19th century: evolution and science (the third one would be democracy). While evolutionary theory brought forth the idea of a continuity between humans and the animal kingdom, positivism stressed the value of scientific knowledge in all empirical domains and stimulated the creation of a natural science of the human mind. In what relates to the third idea, democratic values would soon demand the application of that new knowledge to solve the problems of societies developing under the impulse of the industrial revolution.

Two main lines, biological and sociological, characterize the early developments of psychological research. While certain groups focused on physical and physiological determinants of human behavior, others insisted on socio-historical factors mediating mental processes. On the one hand, work in laboratories and theory construction were stressed; on the other hand, researchers relied on social and ethnological comparisons and became largely interested in practical knowledge.

In the early days of the 20th century, applied research was not an easy task. In the United States, Edward B. Titchener (1867–1927), a “pure scientist,” strongly opposed psychologists emerging from laboratories and facing daily life problems. He feared that their theories were not yet prepared for such a huge task. But demands could not wait unattended for long. Moreover, funds for research soon became dependent on practical results accredited by scientists. Little by little, applications began to follow theory, and science gave way to practical interventions; scientists were accompanied by practitioners, and psychology became a science as well as a profession, as it is at present.

Since its beginnings, interventions in psychology were characterized by the need for combining both a scientific background and information on the idiosyncratic characteristics of every practical problem. Applied science, in the case of psychology, has not only consisted of translating general principles to concrete circumstances, but also of creating new theoretical models that are fully in accordance with concrete data, trying to meet the particular demands, and gaining through such intervention new theoretical insights. Such circular interactions between research and theory were clearly stated by Kurt Lewin in his “action

research” doctrine and have been frequently found by researchers in most of the practical fields.

Developments in the field have been deeply influenced by national trends. Progress cannot be reduced here to a single common process covering all countries and specialties. Three main areas did emerge, however, as their problems were found in all well-developed societies: (1) children’s adaptation to school, (2) worker’s adaptation to job requirements, and (3) people’s adaptation to the requirements of “normal” life in a social world. These areas led to the formation of school and educational psychology, industrial and organizational psychology, and clinical psychology, respectively. As time passed, the mapping of specialties became more and more complicated.

Certainly, among all these differences there are some common traits in these applications that may serve to unit and define the field of applied psychology. These traits will be considered after a general discussion of the field.

## **2. EARLY LANDMARKS IN APPLIED RESEARCH**

While the establishment of Wundt’s laboratory is a landmark that represents the first step, there is no formal beginning for applied psychology as such. In each specialty, however, the first steps have had some unique characteristics of their own that will be considered here.

### **2.1. Clinical Psychology**

It is not easy to disentangle clinical psychology from the fields of mental hygiene and psychiatry, but it might be accepted that problems and deficits in mental aptitudes can be seen as psychological problems, whereas deep mental disorders require medical treatment. The former will impair normal adaptation to life, while the latter will prevent normal functioning of the person among his relatives and in his private world.

The first steps in the field of clinical psychology stem from the introduction of the idea that a mental process and its abnormal functioning could be treated through behavioral/mental means, instead of using somatic or chemical procedures. An empirical approach to the study of mental disorders was demanded in this new science of the mind, as these disorders began to be seen as natural disturbances capable of being treated. Some epochal events took place in the year 1896. In Germany, a student of Wundt, psychiatrist

Emil Kraepelin (1856–1926), first proposed the introduction of the psychological laboratory and the experimental study in psychiatry, in a seminal paper that has been called the first German program for clinical psychology. Kraepelin divided illnesses into two categories: those due to endogenous factors and those caused by exogenous ones; the former were viewed as usually incurable, while the latter had a better prognosis. Based on empirical grounds, he also worked on a classification of mental disorders that served for decades as a nosological system.

Also in 1896, an Italian psychiatrist, Giulio Cesare Ferrari (1868–1932), with G. Guicciardi, published a paper on mental tests that in their opinion could be employed in psychiatry to improve diagnosis. They tried to continue the pioneer work of Gabriele Buccola (1854–1885), who, working with patients suffering from mental disorders, found interesting variations in their reaction time measures. Ferrari and Guicciardi examined attention, associative abilities, and other mental capacities in patients in order to more accurately characterize their pathologies.

And it was in the same year when, in the United States, Lightner Witmer (1867–1956), another student of Wundt and Cattell, founded the first psychological clinic, at the University of Pennsylvania. Witmer became an apostle of the psychological intervention. At the American Psychological Association (APA) convention of that same year (1896), he maintained that psychologists had to assume two tasks: to study mental processes and to apply their knowledge to solve personal problems and social needs. He had to face the negative attitude of his colleagues, who did not wish to accept practical responsibilities. He decided to organize regular training courses at his university for people interested in applications of psychology, with case presentations and formation in diagnostic techniques, and also founded a journal, *The Psychological Clinic* (1907), that claimed to promote “orthogenics,” the knowledge of the correct development of individuals, both in physical and mental capabilities.

A different approach was carried out by William Healy (1869–1963), a physician who treated juvenile delinquency as a phenomenon he believed was rooted in certain psychopathological factors. He created a battery of proofs and applied them to young offenders, trying to determine their personalities and social factors in order to implement preventive and therapeutic interventions. His first clinic was established in Chicago (1909) and then he moved to Boston, where he retired.

In sum, demands for help from psychologists, coming from the social world, were based on the assumption that the new science of psychology, which was dealing with the normal functioning of the mind, could also say something about those pathologies now interpreted as mental disorders.

## 2.2. Educational Psychology

The 19th century saw the democratic rise of masses in Western countries, and basic education and training spread out to all individuals and classes. Many people became teachers, but they soon discovered that their job was not an easy one. Individuals are endowed with unequal learning capacities, and those prevented from normal schooling caused great disturbances in the classroom. Because their need for special treatment was ignored by most teachers, they turned to psychologists for orientation and help. Whereas old-fashioned schooling was based on memory and rote learning, some new tendencies stressed the role of an active process of learning and understanding, a true psychological topic. Teachers viewed psychology as the scientific basis for their own art of teaching.

The association paradigm became largely influential. In 19th century Germany, Johann F. Herbart's (1776–1841) association became the basis of the new science of education; in the United Kingdom, interest in these matters went back to empiricist philosophers, from John Locke to James Mill, whose ideas inspired 19th century psychologists such as Alexander Bain (1818–1903) and James Sully (1842–1923). Sully made careful empirical studies of language, perception, memory, and other faculties in children, and wrote one of the first textbooks, *Teacher's Handbook of Psychology* (1886).

French pedagogue Gabriel Compayré (1843–1913) wrote another textbook for teachers that included not only theoretical concepts on association and other mental aspects, but also practical topics, such as children's plays, how to improve curiosity and memory, and the technology of teaching. His book was translated into Spanish by the great Argentinian educator and president, D. F. Sarmiento, and became influential among Latin American teachers and pedagogues.

Largely under the influence of Wundt's psychology, an experimental pedagogy began to spread in Germany. Scientific knowledge of child development, based on observation and experiments, was systematized in the pioneering work of Wilhelm Preyer (1841–1897),

*The Soul of a Child*. Pedagogues such as Wilhelm A. Lay (1862–1926) and Ernst Meumann (1862–1915) multiplied their efforts to provide psychological support to educational practices. Some of them stressed the need for building a global knowledge of child life, and a first draft was accomplished by German Oscar Christmann in 1889, with his proposal of a “paidology”—or “pedology”—conceived of as a general science of the mental and physical development of the child. Centers for research and teaching were then founded, such as the Institute for Pedagogy and Experimental Psychology of Leipzig, and the Institute for Applied Psychology, created by William Stern and Otto Lipmann in Breslau, both founded in 1906. Specialized journals such as *Zeitschrift fuer Kinderforschung* (1895) also began to publish the results of their research.

Some educators spoke up against undue expectations of psychology. William James (1842–1910), in *Talks to Teachers* (1899) made clear the distinction between the science of psychology and the art of teaching, affirming that “sciences never generate arts directly out of themselves.” But some of his students, like Granville Stanley Hall (1846–1924), assembled a large body of knowledge about lifespan psychology (e.g., childhood, adolescence, senescence), and wrote a seminal work on “The Content of Children’s Mind on Entering School” (1883) in which he empirically explored the average knowledge of children in his society. This field soon became an active area of research. Hall expanded that paper into a book (1894), and founded the journal *Pedagogical Seminary* (1891), later renamed the *Journal of Genetic Psychology*.

Interactions between psychology and education multiplied rapidly, and benefited from the growing knowledge about children promoted by evolutionism. In Belgium, paidological topics received great attention, especially topics related to mental retardation. A. Sluys, J. Demoor, and Ovide Decroly (1871–1932), among others, made lasting contributions that became influential all over Europe. The latter encouraged the adoption of the New School model, which promoted the ideals of active learning and the view of the school as a laboratory.

Great attention was also paid to these topics in socialist countries. In Russia (in those days the Soviet Union), many centers oriented toward experimental study of child development and school education were founded, as the building of a “new man” was a political goal according to Marxist views. Lev Vygotsky (1896–1934), Pavel Blonsky (1884–1941), Aron Zalkind (1888–1936) and others worked on this, stressing the

social and practical roots of human consciousness. A Congress of Pedology gathered in 1928, a journal (*Pedology*, 1928–1932) was founded, and activity grew rapidly, until July 4, 1936, when a decree from the Communist Party headed by Josef Stalin halted testing and deemed nearly all psychological practices counter-revolutionary activities.

The ideals on active education were at the core of an institution created in Geneva, Switzerland in 1912 by Edouard Claparède (1873–1940), Pierre Bovet, and some collaborators, the Institute Jean Jacques Rousseau, which soon became a model for all who were interested in those topics. Claparède offered an overview of the field in his *Psychologie de l’Enfant et Pédagogie Expérimentale* (Child Psychology and Experimental Pedagogy, 1905), in which he stressed the idiosyncratic nature of child life and the need of an educative system that is well adapted to the individual nature of every child and capable of giving effective guidance to students.

Some early studies had great impact on the field, such as research on formal discipline and its effects on concrete learning. In the United States, Edward L. Thorndike (1874–1949) and Robert S. Woodworth (1869–1962) performed experiments on the transfer of training that showed poor or negative aftereffects of a first learning task upon another following it. In doing so, they destroyed the belief that some benefits might be obtained from disciplines such as Latin or technical drawing upon other matters included in the curriculum. But the study of individual differences among children and a more detailed knowledge of the evolution of intellectual abilities and interests enabled teachers to adjust their teaching to the children’s capabilities.

Apart from these descriptive studies, more effective instruments for assessing children in schools were demanded. In response, some psychologists built instruments for screening deficits and assessing effective teaching. They initiated the application of proofs and measures to education. This will be discussed in a following section on testing.

### 2.3. Early Work in Organizational Settings

Western societies have created all sorts of organizations in which the principles of impersonal work, labor division, cooperation, and leadership applied. In modern states, people operate under the leadership of certain elites, and technology introduced during the industrial revolution is changing the world through the work of

people working together. In the early decades of the 19th century, the need for a science of social movements and phenomena stimulated the creation of sociology, an accomplishment of French theorist Auguste Comte (1798–1857), the founder of positivism. Soon after, theorists tried to explain some dimensions of social life as mental phenomena, as well explain some psychological facts by social forces. Interactions between these fields multiplied, and psychological knowledge of individuals operating in a group or organization seemed more and more convenient. Aspects such as job–person fit, efficacy, performance quality, accident prone operators, and social competition were not without economic consequences. Enterprises in search of good operators soon incorporated psychological testing devices in order to find reliable and valuable workers. Psychology appeared as a science dealing with problems closely related to person–job adaptation, and soon society asked psychologists for help in the task of improving industries and social organizations.

The concept of applied psychology was first delineated by German William Stern (1871–1938). According to him, it was “the science of psychological facts that are relevant by their practical utility,” including both assessment and intervention (Dorsch, 1963, p. 9). Early essays on this topic are characterized by their spontaneity and diversity; frequently they were written by laymen interested in practical questions. For instance, Italian physiologist Angelo Mosso (1846–1910) stressed the importance of analyzing the fatigue curves of workers to improve their performances; his findings and methods spread rapidly in Europe. On the other hand, a U.S. engineer, Frederic W. Taylor (1856–1915), working for important steel companies, studied their production from the point of view of the productive behaviors and their timing and morphologies, giving place to a movement called the “scientific organization of labor,” that applied the scientific methodology to the study of people’s efficiency at work, and the ways to improve it. Although not a psychologist himself, he applied psychological techniques—reaction time, movement studies, cooperation among workers, economic incentives—in trying to combine personal satisfaction with greatest worker efficiency at the risk of being judged an inhumane manager. His proposals introduced a new concept to industries, and while social movements and trade unionism shook developed societies, the problems of people at work began to be viewed under the light of psychology. Taylorism paved the way for a more specialized organizational psychology.

The new climate brought new fruits. Frank Parsons created the Boston Vocation Bureau (1908), convinced that a test for discovering motives and aptitudes in individuals could improve their professional adaptation. As a philanthropist, he wanted to help people in search of a job, and his center considered many aspects: people’s abilities, technical and moral training needed, self-understanding, and the material conditions of the job.

A world becoming dominated by technical apparatuses and procedures demanded efficient people for using them. Car driving and traffic security also stimulated the screening of potential accident prone drivers via psychological testing, a task in which German researchers Walter Moede (1888–1958) and Curt Piorkowski (1888–1939) did significant work employing driving simulators. Motion studies of bricklayers done by Lillian M. Gilbreth (1878–1972) and Frank Gilbreth (1869–1924), and of typist’s abilities done by Jean M. Lahy (1872–1943), among others, stimulated the research, but it was Hugo Münsterberg (1863–1916), the “father of psychotechnics,” who was the first to face the problem in its entirety and gave a certain solution.

Trained under Wundt, then working with James at Harvard University, Münsterberg conceived of psychotechnology as the “science of practical applications of psychology in the service of culture” (1914). His books on industrial efficiency, teaching, and eyewitness testimony offered a broad view of psychology as a practical science. He devised proofs and testing situations that proved useful for personnel selection, and received much support from businessmen and industry managers. His works describe the new field of psychological intervention as a coherent whole, in which demands were idiosyncratic, but techniques were interspecific and largely versatile in themselves. Person, situation, and means-ends relations were the significant elements to take care of in every practical intervention.

## 2.4. Early Testing

Early psychophysics—the scientific study of stimuli and the correlated sensory experience—found systematic variations in sensory and motor abilities among individuals. One of the pillars of Darwin’s evolutionism was the existence of spontaneous variations among individuals who were unequally endowed, in order to cope with the challenges of their environment and subject to natural selection.

Francis Galton (1822–1911), who was deeply influenced by evolutionism, studied such variations and devised a battery of tests to measure differences in various types of capacities among subjects. He applied such proofs in an anthropometric laboratory at the International Health Exhibition of South Kensington in 1884, and created some statistics to sum up his data.

James McKeen Cattell (1860–1944), a U.S. student of both Wundt and Galton, had a long-lasting invention: the construction of standard proofs to compare the performance of different people. In 1890, he coined the expression “mental test” for measuring mental capacities, which spread out all over the world and became a symbol for psychology as a profession. His presidential address at the APA convention in 1895 enforced the need to present to society not only scientific but also material successes in the practical field, in order to enhance the status of psychology; the simplest experimental situation in which a person could be studied and measured was the test. For that reason, tests grew without limitation in the 1920s and 1930s.

In Germany, Hermann Ebbinghaus (1850–1909) devised a completion test, in which some missing letters had to be found, for assessing the effects of fatigue in school children, a type of proof frequently used since then. But it was Alfred Binet (1857–1911) who, with his collaborator Victor Henri (1872–1940), made an epochal contribution: the metric scale for intelligence (1905), a worldwide test for assessing intellectual abilities, allowing the determination of the relative position of a child (his “mental age”) among a certain population. Moreover, intelligence soon appeared to be highly correlated with most mental abilities, as a kind of broad cognitive strength or general factor influencing all types of performances in which those abilities operate, as found by Charles Spearman (1863–1945).

Evaluation and comparisons between individuals became then possible. Age was the basis for comparisons in Binet’s system; a further step to get a stable index across ages came from German psychologist William Stern (1871–1938), who introduced the concept of the intelligence quotient (IQ), the ratio of mental age to chronological age, a score that seemed to remain unchanged along the years. Its measure became a highly valuable index of a person’s capabilities and a good predictor of future performances.

Benefits from the possibility of measuring intelligence began to appear everywhere. School was only one field interested in such measures; social life, job selection, even clinical practice soon could profit in

different ways from this measurement. The various branches of psychological practice began, little by little, to cross-fertilize each other, and their methods and purposes approached each other, producing a certain unification of the applied field as a whole.

## **2.5. Interactions across Specialties**

Three main facts caused the explosion of applications of psychology to social problems: the usefulness of intelligence testing for successful prediction of future performance; the growing need of a scientific explanation for the behavior of people in organizations shaken by social and political unrest; and the rise of a clinical view of abnormality in the light of psychological conflicts of personality, as in the contributions of Sigmund Freud (1856–1938). All these historical innovations converged in viewing mind and behavior as the Gordian knot of the problems of modern societies. Psychologists were asked to respond to those problems, and an external event, World War I (1914–1918), provided the opportunity to prove their social efficiency. As Cattell wrote, it “put psychology on the table,” and nobody could ignore it.

War and military needs can be considered as some of the strongest stimuli for developing psychology in societies. World War I gave a great impetus to all the applied branches; the areas of assessment and personnel selection, and also clinical intervention upon war-caused psychopathologies, quickly developed. Three main problems emerged during wartime: war neuroses (then called shell-shock), the health of munitions workers (menaced by industrial fatigue), and military personnel selection.

In the United States, psychologists from the APA, headed by R. Yerkes (1917) and considered as a part of the medical division, worked out a test battery (the Army Alpha for literates, and Army Beta for illiterates) for assessing soldiers. Both proved to be effective. Examinations were given to more than one and a half million people, and since then, tests have been an indispensable part of any personnel selection. The war also was the motivation for the creation of some clinical instruments, like the personal data questionnaire (1918) devised by R.S. Woodworth (1869–1962) to detect symptoms of a neurotic tendency. With the end of the war, these particular activities greatly diminished.

The war also promoted the introduction of psychology as part of the Army services in Germany. There some regular positions for psychologists were created

in 1925, mainly oriented toward officer selection. This number increased after 1935, as military psychology became synonymous with psychology under the Nazi government. In Great Britain, the war brought the problem of industrial fatigue to the forefront. In munitions factories, people had doubled their working time, and physical and mental disorders rapidly appeared, affecting the subjects' health and efficiency. To face the situation, a Health of Munitions Workers Committee was set up by the government in 1915; it was renamed the Industrial Fatigue Research Board (1917) and then became the National Institute of Industrial Psychology (1921), headed by Charles S. Myers (1873–1946). Similar efforts were carried out in other countries. Airplane pilots were urgently needed, and they were examined by psychologists. In Italy, Father Agostino Gemelli (1878–1959), founder of the first psychophysiological laboratory of the Italian Army, did his best to develop applied psychology in his country in the 1920s and 1930s; in France, Camus and Nepper (1915) did similar work.

Around the same time, the first centers for applied psychology began to appear. In the United States, a psychology department was established in 1915 at the Carnegie Institute of Technology, headed by Walter Van Dyke Bingham (1880–1952), and a year later, Walter Dill Scott (1869–1955) was appointed as professor of applied psychology, the first to receive that title in that country. Vocational counseling, industrial efficiency, and related topics began to be treated as academic matters, so the time was ripe for institutions to specialize in those questions. For example, the Kaiser Wilhelm Institut, the Laboratory for Industrial Psychotechnology from Charlottenburg, the Institute for Work in Moscow, the Institute of Professional Orientation of Barcelona, the Institute Jean-Jacques Rousseau at Geneva, the Psychological Laboratory at the Catholic University of Sacred Heart (Milan), the Institute of Scientific Organization at Warsaw, the Intercommunal Center for Professional Guidance at Bruxelles, the Laboratory for Ergonomics of the Institute of Higher Studies of Belgium, the Laboratory of Experimental Psychology of the School for Higher Studies at Paris, the Center for Professional Orientation, and the Bos Institute (the Netherlands), among others, became active centers for training, testing, and research. These institutions and centers were created to face social demands, and utility was the dominant criterion accepted by researchers, although in the long run many theoretical contributions were produced from such practical work.

### 3. THE INSTITUTIONALIZATION OF THE FIELD: EMERGING SOCIETIES OF APPLIED PSYCHOLOGY

The period that followed World War I brought with it the normal problems in a rapidly growing field without any internal regulation, mainly under external pressures. Psychology was acknowledged as a useful instrument for social work, and its techniques for personnel selection spread both to private firms and to governmental agencies.

To face the new demands, psychologists tried to set up several professional networks. The need for coordination and the convenience of employing common instruments for assessment stimulated the foundation of an international society for applied psychology, was the *Association Internationale des Conférences de Psychotechnique* (International Association for Psychotechnology). Rooted in Europe but open to the whole world, it held its first meeting, chaired by Swiss Edouard Claparède (1873–1941) in Geneva in 1920; in 1955 it changed its name to the International Association of Applied Psychology (IAAP). The society provided a useful platform for cooperation, but was greatly affected by the political circumstances that led to World War II; 15 years passed between the 8th Congress, held in Prague in 1934 and the 9th, held in Bern in 1949. Since then, it has gathered regularly. The IAAP has been headed by leading scientists: Edouard Claparède (1920–1941), Henri Piéron (1947–1953), Clifford Frisby (1953–1958), Morris S. Viteles (1958–1968), Gunnar Westerlund (1968–1974), Edwin A. Fleishman (1974–1982), Claude Lévy-Leboyer (1982–1990), Harry C. Triandis (1990–1994), Bernhard Wilpert (1994–1998), Charles D. Spielberger (1998–2002), and Michael Frese (2002–present).

Many other international associations have been created to serve the multiple interests of psychologists. The International Union of Psychological Science (IUPsyS) was created in 1951, grouping together many previous national associations. While its roots go far back (a first International Congress of Psychology took place in Paris in 1889), the IUPsyS now represents psychology in international councils of science and culture. Also, the Interamerican Psychological Society (*Sociedad Interamericana de Psicología*, SIP), was founded in 1951 to cover the needs of psychologists from the entire American continent; it is very active at present. Many other specialized societies have flourished in parallel to the growth of the profession and the scientific field.

An outstanding society as measured by its size and activities is the American Psychological Association (APA). Founded by G. S. Hall in 1892, in its beginnings it covered only scientific matters. Therefore, applied oriented psychologists created the American Association of Clinical Psychologists in 1917, but finally this group was included in the APA in 1919 as a clinical section dedicated to professional issues. A new group called the American Association for Applied Psychology was founded in 1937 to better serve practical interests. It edited a journal (*Journal of Consulting Psychology*), but eventually it too joined the APA in 1945, forming a strong society, organized in various divisions, oriented to both science and profession.

The growing complexity of the applied field necessitated a communication network that could disseminate specialized information. During the years of war, some periodicals began to appear. In Germany in 1916, the Institute for Applied Psychology of Leipzig edited the *Zeitschrift fuer Angewandte Psychologie* [Journal of Applied Psychology] under the care of William Stern and Otto Lipmann. In 1917 in the United States, G. Stanley Hall initiated the publication of the *Journal of Applied Psychology*, another one of the leading journals in the field. It was soon managed by James P. Porter (1920–1942), and later owned by the American Psychological Association. Other significant journals have been the *Revue de Psychologie Appliquée* [Review of Applied Psychology] (1920), and *Human Factor* (1932), under the sponsorship of the National Institute of Industrial Psychology and the initial editorship of C. S. Myers. Since then, the network has developed enormously, and the creation of the *Annual Review of Psychology* (1950) and more recently the data banks that have been built in many specialties have introduced a new order into the information chaos.

#### 4. SIGNIFICANT PROBLEMS IN EARLY APPLIED RESEARCH

Some basic problems arose during the evolution of applied psychology. The following sections briefly describe some examples of these problems.

##### 4.1. Inborn versus Learned (Acquired) Abilities

Theorists had to decide whether human abilities were innate or learned; testing procedures and measurements

raised suspicions and opposition among socially involved groups on political grounds. Large differences in measurements between various classes and ethnicities raised political doubts about the fairness of procedures that, in fact, introduced discrimination between races, classes, and individuals. In addition, an increasing number of dimensions of life were subject to study under this methodology.

One subject of intense debate was intelligence. Did the result of testing stem from the inborn characteristic of the person or from the social conditions of the person's life?

Early on, Francis Galton (1822–1911) was inclined toward a genetic solution to the dilemma. In the United States, H. H. Goddard (1866–1957), a Binet follower, reinforced that opinion. He compared two lines of descendants, both legitimate and illegitimate, flowing from a common ancestor and examined their achievements and life records. He concluded that hereditary factors explained most of the variance, as the legitimate family was largely superior to the other one in intellectual and moral qualities.

Different races and cultures were also compared on the basis of their performance on intelligence tests. Data from the Army Alpha and Beta tests were studied and reanalyzed in the United States. Significant differences in IQ among people from various ethnic origins were supposedly found. Among other questioned results, the intellectual weakness of African Americans seemed to be firmly established on these grounds. According to Garrett (1951), "Negro soldiers scored lower than whites on Alpha, southern Negroes lower than northern Negroes." Following these results, democracy was criticized by some groups on the grounds of offering political equality to unequally mentally endowed people. Racist arguments seemed to flow from psychological data. As a consequence, U.S. federal law regulating immigration excluded all "persons of constitutional psychopathic inferiority."

Other voices joined the choir. In Europe, data obtained by H. English and others suggested a strong correlation between economic level and IQ and pointed to non-inborn factors in intelligence scores. The discussion was deeply affected by political prejudices and attitudes. IQ testing was then criticized on political grounds. In the United States, an influential journalist, W. Lippmann (1899–1974), strongly criticized the testing procedures for taking scores as true measures instead of rough estimates of variable qualities in subjects.

In Russia, then under the Soviet regime, a ban on all testing activities in schools was imposed by the

Communist Party's Central Committee. They deemed such activities a "pedological perversion," considering tests as reactionary and anti-egalitarian techniques. Given the thesis of the social determination of consciousness caused by socioeconomic factors, the idea of inborn abilities was seen as unacceptable. Such a decision greatly impaired the development of empirical psychology in the country for decades.

As is well known, this debate has endured for an entire century and at certain moments has become intertwined with political attitudes. The works of A. R. Jensen (1967), H. J. Eysenck (1971), L. Kamin (1974), S. J. Gould (1981), R. Herrnstein, and C. Murray (1994) represent only a small part of the vast literature generated by this controversy.

Today, heredity and environment are placed in a completely different relationship, as genetic engineering has resulted in a complete mapping of genes. Proteins related to organic and mental functioning are now beginning to be discovered, and intervention at such microscopic levels will be possible in the future, opening the way to biological techniques of medical and psychological value.

#### 4.2. Types of Measurement and Types of Diagnosis

A long-ranging debate commenced between those inclined to build diagnoses upon a collection of measures, including personality traits and abilities, and those who limited themselves to the study of the subjects' aptitudes directly related to the job they were applying for. In the early days, theorists such as German O. Lippmann (1880–1933) tended toward a holistic approach, while many others (e.g., Belgian A. Christiaens) favored the opposite view.

The debate evolved into the opposition between qualitative and quantitative (psychometric) procedures for assessing people. The former was frequently called clinical methodology, because it focused on concrete phenomena affecting a single subject (idiographic methods), while the latter was frequently referred to as statistical methodology, as far as it employed quantitative measures obtained from groups of subjects through psychometric tests (nomothetic methods). For instance, some theorists, such as Swiss cognitivist Jean Piaget (1896–1980), firmly rejected the use of psychometric proofs, convinced they gave wrong information about the child's mind, and systematically employed clinical procedures.

Researchers also developed instruments to generate and appreciate qualitative experiences that might be directly related to mental forces and impulses—the projective techniques, which included well-known proofs such as Rorschach's inkblot test (1921) and Murray's Thematic Apperception Test (1938). Hermann Rorschach (1884–1922), a Swiss psychiatrist, wrote *Psychodiagnostik* based on the experimental study of characteristic answers to some inkblots provided by different criteria groups of psychiatric patients. On the other hand, many psychologists tried to devise laboratory situations closely related to those found in real situations.

Opposition to mental measurement also grew from theoretical grounds. Behaviorism, which largely dominated U.S. psychology between the 1920s and 1960s, banned all mental-based concepts from its system, especially from all its clinical topics. It stated that psychology should deal only with behavioral facts and laws, and each individual should be tested in definite settings, in order to establish those precise S-R associations determining his or her behavior. Adaptive or maladaptive habits were substituted for old personality traits, and behavioral changes were evaluated in order to assess the effectiveness of various procedures of behavior modification. New instruments based on observational procedures were developed, employing sampling recording of target behaviors in single subject ( $n = 1$ ) cases; these techniques substituted for traditional tests and questionnaires.

#### 4.3. Studying People or Analyzing Situations

Early psychotechnology tried to study the adaptation of a person to a situation by comparing the person's qualities to the requirements of each type of situation. Typologies and professions were its main fruits. But some unfitting data opened the door to the study of work as a global social situation. Surprising data emerged from some studies on work efficiency carried out by George Elton Mayo (1880–1949) and collaborators on employees of the Western Electric Company in Hawthorne, Chicago. An in-depth study with those workers showed that their efficiency at work was mainly dependent not on physical factors at the workplace but on social characteristics of the group—informal rules, leadership, social climate, expectations, and fears—and also from the fact that they were receiving attention from researchers, a fact that influenced the feelings of those

workers. Mayo, F. G. Roethlisberger, W. J. Dixon, and G. Homans placed emphasis on human relations in industry and stressed the relevance of the “human meaning” of work and the necessity of a holistic approach to work problems. Significant developments in this theory were obtained when extra-organizational variables, such as technological and market changes, religion, race, and social class, were also taken into account by the Chicago school (W. L. Warner and colleagues) and other groups.

All these experiences showed the need to study the characteristics and behaviors of groups. Gestalt psychologist Kurt Lewin (1890–1947) emerged as a leader in that field. His ideas influenced not only the general field of social psychology, but also the study of applied aspects, such as the relationship between school climate and teaching efficiency. Lewin, R. Lippitt, and R. K. White evaluated the effect of different group atmospheres—democratic, authoritarian, and *laissez faire*—upon school teaching and group interactions. Leadership proved to be an important variable, and various models of it have been devised to explain empirical data. Results from industrial settings were transferred to educational and clinical fields. Group psychotherapies have been applied since the 1920s by Alfred Adler (1870–1937) and coworkers, and group interactions proved fruitful in handling feeling and personality disorders. Well into the 1950s, group psychology developed a variety of forms and applications: for instance, T (training) groups were created that offered training in skills and interpersonal communication; encounter groups were employed for therapeutic discussions in humanistic psychotherapies. A special place in this line of thought belongs to the Tavistock Institute (UK) founded in 1947. It focused on the interactions between technological and psychosocial aspects of organizations, and tried to combine psychoanalytic concepts and the group dynamics as developed by the Lewinian group of theorists. A. K. Rice, E. L. Trist, W. R. Bion, and many others helped the development of this movement, and its journal, *Human Relations* (1947) maintained a comprehensive perspective in the field.

#### 4.4. Intervention versus Goal-Oriented Research

As applied psychology rapidly developed, more and more concrete questions absorbed the interest of researchers, who began to specialize in a small range

of topics and to sight of the entire field. An increasing distance began to appear between theory and applications, and as a result, some tried to restore the communication between both fields.

After World War II, clinical psychologists in the United States felt the necessity of combining practice with research. In a conference at Boulder, Colorado in 1949 they devised the so-called Boulder model, which recommended “the training of the clinical psychologist for research and practice, with equal emphasis on both” (Raimy, 1950). This model has been widely accepted, and is still in force among specialists.

Similarly, social psychologists doing research and intervention in social settings were warned by K. Lewin in 1948 that their interventions had to be organized “in a spiral of steps each of which . . . composed of a circle of planning, action, and fact-finding about the result of the action.” In combining planning, action, and evaluation, researchers will adjust themselves to a Lewinian “action research” model that puts together theory and action into a single operation, a model assumed by most social scientists. These innovations perceived practice and theory as two dimensions of the same impulse toward knowledge.

### 5. THE INFLUENCE OF WORLD WAR II (1939–1945)

World War II deeply changed the social scene, and all aspects of human life, including psychology, felt its impact. Numerous institutions, researchers, and schools of thought were destroyed or silenced by the war. In Europe, an enormous number of scientists and intellectuals belonging to Jewish families were persecuted and even killed in countries placed under the Nazi regime. Psychologists were no exception. Many of them fled to the United States or to other countries, and European psychology, until then considered the best in the field, lost its position of leadership to the United States.

In Germany, some psychologists carried out research for the new regime. The German army established a *Wehrmacht* psychological division, which performed the selection of specialists and officers; the Reich Institute for Psychological Research and Psychotherapy trained “doctor psychologists” for psychotherapeutic labor. Psychologists found new professional perspectives not limited to the academic field; at

the same time, people teaching in universities had to show their loyalty to the racist ideology then dominating the country.

Likewise, many psychologists in the United States worked for the U.S. army, doing personnel selection, training, and psychotherapy. Limited but important research was also carried out on topics like morals, propaganda, and psychological warfare. Well-known researchers such as R. M. Yerkes (1876–1956) and E. G. Boring (1886–1968), among others, made great contributions to that effort; more than a thousand professionals worked full time to face the emergency. During that time, a new test (the Army General Classification Test) was administered to some 12 million soldiers. Its results favored an in-depth study on aptitudes and abilities.

The war also brought to the fore new problems and research topics for psychologists. Studies on visual perception in flight situations fostered new research in ecological theory like that carried out by James J. Gibson (1904–1979); new techniques for radar detection systems paved the way for a new psychophysics, based on signal detection and decision theory; vigilance and task monotony induced new research on stimulus deprivation and attention processing; race prejudice and antisemitism stimulated research on authoritarian personality by Theodor Adorno (1903–1969) and colleagues; and information theory, developed by Claude Shannon (1916–2001), and cybernetics, by Norbert Wiener (1894–1964), were theoretical advances born from practical needs that soon induced new developments in psychological theory as well as in the world of automation; post-traumatic stress disorders and the demand for help from clinical psychologists stimulated research in this field emphasizing the relationship between organic (somatic) reactions and social experiences. War tied together theory and practice, after all.

## **6. THEORY AND PRACTICE: THE INFLUENCE OF THE ERA OF SCHOOLS**

Theoretical psychology experienced in the first half of the 20th century an extraordinary diversification of its viewpoints. That period has been frequently called the era of schools. Some viewpoints became largely dominant, and conflicts between them arose both in theoretical and practical questions. Behaviorism, Gestalt,

psychoanalysis, and the humanistic approach were among the most influential perspectives in the United States; in the European countries, on the other hand, functionalism and factorialism became very influential. In certain cases, a special doctrine was dominant—as is the case of the Marxist-based Soviet psychology that monopolized Russia under the Soviet regime, or the Piagetian school that influenced European child and school psychology in the middle of the century. These schools of thought are briefly described in the following sections.

1. Behaviorism, created by John B. Watson (1878–1958), made behavior the proper object of study for psychologists, while excluding mind and consciousness as “subjective non-public entities” and, as such, unsuitable for science. Built upon an S-R (stimulus-response) paradigm, theory was conceived of as a guide for exerting technological control over behavior. Since its early days, behaviorism advocated significant applied research. Watson himself wrote a book on child training that was largely influential among U.S. families in the 1930s; he was also for many years a practitioner of commercial and advertising techniques.

Predicting and changing behavior became an essential function of psychologists; the practical value of such control gave birth to an entire field, “behavior modification,” that applied the laws of learning and motivation to changing people’s behavior. Pioneering work was carried out by one of Watson’s students, Mary Cover Jones, who, based on Watson’s ideas of conditioning, devised a successful treatment of a phobia in a child.

Behavior modification appears to be a case of a multiple discoveries that took place in the early 1950s. Burrus F. Skinner (1904–1990) in the United States, Joseph Wolpe (1915–1997) in South Africa, and Hans J. Eysenck (1916–1997) in Great Britain all had the same idea of conceiving mental disorders as learned (maladaptive) behaviors, subject to change for the better (re-learning) under adequate manipulation of effective variables. The new approach rapidly spread to the Western countries. In the United States, a large group of researchers oriented themselves toward the applied behavior analysis, devising interventions on social problems according to its principles. In Latin America, where psychoanalysis largely dominated in the 1940s and 1950s, behavioral techniques appeared as a greatly desired alternative in clinical practice. Under the influence of some U.S. specialists, the movement began to consolidate in the 1960s in Brazil, Mexico, Chile,

Venezuela, and Colombia, and the Latin-American Association for the Behavior Analysis and Modification (ALAMOC) was founded in 1975. A similar phenomenon took place in European countries, where these techniques proliferated in the 1960s, and the European Association for Behavior Therapy was created in 1970. Adepts to these principles were oriented toward practice or toward theoretical questions. Techniques for handling stimuli and motives have been devised and applied to nearly every human situation—schools, family, industries, therapies, and so on. While very effective in multiple settings, behaviorism has also been criticized for its appearance of merely dealing with symptoms, with total neglect for the causes of behavior. After the emergence of the cognitive paradigm, a combined approach—the cognitive behavior modification—was generated; it focused on cognitive control and influence upon behavior as a means of rebuilding a healthy state.

2. Psychoanalysis, created by Austrian psychiatrist Sigmund Freud (1856–1939), has turned into one of the biggest intellectual and social constructions of the 20th century, influencing all subject matters related to humankind, including religion, society, art, sex, and morality. In its applied version, it has largely influenced the clinical approaches to the treatment of mental and behavioral disorders. Two points should be mentioned here: (1) as a form of psychotherapy, psychoanalysis has a few core concepts—the unconscious determination of conscious behavior, the presence of dynamic conflicts between instinctual forces and other social and moral repressive ones, etc.—that have adopted an endless variety of forms maintained by hundreds of schools and groups, but that have raised strong doubts about the therapeutic efficacy and the scientific status of their basic assumptions. In many cases, hermeneutic and philosophical nature of psychoanalysis has been acknowledged, while leaving in suspense its scientific claims. And (2), most of its practitioners have received medical training and licenses, and their techniques are viewed as an essential part of contemporary psychiatry.

3. Gestalt psychology, initiated in Germany in 1912 under the inspiration of German psychologist Max Wertheimer (1886–1943), claimed that all behavior always follows a mental structure in which all stimuli coming from the outside world and from the subject's body form a mental whole that determines open responses to situations. Although it is mainly a theoretical construction, a large number of applications were devised by one of its leading figures, K. Lewin,

to solve problems in school and organizational settings. Gestalt theorists put great emphasis on mental variables such as perception, expectations, values, need for achievement, leadership, and group pressure that proved useful in predicting and explaining behavior in different types of social situations. Important developments on group processes were carried out by former Lewin students Dorwin Cartwright, Leo Festinger, and Stanley Schachter, among others; they received much attention in social and clinical settings.

4. Humanistic psychology (sometimes called “the third force”) is a reaction to the reductionism imposed by behaviorism and to the irrationality and unconsciousness emphasized by psychoanalysis. It stresses meaning and values as essential determinants of human behavior. Self-fulfillment, life goals, and personal growth are some of the concepts that are at the core of this movement. Born in the United States, it has been largely influenced by European philosophies such as M. Heidegger and J. P. Sartre's existentialism and E. Husserl's phenomenology. Abraham Maslow (1908–1970), Carl Rogers (1902–1987), and Charlotte Bühler (1883–1974) are some of the names associated with it; all of them significantly contributed to psychotherapy, and their insights were taken as bases for a technique of personal development that proved fruitful in the world of management and human relations. Close to these views are some other clinical approaches that also emphasize the role of subjective feelings, cognition, and emotions; these processes deeply influence the subject's mind in its interactions with the world. A. T. Beck's cognitive therapy, A. Ellis's rational emotive therapy, and F. Perls's Gestalt therapy, among others, have widened the spectrum of ways to intervene in psychopathological problems.

5. While these schools largely dominated the U.S. scene, different lines of thought were prevalent in Europe, such as functionalism, a movement with American roots (e.g., William James, John Dewey) but with robust development in Great Britain, Switzerland, and other Western countries. Functionalism stressed the relevance of the “what for” considerations when applied to psychological acts and aptitudes, and mainly focused on life's needs and demands. Industrial psychologist Charles S. Myers (1873–1947) exerted great influence on some of his students; among them was Sir Frederic Bartlett (1886–1969), who was interested in real life problems, and who did significant work on memory and thinking inside a functionalist framework. Swiss functionalist Edouard Claparède applied these views to child and educational psychology, and one of

his collaborators, the Swiss theorist Jean Piaget (1896–1980), deepened those insights and built a major system in epistemology and knowledge, offering a solid view on the biological bases of cognitive development—from reflex action to logical operations—and its sequential stages. As is well known, Piagetian ideas have greatly influenced education in many countries.

6. Last but not least, the Marxian philosophy (historical materialism) that inspired the Soviet regime in Russia (the Soviet Union between 1917 and 1990) also was the basis for a worldview embracing humankind and society that influenced all dimensions of knowledge. Psychology was no exception.

Political theorists such as Karl Marx (1818–1883) and Vladimir Lenin (1870–1924) were considered conceptual leaders in every scientific and cultural field. Human life, guided by the conscious mind, was organized through a process of active interactions with the world, but human actions received their ultimate meaning from the socioeconomic structure in which people lived. Consciousness ultimately depended on the production system dominating society. As a consequence, it stressed the unity between consciousness and activity, and, furthermore, between theory and practice.

Lenin, while denouncing “bourgeois exploitation” in Taylorism and other Western labor theories, gave priority to those studies related to work. In 1927, the All-Russia Psychotechnical Society was established, and in 1931 the 7th International Psychotechnical Conference took place in Moscow. Well-known specialists in educational matters, such as Nadezhda Krupskaya (1869–1939), Isaak N. Spielrein (1891–1936?), and Lev Vygotsky (1896–1934), tried to synthesize the Marxist view with recent Western developments; all of them suffered from prosecution in the days of Jozef Stalin’s personality cult in the late 1930s, and for nearly two decades their ideas and works were banned and silenced. During this time, psychology focused on practical (social) activity as causing mental organization through brain structures. Marxian views seemed to fit well with Pavlovian psychophysiology, and theoretical and applied questions became more closely connected. The significant growth of Soviet psychology since the 1950s was based on the cooperation of these two lines of thought. Alexei Leontiev’s (1903–1979) idea of activity, largely based on Vygotskian analyses, spread out to other fields, such as Alexander Luria’s (1902–1977) defectology and neuropsychology. The world race for space supremacy, that in the 1960s involved the United States and the USSR, stimulated man–machine

research, with outstanding contributions such as Boris Lomov’s (1927–1989) industrial psychology and Vladimir Zinchenko’s ergonomics, among others.

The end of the Soviet regime in 1990 brought with it the fall of its previous political and philosophical framework, and a new freedom for thought emerged, in which many Western schools and tendencies then flourished.

## 7. APPLIED PSYCHOLOGY IN THE SECOND HALF OF THE 20TH CENTURY

In the decades that followed World War II, enormous growth took place in psychology that covered most of the facets of human activity and work, in many cases producing interdisciplinary research. Theory also experienced significant changes that have been frequently interpreted as the rise of a new, cognitive paradigm. According to this paradigm, humans should be viewed as information-processing systems, and consciousness, cognitions, affections, and emotions reflect the richness and complexity of such a system. Substantial inspiration for the new paradigm came from applied areas. Swiftly developing computer technology opened new ways to the study of behavior based on the computer model (metaphor) of the brain.

Structure and qualities of behavior referred both to organism and to environment, whose complex interactions had to be analyzed into terms of plans, projects, expectations, and values. Plans and structure of the human operator parallel the software–hardware dimensions of a computer. A vast program launched to examine analogies and possibilities of the new metaphor for psychological research brought up new lines of thought: behavior simulation by computers, artificial intelligence, computer-assisted learning, expert programs of various specialized tasks (e.g., medical diagnosis), and mental activity viewed as the operation of an information-processing system (IPS). Such a system should be viewed as an “open system,” in which the operator unit and its environment exchange information signals, energy, materials, and causative interactions. The entire complex (P-E) integrates both person (P) and environment (E) in continuous interaction.

A cursory review of some of the more significant developments that took place in this period are presented here, without any pretension of exhaustiveness.

### **7.1. Ecological Psychology**

The P-E model, as described above, is at the core of an ecological approach to psychology. The environment is mainly viewed in its material aspect, and human behavior is then examined in the light of the physical determinants of its medium and their influences and constraints upon the activity of the subject. Roger Barker (1968), influenced by Lewin's ideas, conceived of the field as a study of objective contexts influencing behavior, a hypothesis full of consequences for other disciplines such as urban and home designs, geriatrics, and child care, among others. At present, environmental psychology, defined as a study of the relationships between people and their physical environments, examines a large range of questions, from coping with disaster or crime and environment to developing positive attitudes toward nature and protecting nature against human destructiveness.

### **7.2. Traffic Safety Psychology**

Closely connected with the preceding, psychology has developed conceptual and technical approaches to behavior of drivers and pedestrians in traffic. Many psychological functions (perceptual, motor, emotions, risk-seeking, and so on) have a strong influence in determining traffic accidents. Although this field is as old as applied research, recent developments in motivation, risky decisions, and accident-prone personality have added new dimensions to it. In several countries, psychologists are now working on the screening of potentially dangerous drivers and in the design of public campaigns for the promotion of safe attitudes in driving situations.

### **7.3. Organizational Psychology**

Organizations are the most frequent social environment for women and men in developed societies. Industries, educative and health institutions, and political organizations are social structures imposing facilities and constraints on their members, which in turn activate their psychological abilities to adapt and to achieve their goals inside those structures. As a matter of fact, this denomination has nearly come to represent what industrial psychology was covering some decades ago (e.g., personnel selection, human relations), stressing its structural approach to these types of problems.

Some significant lines of thought on this field that developed in the early decades of the century were

noted earlier. In the second half of the century, organizational psychology has grown, and in some cases it has dominated the sociotechnological approach. Other groups stressed the relevance of humanistic aspects and of combining the needs of both the individual and the organization, trying to arrive at a situation of balanced forces contributing to the whole structure and viewing organizational behavior as based on bounded rationality.

Organizations operate in a changing world, and need to maintain their adjustment to the environment through changes taking place inside and outside their own boundaries. A vast technology has been developed to provide them with techniques and means for their development. Organizations, through changes in structure, culture, competence, and processes, aim at improving their problem solving and conflict resolution and their own climate and efficiency. Far from the old lines of rationalist scientific management, the organizational development, rooted in the action research theory designed by K. Lewin in the 1940s, operates with psychological variables through micro and macro techniques, including program assessment and redesigning of structures; its topics have become central in the research field.

Special cases of organizations are armies, churches, national administrations, and so on. These large social structures generate enough questions to become the particular object of psychological specialties covering them. The use of telematics and new information and communication technologies applied to organizations have also raised new demands on workers' formation and capacities, an interdisciplinary field that many psychologists are now entering.

Other fields, such as the large one of sport behavior or military psychology, can also be mentioned in this regard. In the former, psychologists have proved to be efficient technicians in dealing with variables that determine high quality sports performances (e.g., achievement motivation, emotional and attention control). The latter includes the study of specific organizations (armies) that were among the first clients of psychology during wartime. In this field, there is clear overlap with similar areas such as engineering psychology and equipment design and ergonomics. In recent times, emphasis has turned to relevant personality dimensions like authoritarianism, leadership, conformity, and prejudice that set the basis for social conflicts. During the so-called Cold War between Western democracies and the Soviet world, some topics also gained relevance, such as the "war of nerves," national morals, and conflict simulation.

In their approach to such complex problems, some theorists have explored the multiplicity of determinants that influence human action, trying to understand the cues and factors that operate in effective choices and decisions; such insights have enlarged the well-known Lewinian “action theory,” which tries to strengthen theory and practice bonds.

#### 7.4. Health Psychology

This branch, which is closely related to the clinical field, has rapidly developed since the 1980s. It concentrates on all psychological questions related to health maintenance, illness prevention, and patient-practitioner interactions in treating diseases. This specialty deals not only with problems of the individual, but also with others of an organizational nature linked to health care systems and health policy. Prevention, adherence to doctor’s instructions, and people’s healthy life habits are now among the goals of psychologists working in this field.

Connected to this field is contemporary psychogerontology, a specialty that emerged from studies on lifespan developmental psychology and turned into one of the most active areas of research and social priority. In many Western countries, the elderly occupy one of the largest segment of the population, and retirement arrives at an age still able for productive activity. Leisure, social deprivation, loss of power and influence, depressive moods, and psychological losses became relevant targets for psychological intervention.

A special chapter has recently developed and is receiving great attention from society: stress research. Growing evidence shows a causal influence of psychological factors upon the immune system of organisms. Stress research largely deals with the relationship between psychological and organic variables and stressing experiences. The style of coping with negative effects seems not unrelated to serious diseases such as cancer, heart attacks, and AIDS, and with those organic processes connected with the end of such processes. For instance, much literature has been produced on A personality types, considered by some researchers to be risk factors of many serious diseases, mediating some changes in the bodily immune system.

It is also the important and complex role assumed by psychologists working on drug-dependence assistance, one of the most serious problems in contemporary society. Abuse and addiction therapies demand both physical and psychological attention, and patients require social support and advice for changing their

attitudes toward dangerous substances and to adhere to treatment.

Finally, the clinical field is also neuropsychology. This is one of the areas that has experienced extraordinary developments in recent times. It is usually described as the study of the brain machinery underlying psychological functions in patients suffering from brain damage, carried out for practical purposes of rehabilitation. Its conceptual roots can be found in the works of Alexander R. Luria (1902–1977) and Donald O. Hebb (1904–1985). Technical advances in brain diagnosis through direct imagery (positron-emission tomography and the like) have combined with systematic testing of behavioral effects of damage through batteries such as the Halstead-Reitan Neuropsychological test battery, the Luria/Christensen test battery, and many others. Cognitive and emotional deficits and perturbations, while significant for the clinician, are also relevant for the general theorist, who tries to combine the experimental cognitive approach with new neuropsychological information in order to deepen the knowledge on the mind–brain relationship. This is clearly an interdisciplinary field that needs the joint labor of neurologists, psychologists, and other scientists.

#### 7.5. Law Psychology

Since its very beginning, applied psychology has been charged with the task of establishing the degree of credibility of eyewitness testimonies presented in court. Karl Marbe (1869–1953) studied child charges of sexual abuse, and, interested in the topic, wrote a pioneering handbook on forensic psychology. U.S. clinics for offenders’ rehabilitation created by Healey have been mentioned previously. Since then, questions of deception in testimonies, jury selection, legal evidence in court, deterrence from crime, rehabilitation of offenders, and assessment of convict personalities, among others, are under scrutiny and discussion. In the United States, an important court sentence that accepted the psychologist as an expert on mental states of defendants in court strongly promoted the specialty.

#### 7.6. Other Specialties

Human behavior offers endless dimensions and aspects that may be considered by the psychologist. Apart from those mentioned above, other specialties that are now flourishing in developed countries include community

psychology, economic psychology, and political psychology. Psychological care is increasingly a right for people living in developed countries. These provide their citizens with a welfare system that offers security and support when individuals and groups have to cope with stressful events, which frequently include psychological connotations. Problems of family life, such as abuse and neglect of children and women, victimology, and crime and drug prevention, require planning that should include psychological assessment and measures in order to be effective.

Modern societies rely on a large economic system, that through tax policies and contributions supports the major needs of administrative and social services that are in charge of the state. While societies for centuries employed constrictive measures, more attractive ones have been introduced by democratic regimes, which try to persuade people to act lawfully by spontaneous motivation and reflection instead of by compulsion. Individuals who are forced to participate in supporting the state budget, and who are active members of the market system, are basically guided by their interests, beliefs, and fear of penal enforcement. Techniques employed to increase efficiency and motivation in business in the early days of applied psychology by W. D. Scott and other pioneers have recently been transferred and adapted to social policy by governments and social agencies. Economic psychology today covers a wide variety of topics, including behavior in market settings, taxes, investing, risk assumption, bargaining, publicity, and consumer behaviors.

## 8. GENERAL TRENDS

Most psychological professionals work in the field of applied psychology, and the field produces a large portion of the research in psychology. While it is usual to think that it deals with the mere application of some existing principles to certain circumstances, practical intervention frequently gives birth to new concepts and techniques in a continuous interaction among proposed goals, effective applications, and evaluations.

Applied psychology systematically uses evaluation techniques in order to determine the initial problem, to measure the changes produced in it by psychological intervention, and to assess the resulting costs and benefits. According to Gilgen and Gilgen (1987), "Whatever the country... it seems clear that support for academic research psychology is largely dependent

on how successful psychologists are at demonstrating their usefulness to society."

Modern technological devices (e.g., complex computer-assisted simulation, modeling and data analyses) have been introduced in all social organizations and working groups, and psychologists have assumed the demands created by that technology, and have incorporated them in their own practice. This has resulted in the task of training and recycling specialists as well as a redesign of groups in view of the new goals to be reached.

In every field, interventions have generated the creation of micro-level models that try to fit with concrete data, and whose roots on a general view of mind and behavior are loose and flexible. Assessment runs parallel with intervention, in order to introduce the needed corrections to the planned operation.

Interdisciplinarity has become dominant, and psychologists are approaching the views and methods of those groups with which they have to cooperate. At the present, tremendous development in fields such as genetics, bioengineering, and computer sciences demand further and deeper interaction with psychology, for both theoretical and practical reasons. The new situation demands specialization and competence in many scientific disciplines, but at the same time, all progress will depend upon the ability to connect these new ideas to the questions raised by the understanding of the human mind.

In recent decades, psychologists involved in professional practice clearly outnumber those oriented toward academic research. Their relative weight has turned upside-down the situation that existed in the early days. In addition to the debates on scientific topics, new questions are demanding solutions from professionals: the implementation of adequate training (in concepts and skills) and a qualifications for licensing in different specialties; the adaptation of some minimal technical standards to the peculiarities and idiosyncrasies of the different societies and cultures in which psychology is now applied; communication between academics and the practical world, benefiting from recent findings; the creation of mechanisms for accumulating and disseminating the growing body of knowledge in useful and standardized ways; the updating of ethical codes in view of the multiplicity of communication systems and computational technologies; and the increasing penetration of psychology in different aspects and dimensions of modern life.

Training for a professional career in psychology increasingly requires a specialization in concrete techniques that run the gamut of related topics. In many cases, this is deferred until the postgraduate level.

International cooperation provides additional possibilities for effective work. Associations such as the International Association of Applied Psychology, the European Federation of Professional Psychologists (recently renamed the European Federation of Psychological Associations), and the American Psychological Association are considering the requirements to become a well-trained professional in the various branches of psychology, and the technical and conceptual requirements required by the growing complex interactions between this professional and its clients.

### See Also the Following Articles

Clinical Assessment ■ Cross-Cultural Psychology, Overview ■ Cultural Psychology ■ Environmental Psychology, Overview ■ Health Psychology, Cross-Cultural ■ Industrial–Organizational Psychology across Cultures

### Further Reading

- Adler, N., & Matthews, K. (1994). Health psychology: Why do some people get sick and some stay well? *Annual Review of Psychology*, 45, 229–259.
- Ardila, R. (1978). *La profesión del psicólogo*. México, Trillas.
- Baker, D. B., & Benjamin, L. T., Jr. (2000). The affirmation of the scientist-practitioner. A look back at Boulder. *American Psychologist*, 55, 241–246.
- Baltes, P. B., & Baltes, M. M. (1990). *Successful aging*. New York: Cambridge Univ. Press.
- Barker, R. (1968). *Ecological psychology*. Stanford: Stanford Univ. Press.
- Baumgarten, F. (1957). *Exámenes de aptitud profesional. Teoría y práctica*. Barcelona, Labor.
- Birren, J. E., & Schaie, K. W. (Eds.) (1990). *Handbook of aging and the individual*. Chicago, Chicago Univ. Press.
- Boden, M. (1988). *Computer models of mind*. New York, Cambridge Univ. Press.
- Brigham, C. (1923). *A study of American intelligence*. Princeton, NJ: Princeton Univ. Press.
- Brozek, J. (1973). Soviet psychology, In Marx, M. H., & Hillix, W. A. (Eds.), *Systems and theories in psychology* (2nd ed., pp. 521–548). New York: McGraw-Hill.
- Buckley, K. W. (1989). *Mechanical man: John Broadus Watson and the beginnings of behaviorism*. New York: Guilford Press.
- Bühler, C., & Allen, M. (1972). *Introduction to humanistic psychology*. Monterey, CA: Brooks & Cole.
- Carpintero, H. (2001). The psychology of aging in historical perspective, In Leon-Carrión, J., & Giannini, M. (Eds.). *Behavioral neurology in the elderly* (pp. 3–22). Boca Raton, FL: CRC Press.
- Chanderssais, C. (1961). *La psicología en las fuerzas armadas*. Buenos Aires: Kapelusz.
- Cohen, S., & Herbert, T. (1996). Health psychology: Psychological factors and physical disease from the perspective of human psychoneuroimmunology. *Annual Review of Psychology*, 47, 113–142.
- Compayré, G. (1898). *Psicología teórica y práctica aplicada a la educación*. Paris: Lib. V. C. Bouret.
- Dorsch, F. (1963). *Geschichte und Probleme der angewandten Psychologie*. Bern: Hans Huber.
- Du Bois, P. (1970). *A history of psychological testing*. Boston: Allyn and Bacon.
- Eysenck, H. J. (1971). *Race, intelligence and education*. London: Temple Smith.
- Fernandez-Ballesteros, R. (1997). *Psicología del envejecimiento: crecimiento y declive*. Madrid: Univ. Autónoma de Madrid.
- Fine, R. (1990). *The history of psychoanalysis* (2nd ed.). New York: Continuum.
- French, W. L., & Bell, C. H. (1978). *Organizational development* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Geuter, U. (1992). *The professionalization of psychology in Nazi Germany*. Cambridge, UK: Cambridge Univ. Press.
- Gilgen, A. R. (1982). *American psychology since World War II: A profile of the discipline*. Westport, CT: Greenwood Press.
- Gilgen, A. R., & Gilgen, C. K. (1987). *International handbook of psychology*. New York: Greenwood Press.
- Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.
- Gundlach, H. (1998). An outline of the history of the IAAP, In *Applied Psychology: The First-Thirteenth Congress Proceedings of the International Association of Applied Psychology*, Vol. I (pp. 1–24). London: Routledge.
- Hearnshaw, L. (1964). *A short history of British psychology*. London: Methuen.
- Herrnstein, R., & Murray, C. (1994). *The Bell Curve: Intelligence and class structure in American life*. New York: The Free Press.
- Holtzman, W. H. (1992). Health psychology, In Rosenzweig, M. (Ed.), *International Psychological Science* (pp. 199–226). Washington, DC: APA-IUPsyS.
- IAREP (1997). *IAREP XXII International Colloquium of Economic Psychology*. Valencia, Promolibro.
- Jaeger, S., & Staeuble, I., (1981). *die psychotechnik und ihre gesellschaftlichen entwicklungsbedingungen*, in *Die Psychologie des 20* (pp. 53–95). Jahrhunderts: Zurich.
- James, W. (1899/1962). *Talks to teachers on psychology and to students on some of life's ideals*. New York: Dover.
- Jensen, A. J. (1973). *Genetics and education*. New York: Harper & Row.
- Johnson-Laird, P. N. (1988). *The computer and the mind. An introduction to cognitive science*. Glasgow: Collins & Sons.
- la Vaissière, J. (1915). *Psychologie pédagogique*. Paris: Beauchesne.
- Leon-Carrion, J. (Ed.) (1997). *Neuropsychological rehabilitation: Fundamentals, innovations, and directions*. Delray Beach, FL: GR/St. Lucie Press.

- Leonova, A. (1995). Applied psychology in the European North-East: Historical traditions, stages of development and current perspectives. In Jaeger, S., et al. (Eds.), *Psychologie Im Sozialkulturellen Wandeln.-Kontinuitäten und Diskontinuitäten* (pp. 186–193). Frankfurt: Peter Lang.
- Lévy-Leboyer, C. (1980). *Psychologie et environnement*. Paris: PUF.
- Lewin, K. (1948). *Resolving social conflicts*. New York: Harper & Row.
- Lopes, L. (1994). Psychology and economics. *Annual Review of Psychology*, 45, 197–227.
- Lunt, I. (1999). The professionalization of psychology in Europe. *European Psychology*, 4, 240–247.
- Marbe, K. (1913). *Grundzüge der forensischen Psychologie*. München: Beck.
- Matarazzo, J. (1980). Behavioral health and behavioral medicine: Frontiers for a new health psychology. *American Psychologist*, 35, 807–817.
- Monahan, J., & Loftus, E. (1982). The psychology of law. *Annual Review of Psychology*, 33, 441–475.
- Montoro, L., et al. (1995). *Seguridad vial: Del factor humano a las nuevas tecnologías*. Madrid: Síntesis.
- Näätänen, R., & Summala, H. (1976). *Road-user behavior and traffic accidents*. Amsterdam: North Holland.
- Otto, R. K., & Heilbrun, K. (2002). The practice of forensic psychology: A look toward the future in light of the past. *American Psychologist*, 57, 5–18.
- Peiró, J. M. (1983). *Psicología de la organización*. Madrid: UNED.
- Perusia, F. (1994). *Psicologo: Storia e attualità di una professione scientifica*. Torino: Bollati Boringhieri.
- Petrovsky, A. (1990). *Psychology in the Soviet Union*. Moscow, Progress.
- Raimy, V. C. (Ed.) (1950). *Training in clinical psychology*. New York: Prentice Hall.
- Reisman, J. M. (1991). *A history of clinical psychology* (2nd ed.). New York: Hemisphere.
- Rodriguez Marin, J. (1995). *Psicología social de la salud*. Madrid: Síntesis.
- Rüegsegger, R. (1986). *Die Geschichte der angewandten Psychologie, 1900–1940*. Bern: Huber.
- Sexton, V. S., & Hogan, J. D. (1992). *International psychology. Views from around the world*. Lincoln, NE: Univ. of Nebraska Press.
- Sollier, P., & Drabs, J. (1935). *La psychotechnique*. Bruxelles: Comité Central Industriel de Belgique.
- Sprung, L., & Schönplflug, W. (Eds.). (2003). *Zur Geschichte der Psychologie in Berlin* (2nd ed.). Frankfurt: Peter Lang.
- Stokols, D. (1978). Environmental psychology. *Annual Review of Psychology*, 29, 253–295.
- Sully, J. (1900). *Psicología pedagógica*. New York: Appleton (originally *The teacher's handbook of psychology*. London: Longmans Green, 1886).
- Sundstrom, E., et al. (1996). Environmental psychology, 1989–1994. *Annual Review of Psychology*, 47, 485–512.
- Thorndike, E. L. (1910). *Educational psychology* (2nd ed.). New York: Columbia Univ.
- Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon efficiency of other functions. *Psychology Review*, 8, 247–261, 384–395, 553–564.
- Wilpert, B. (1995). Organizational behavior. *Annual Review of Psychology*, 46, 59–90.
- Wilpert, B., & Lunt, I. (1998). *World study of applied psychology*. Unpublished report to IAAP.
- Wrightsmann, L. (2001). *Forensic psychology*. Belmont, WA: Wadsworth.
- Zilboorg, G., & Henry, G. W. (1941). *A history of medical psychology*. New York: Norton.



# Holland's Theory (Vocational Personality Types)

W. Bruce Walsh

*The Ohio State University, Columbus, Ohio, USA*

1. The Personality Types
  2. The Model Environments
  3. Person–Environment Congruence
  4. Consistency, Differentiation, and Identity
  5. Research
  6. Summary
- Further Reading

## GLOSSARY

**congruence** The degree of fit between personality type and work environment.

**consistency** The degree of relatedness between personality types or between environmental models.

**differentiation** The degree to which a person or environment is well defined.

**identity** The clarity and stability of a person's interests, abilities, and goals.

**type** A cluster of personal attributes.

John Holland's theory of personality types and model environments is a theory of vocational psychology, personality, and person–environment psychology. Holland's theory, as viewed by Helson and Mitchell, is tough, practical, and useful. The theory is operational, and there is much data to support it. In essence, the theory suggests that behavior is a function

of the complementary match between an individual's personality style and the psychological environment. Holland suggests that individuals enter environments because of their personalities and remain in those environments because of the reinforcements and satisfactions obtained through interpersonal interactions in the environments.

Holland's theory assumes that behavior is a function of personality and social environment. The theory emphasizes interests and personality since to Holland the choice of an occupation is an expression of personality; thus, interest inventories are personality inventories. The environmental component of the theory is linked to the notion that people in a particular environment tend to have similar personalities and histories of personal development. Because people in a given environment have similar personalities, Holland believes they tend to cope in similar ways. In general, Holland holds that complementary person–environment interactions are reinforcing and satisfying and contribute to job stability, work quality, and psychological well-being.

## 1. THE PERSONALITY TYPES

Holland attempts to explain behavior by using a few well-defined ideas. These assumptions of his theory

tend to be fairly practical, straightforward, and operational in terms of data-collection procedures. The first of these assumptions is that people learn to be one or more personality types. A type is defined as a cluster of personal attributes that may be used to assess the person. Holland's six personality types are described as follows:

1. The realistic (R) type likes realistic jobs such as automobile mechanic, farmer, or electrician. R's have mechanical abilities and tend to be conforming, honest, materialistic, natural, persistent, practical, modest, and stable.

2. The investigative (I) type likes investigative jobs such as chemist, physicist, biologist, or geologist. I's have mathematical and scientific ability and are described as analytical, independent, intellectual, introverted, methodological, and rational.

3. The artistic (A) type likes artistic jobs such as musician, writer, or actor. A's have writing, music, and/or artistic abilities and are described as complicated, emotional, expressive, imaginative, impulsive, nonconforming, and original.

4. The social (S) type likes social jobs such as teacher, counseling psychologist, clinical psychologist, or speech therapist. S's have social skills and are described as cooperative, friendly, helpful, insightful, responsible, sociable, and understanding.

5. The enterprising (E) type likes enterprising jobs such as salesperson, manager, or business executive. E's have leadership and verbal abilities and are described as ambitious, domineering, pleasure seeking, self-confident, and sociable.

6. The conventional (C) type likes conventional jobs such as bookkeeper, banker, or tax expert. C's have clerical and arithmetic ability and are described as conforming, conscientious, orderly, persistent, practical, and self-controlled.

A person's resemblance to each of the personality types is assessed by the use of the Vocational Preference Inventory, the Self-Directed Search, and the Strong Interest Inventory. All of these inventories are objective in nature and collect interval kinds of data that may be used in a variety of research settings. Much of the research on Holland's theory has used one of these instruments in a field research setting.

## 2. THE MODEL ENVIRONMENTS

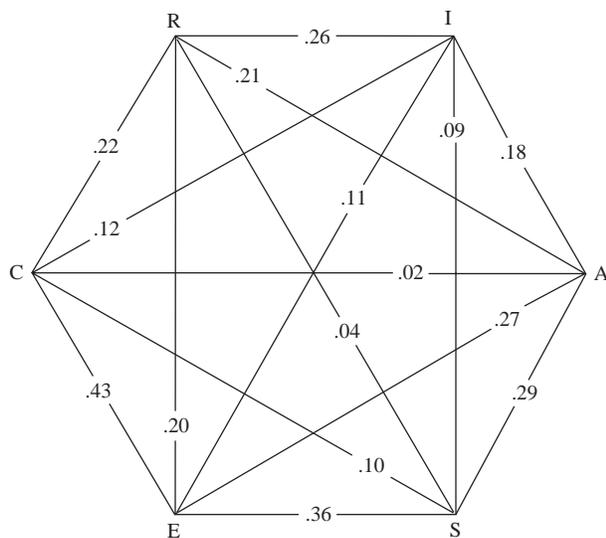
A second idea used by Holland to explain behavior is that environments in which people live and work may be

characterized by their resemblance to one or more model environments. Holland suggests six model environments corresponding to the six personality types and believes that people tend to move toward environments congruent with their personality types: Realistic types tend to move toward realistic environments, enterprising types tend to move toward enterprising environments, and so forth. To assess environments, Holland and Astin developed the Environmental Assessment Technique. This assessment technique simply makes a count or a census of the number of different personality types existing or interacting in a given environment. The personality type occurring most frequently in a given environment is thought to be dominant in the environment and to have a psychological influence in that specific environment. In summary, Holland defined the environment psychologically according to the people who were actually interacting in it, or, as he would say, people define the environment. Thus, although the environment is determined by a count, the influence is thought to be psychological in nature.

## 3. PERSON-ENVIRONMENT CONGRUENCE

A third idea used by Holland is that person-environment congruence tends to be associated with satisfaction, productivity, personal stability, vocational stability, and well-being. The thought is that, for example, a realistic type in a realistic environment would tend to be more productive, more stable, more vocationally satisfied, and psychologically healthier. Complementary person-environment links are reinforcing and satisfying. Uncomplementary or incongruent person-environment interactions are punishing and contribute to change. This is not to say that congruence is good and incongruence is bad but suggests that a state of incongruence may be part of the learning and development process. Thus, there seems to be periods of incongruence in an individual's life span that need to be worked through in order to pursue personal and vocational goals and well-being.

Based on his six personality-environment types, Holland further suggests that the relations among types, among environments, and between types and environments tend to be hexagonal in nature (Fig. 1). Types and environments that are in closest proximity in Fig. 1 are more psychologically related. Types and environments that are further removed from one another are more psychologically different. For example, an artistic person



**FIGURE 1** Hexagonal model for interpreting inter- and intra-class relationships. Correlations are between summary scale scores for females ( $n = 1600$ ) in the 1994 normative sample. Reproduced by special permission of the publisher, from the *Self-Directed Search Technical Manual* by John L. Holland, Ph.D. Copyright 1985, 1987, 1994, by Psychological Assessment Resources, Inc. Further reproduction is prohibited without permission of PAR, Inc.

in an artistic environment is in a more congruent situation than an artistic person in a social environment. An artistic person in a conventional environment is in the most incongruent situation possible. In this way, the hexagonal model may be used to obtain and estimate different degrees or levels of person–environment congruence. The hexagonal model for the six personality types and environments has also been used to identify individuals in vocational or personal conflicts, compare different tests and inventories, organize occupational environments, and guide the layout of the Strong Interest Inventory.

#### 4. CONSISTENCY, DIFFERENTIATION, AND IDENTITY

The previous assumptions, as noted by Holland, are extended by several secondary assumptions that may be applied to both persons and environments. The secondary assumptions include consistency, differentiation, and identity. Consistency is the degree of relatedness between personality types or between environmental models, as noted previously. Some pairs of types and

environments are more closely related than others. Differentiation refers to the degree to which a person or an environment is well defined. For example, a person closely resembling a single type is clearly differentiated (well defined). Identity is primarily concerned with estimating the clarity and stability of a person's vocational identity. High identity suggests a limited set of consistent and explicit goals.

#### 5. RESEARCH

In general, research testing the theory (more than 500 studies) reported by Holland and Walsh and Holland clearly indicates that individuals tend to choose and enter environments consistent with their personality types. For example, enterprising types tend to enter and remain in enterprising kinds of environments. Furthermore, evidence suggests that congruent person–environment interactions are conducive to personal and vocational stability and satisfaction and, to a lesser extent, actual achievement. Specifically, the pattern of findings suggests to some extent that person–environment congruence is related to measures of job satisfaction and stability, job involvement, work quality, productivity, and well-being but not necessarily to measures of decision making, sociability, and problem-solving ability. In addition, meaningful data about adolescents and adults imply that person–environment congruence and job satisfaction may increase with age. Thus, it seems that people in congruent work situations will probably change little since their expression of interests, abilities, and personality is rewarded. Persons in incongruent work environments will probably change the most since they tend to be ignored and punished. This work, then, indicates that students and workers who tend to be functioning in environments congruent with their personal characteristics are probably psychologically healthier and more satisfied than people in incongruent situations.

#### 6. SUMMARY

In Holland's theory, behavior is viewed as a function of person–environment interactions, and from this perspective person–environment congruence becomes a very important concept. Thus, Holland suggests, and the data tend to support, the notion that people in environments congruent with their personality types tend to be psychologically healthier, more satisfied,

and more productive than people in incongruent environments. In other words, person–environment congruence tends to facilitate personal growth, self-acceptance, and life purpose.

### See Also the Following Articles

Person–Environment Fit ■ Traits ■ Vocational Interests  
■ Work Adjustment

### Further Reading

Astin, A. W., & Holland, J. L. (1961). The environmental assessment technique: A way to measure college environments. *Journal of Educational Psychology*, *52*, 308–316.

Harmon, L. W., Hansen, J. C., Borgen, F. H., & Hammer, A. L. (1994). *Strong Interest Inventory: Applications and technical guide*. Stanford, CA: Stanford University Press.

Helson, R., & Mitchell, V. (1978). Personality. In M. R. Rosenzweig, & L. W. Porter (Eds.), *Annual review of psychology* (pp. 555–585). Palo Alto, CA: Annual Reviews.

Holland, J. L. (1985). *Vocational Preference Inventory*. Odessa, FL: Psychological Assessment Resources.

Holland, J. L. (1994). *Self-Directed Search (SDS) Form R*. Odessa, FL: Psychological Assessment Resources.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Spokane, A. R. (1985). A review of research on person–environment congruence in Holland's theory of careers. *Journal of Vocational Behavior*, *26*, 306–343.

Spokane, A. R., Meir, E. I., & Catalano, M. (2000). Person–environment congruence and Holland's theory: A review and reconsideration. *Journal of Vocational Behavior*, *57*, 137–178.

Walsh, W. B., & Holland, J. L. (1992). A theory of personality types and work environments. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person–environment psychology*. Mahwah, NJ: Erlbaum.



# Home–School Collaboration

Richard J. Cowan

Kent State University, Kent, Ohio, USA

Susan M. Swearer and Susan M. Sheridan

University of Nebraska, Lincoln, Nebraska, USA

1. Home–School Collaboration Defined
  2. Background and History of Home–School Collaboration
  3. Factors Influencing Home–School Collaboration
  4. Approaches to Fostering and Maintaining Home–School Collaboration
  5. Conclusion
- Further Reading

## GLOSSARY

**collaboration** Working together toward a common goal or set of goals.

**educators** General education instructors, special education instructors, teaching aides, classroom volunteers, administrators, support staff, school psychologists, and any other persons within a school setting who contribute to the education of students within that environment.

**parent** The adult who is responsible for the care and well-being of the child (e.g., biological parent, caregiver, legal guardian).

**partnership** A coequal interdependent relationship that is established and developed over time, with the primary purpose of working together over an extended period of time toward a mutually determined set of goals and objectives.

**system** A particular environment with an emphasis on specific components of the environment, including people, physical objects, placement of people and objects, physical arrangement, and interactions among these components (e.g., home system; school system).

Home–school collaboration is a reciprocal dynamic process that occurs between at least one parent

(or guardian) and at least one individual within the school system (e.g., educators, administrators, psychologists) who share in decision making regarding mutually determined goals and solutions related to a student for whom all parties share interest and responsibility. Home–school collaboration may occur at three levels: system (i.e., between the home and school settings), school/classroom, and individual. With all participants (e.g., parents, educators) offering their unique roles and contributions, the collaboration process is guided by the primary emphasis on mutually determined academic and behavioral goals serving as the standard for progress toward overall outcomes.

## 1. HOME–SCHOOL COLLABORATION DEFINED

A common definition of the term “collaboration” is to work together toward a common goal or set of goals. As applied to relationships among parents, educators, and the systems in which they operate, the term has a rich meaning. Driven by academic, cultural, and political influences, the term “home–school collaboration” is a reciprocal dynamic process that occurs among systems (e.g., families, communities, partnerships), schools/classrooms, and/or individuals (e.g., parents, educators, administrators, psychologists) who share in decision making toward common goals and solutions related to students. Inherent in this definition is the notion that parents and educators involved in collaboration pool

their resources to create a cooperative interdependent relationship. For example, parents offer what they know about their children's academic and behavioral strengths and limitations in the home and other non-school settings, whereas teachers offer their knowledge of students' relative strengths and limitations in the classroom and other educational settings. The collaboration process is guided by a primary emphasis on specific cooperatively predetermined outcomes for students, with mutually established academic and behavioral goals.

## **2. BACKGROUND AND HISTORY OF HOME-SCHOOL COLLABORATION**

### **2.1. Traditional Family Involvement in Education**

During colonial times and throughout the early 1900s, parents had very little interaction with the educational system. In 1995, Welch and Sheridan noted that during that time in educational history, children were sent to school to learn academic skills, and parents were seen as responsible for educating their children regarding moral principles in the home setting. Parents entrusted the education of their children to school personnel and expressed little interest in participating in their children's education. These expectations and roles resulted in educators viewing the school as their territory, and parents who tried to become an active part of the school setting were viewed as intrusive. In response to this situation, the National Parent-Teacher Association (PTA) was founded on February 17, 1897, in Washington, D.C., with the goal of having a PTA in every school. Today, there are more than 6.5 million members of the National PTA, making it the largest child advocacy organization in the United States.

Welch and Sheridan reported that parents continued to assume a relatively "hands-off" approach to education until the late 1960s and early 1970s, when economic and political factors induced parents to change their views regarding their involvement in education. Fueled by economic situations that forced society to reconsider how money was being spent on education, as well as by a political climate in which legislation was identifying and defining specific roles for parents in their children's education, parents became more active in education. The National PTA continued to be a strong advocacy voice for parent involvement in education. In 1994, the federal government enacted "Goals 2000," and in 1997,

the National PTA released its National Standards for Parent/Family Involvement Programs. However, although parents advocated for participation in education, the educational system was not prepared for this type of parental involvement. Despite the efforts of the National PTA, when parents participated in education, their involvement was limited in both scope and depth. Unfortunately, parental involvement in education has continued to be limited. Many parents feel as though they are outsiders in their children's school. Merely opening the doors for parents to participate at an entry level is not enough. Epstein has written extensively about the need to develop family-school-community partnerships to enhance student performance. These partnerships reflect an important process that includes all families. The next subsection describes a philosophy to guide parental involvement in education: the home-school partnership model for working with families.

### **2.2. Partnering With Families in Education: A Modern Approach to Collaboration**

The definition of home-school collaboration provided earlier is derived in part from the characteristics of a partnership philosophy for interacting with families in education. This philosophy represents the current standard for how parents and families can be included as a member of the educational system. A partnership philosophy regarding home-school collaboration recognizes that parents are invaluable members of the educational team. Just as educators and administrators are viewed as possessing educational information and expertise, parents are viewed as experts possessing equally critical information regarding the home system. When schools adopt this attitude, they are more likely to actively pursue meaningful parental participation in education. If parents and educators adopt this philosophy, it affords parents equal status with the school regarding critical activities and decision making related to their children's education.

A primary difference between home-school collaboration and home-school partnership is that home-school collaboration implies a process related to a specific goal or set of goals that may be relatively short term (but not necessarily), whereas home-school partnership implies a long-term, ever-evolving relationship between parents and members of the school setting extending beyond time-limited problem solving and goal achievement. A partnership allows parents and educators to work together to improve the educational

system such that it is possible for them to work toward preventing future problems, meaning that all students benefit from the partnership. Another characteristic of a partnership is that individuals learn from one another and pool together expertise and skills in support of a common focus, that is, improved education for all students. Goals and objectives remain at the heart of interactions between parents and educators as they form educational partnerships; however, with partnerships, these goals are likely to be more long term and more encompassing (e.g., creating a safe and diverse atmosphere that addresses the needs of all students). In addition, the quality of the partnership will depend heavily on the type of communication that occurs between parents and educators. Maintaining open communication is likely to make a meaningful contribution to a healthy balanced partnership.

When schools invest in building partnerships by seeking meaningful ways in which to involve parents, many can be considerably enriched. For example, instead of a traditional one-directional approach to parent-teacher conferences, a teacher might approach this event by asking for parental input regarding behavior patterns across nonschool settings and ask for ideas regarding potential solutions. Furthermore, if parents and teachers interact regularly outside of the designated parent-teacher conference time (e.g., through home-school notes, at school activities), conference time might be used more constructively to address goals and solutions that require face-to-face communication. A partnership is not formed and developed during one interaction or activity; rather, it is established and cultivated over time through multiple interactions during which parents and educators come to embrace the attitude and belief that meaningful relationships form the foundation of in-depth, long-term collaboration.

### **2.3. Potential Benefits of Home-School Partnerships**

There are many potential benefits of home-school partnership. For example, collaboration through partnership can enhance communication between parents and educators. In addition, home-school relationships provide an opportunity for parents and educators to better understand what occurs across settings, and this in turn allows for a greater overall understanding of children's backgrounds and current levels of functioning. Establishing a home-school partnership also allows for "joint ownership" of problems and concerns

as well as joint commitment toward educational and behavioral goals for students. Parents and teachers may come to feel supported by each other as they work toward mutually determined goals. Working together also tends to improve relationships between teachers and parents as they begin to understand each other's perceptions regarding children's behavior across settings. In addition to these potential benefits for parents and educators, students may benefit from the process. For example, they may develop better attitudes toward school, improve their performance both in and out of the classroom (e.g., demonstrate higher rates of work completion at school while at the same time demonstrating an increase in homework completion), get along better with teachers and other school-based individuals, and improve specific academic skills (e.g., reading fluency). As with any relationship, the benefits of a partnership might not be immediate, and it will likely take time for all participants to yield such benefits.

### **3. FACTORS INFLUENCING HOME-SCHOOL COLLABORATION**

The educational system in the United States has been heavily influenced by a systems-based approach to understanding and intervening with regard to the educational process. This approach is grounded in ecological theory. Based on the work of Bronfenbrenner, ecological theory is concerned with the interaction between an individual and various contextual systems: the microsystem, mesosystem, exosystem, and macrosystem. The microsystem is concerned with the individual and his or her immediate environment (e.g., a child interacting in a particular classroom), the mesosystem involves the interrelation between major systems in the individual's life (e.g., between home and school), the exosystem is concerned with environments not directly related to the individual but still influencing his or her life (e.g., a parent's workplace), and the macrosystem includes overall cultural or subcultural patterns and influences (e.g., policies, federal and state legislation, national and global economic factors). This framework illustrates how systems influence, and are influenced by, one another, stressing that individuals do not exist in isolation; rather, each individual's interactions across and within systems are multifaceted and multidetermined over time. Home-school collaboration is influenced by multiple factors, including (but not limited to) educational legislation, family characteristics,

school-based variables, and community-related influences. Factors across these systems are not mutually exclusive and are discussed in the following subsections.

### 3.1. Educational Legislation

Education in the United States has forever been changed due to federal legislation that ensures the right to a free and appropriate public education for all individuals, mandates the individualization of services for students with special needs, incorporates the provision of educational services in the least restrictive environment, provides guidelines for educational planning and programming, specifies procedures for identifying students with special needs, highlights roles and responsibilities for local educational agencies, and mandates specific roles for parents in education. Such legislation has led parents, educators, politicians, and others concerned with education to consider how best to foster relationships between parents and educators. Furthermore, it has changed the ways in which parents are involved in their children's education. The following subsections discuss major legislative and political factors affecting how parents and educators operationalize and designate roles and responsibilities in education.

#### 3.1.1. *Individuals with Disabilities Education Act*

The Individuals with Disabilities Education Act (IDEA) is the latest manifestation of the public education law called the Education for All Handicapped Children Act (EAHCA, also referred to as Public Law 94-142). A major emphasis of the 1997 reauthorization of IDEA is the inclusion of specific requirements regarding the active recruitment and participation of parents throughout the evaluation process. Specifically, school personnel are now required to (a) notify parents in advance to ensure that they will have the opportunity to attend each meeting related to their children's educational programming, (b) schedule meetings at a mutually determined time and place (even in the home if necessary and appropriate), and (c) arrange other means of including parents if they are unable to attend the meeting (e.g., teleconferencing).

In addition to ensuring that parents are able to actively participate in meetings related to their children's educational programming, IDEA mandates and outlines several parental rights and responsibilities. Specifically, parents are entitled to input during evaluation, eligibility, and placement. That is, in addition to contributing to the

pool of information that will be used during the assessment of their children and that may result in special education placement, parents are afforded an opportunity to voice their opinions and contribute to the final decisions when the teams (of which parents are key members) determine (a) eligibility (i.e., declare whether students meet the specific predetermined criteria for placement under a qualifying category of disability as defined in IDEA), (b) accommodations that will be necessary to meet students' individual needs (as determined by goals and objectives outlined in the students' team-generated individualized educational plans [IEPs]), and (c) placement (i.e., where children will receive their educational programming, e.g., resource room, general education classroom, some combination thereof).

Once students are identified and placed under the umbrella of special education and related services, they are entitled to a reevaluation at least every 3 years from their initial or latest evaluation to determine whether they continue to meet eligibility criteria for services. Parental permission for reevaluation is mandatory. Parents are also entitled to receive regular reports on their children's progress, and parents may request changes to educational programming (i.e., revisions of children's IEPs) to address any lack of progress toward annual goals. This level of parent recruitment and participation is significantly different from the hands-off approach popular until the late 1960s and early 1970s described earlier. As such, this legislation represents a powerful impetus for change with regard to parent and teacher roles throughout the educational process.

#### 3.1.2. *No Child Left Behind*

"No Child Left Behind" (Public Law 107-110) is the latest manifestation of the public education law called the Elementary and Secondary Education Act (ESEA). As with IDEA, No Child Left Behind designates roles and responsibilities for parents and educators that necessitate home-school communication and cooperation. Perhaps the most salient feature of this legislation is its emphasis on providing information to parents while at the same time holding schools accountable for the quality of education they provide to all students. In addition to mandating that schools provide information to parents regarding their children's individual progress, No Child Left Behind requires states and school districts to give parents information about the relative strengths of each school. Specifically, for each school, they must provide achievement data (combining the performances of all students) broken

out by race, ethnicity, gender, English-language proficiency, migrant status, disability status, and income status. If, on receiving this information, parents believe that their children are not receiving adequate services, they may opt to place the children in better performing schools or tap additional resources such as tutoring and after-school programs.

Although there are clear benefits to educators and administrators (e.g., they are able to have detailed information about each child from which they can adapt teaching techniques, they are able to determine what does and does not work and make decisions based on the data), the emphasis of this legislation is on protecting individual students, parents, and families through providing critical information and options. This legislation provides critical information for educational partners to use in making decisions regarding how to improve the quality and fairness of the educational system. It also provides equal access to information, ensuring that no party has the upper hand. As indicated earlier, equal participation is a defining characteristic of educational partnerships. In addition, Goal 8 of the 1999 National Education Goals report concerns parental participation and states, "By the year 2000, every school will promote partnerships that will increase parental involvement and participation in promoting the social, emotional, and academic growth of children."

In addition to clauses mandating the types of information to which parents are entitled, No Child Left Behind includes clauses mandating that parents have the right and responsibility to be involved at the school, district, and state levels. Specifically, No Child Left Behind states that every school must have a written parent involvement policy developed with and approved by parents. If developed and executed properly, this policy affords parents much power in decision making and in determining educational policy at a local level. No Child Left Behind also calls for a school-parent compact, developed with and approved by parents, that describes how educators and parents will build a partnership to improve student achievement. If a partnership philosophy guides this compact, it can open to doors to multiple levels and types of home-school collaboration, expanding and evolving over time.

No Child Left Behind has resulted in an increased emphasis on site-based management. Site-based management, an approach with a long history, is concerned with decentralizing decision making by transferring authority and resources from state- and district-level offices to more localized units such as governance councils, committees, and teams located within each school

building. This transfer of authority and power to a local level affords educational partners the opportunity to consider the data generated by schools to meet mandate requirements of No Child Left Behind and to make budget- and curriculum-related decisions at an individual school level. These site-based management teams, often composed of principals, teachers, parents, and other professionals, make decisions regarding policies for the schools. This structure necessitates input from parents as decision-making team members. Similar teams exist at the district and state board levels, whereby parents are involved as active developers and monitors regarding school board functions as well as district- and statewide governing bodies.

No Child Left Behind embodies the state of the nation regarding the pivotal role that parents play in the education of their children. This level of parental involvement requires communication and collaboration between parents and school-based personnel. As with IDEA, a complete discussion of No Child Left Behind related to home-school collaboration is beyond the scope of this article. Specific components of the legislation were discussed previously as a means of describing the rights and responsibilities of parents in education as mandated by federal law. How parents respond to such roles varies considerably across individuals and is likely influenced by multiple factors, some of which are described in the next subsection.

### 3.2. Home- and Family-Related Factors

There are many family-related factors that may influence how a family interacts with educators. The quality of the interaction is perhaps most heavily influenced by the family's history with the school, including how the family has been received by the school in the past. For example, during past interactions with the school, was the school hierarchical, that is, acting as if the parents' presence was an intrusion? Alternatively, was the family made to feel comfortable in the school setting? Has the family been proactively involved with the school (i.e., not just when their child is in need of assistance)? Does the family have a child who has been identified as at risk for, or as an individual with, a specific disability? If so, how was the family included in the related meetings and educational programming sequence? In general, the more positive the family's past experiences, the more likely the family is to feel accepted and welcomed as an educational partner.

Other factors that can potentially influence the nature of home-school collaboration are found under the broad umbrella of family culture. As used here, culture extends beyond race and ethnicity to include factors such as acculturation level, beliefs, values, and expectations. Acculturation level is concerned with families' adjustment to the immediate culture when they are not originally from the United States or when they are new to a particular region of the country. If they come from a place where the educational system varies considerably from that in their new surroundings, it may take them time to adjust to the new local norms regarding parental involvement in education. Some families might not view their role as that of active participants and may leave the majority of decisions and actions to educators, whereas others might take an active directive role in working with educators. Family culture also includes beliefs, values, and expectations regarding education and how to interact with the school system. Stemming from cultural, economic, and political influences, these factors are multidetermined. Regardless of whether families consist of first-time parents or parents who have been involved in the educational system long enough to witness significant changes in roles for parents in education, families will require time to adjust to their new roles.

### **3.3. School-Related Factors**

The school is a dynamic system that is composed of multiple individuals and influenced by multiple forces (e.g., culture, economic conditions, politics). When considering school-related factors that may influence the nature of home-school collaboration, it is important to take into account both variables related to individuals within the school and building-specific traits and characteristics. With regard to school personnel, these individuals acquire beliefs, values, and expectations that are influenced by a myriad of cultural, economic, and political influences. These beliefs, values, and expectations evolve over time and influence interactions with parents. In addition, just as parents' behaviors are influenced by past experiences with the school, teachers' behaviors are influenced by past experiences with parents and families.

In addition to factors related to individuals within the school, factors related to the school itself may affect the nature of home-school collaboration. Just as each family is unique in its history and culture, each school takes on an identity that makes it unique. For example, is the school public or private? The choice to send children to

a private school may represent different values and have implications for how parents and educators interact. It is also true that the size and organizational structure of the school may influence the degree to which parents feel comfortable. In general, a larger school is less personal, and this may cause parents to feel somewhat overwhelmed, resulting in less contact and collaboration with the school. The organizational structure also influences how parents interact with school personnel. For example, are teachers and administrators readily available? Are there individuals who take it on themselves to make regular contact with parents? Some school administrators may encourage teachers to maintain regular contact with parents, whereas others may value a top-down approach in which parent contact occurs through administrator-initiated actions and activities (e.g., newsletters, organized meetings). Some schools are warmer, more inclusive, and generally more family-friendly than are others. How a school "feels" to parents and students is influenced by multiple factors, and this makes pinpointing salient factors a challenge.

### **3.4. Community-Related Factors**

Community factors that influence interactions between parents and educators are difficult to isolate given their identification across multiple systems. However, one cannot overlook the powerful influence of the community in shaping individual and systemic behaviors. Broadening the definition of cultural influences, the community system posits its own set of norms, values, and expectations regarding home-school collaboration. Behaviors, attitudes, and values transmitted from individuals in the home and school settings culminate to form a message that guides community behaviors and actions, and this in turn influences individual parents and teachers engaged in collaboration. These may be based on religious, political, or other influences and may vary over time. In addition to influencing how parents and educators interact, the community has the power to support or hinder goals and objectives identified collectively by parents and educators. For example, if parents and teachers agree that having an after-school program might help to reduce problems related to latchkey children (i.e., children who care for themselves after school), will the community embrace this plan? Does it offer supports to make the vision a reality? Do role models exist in the community to encourage more advanced collaboration? Each community has a history that guides immediate behavior, which in turn affects future interactions between various

systems and individuals. This broad-based system is pivotal and cannot be underestimated with regard to the power it has in guiding home-school interactions.

#### **4. APPROACHES TO FOSTERING AND MAINTAINING HOME-SCHOOL COLLABORATION**

##### **4.1. Underlying Assumptions, Roles, and Responsibilities**

Stemming from a myriad of cultural, economic, and political influences, the educational system has called on parents to actively participate in both the general and special education processes. Parents are increasingly becoming more actively involved in planning and decision making at the local, district, and state levels. In addition to participating in traditional activities such as parent-teacher conferences and back-to-school nights, parental involvement occurs through volunteering, planning, and serving on various committees. In addition to these responsibilities, parents of children who are identified with specific disabilities are involved in the assessment and identification of their children's specific strengths and limitations as well as in co-constructing and facilitating their children's IEPs. Parent involvement also occurs in the home and community. When parents read with their children, observe and participate in play, and support homework completion, this type of parent involvement supports the education of their children. Identifying meaningful ways in which parents can participate in their children's education is a task assigned to both parents and educators. If parents and educators engage in an educational partnership, team members will collectively identify and use specific areas of expertise that each member brings to the collaboration process as they make progress toward a mutually determined set of goals and objectives. Furthermore, if a partnership philosophy guides home-school interactions, parents and educators will seek ways in which members of the school system can support the home system.

##### **4.2. How to Involve Families in Education**

Although state and federal legislation mandates specify roles, rights, and responsibilities for parents in education, they do not prescribe specific means for establishing relationships between the home and school settings. Building a partnership is a complex process that does

not follow a "cookbook" approach; rather, fostering this relationship is a process that must be approached with care based on a specific guiding philosophy. One such philosophy was described by Christenson and Sheridan, who in 2001 maintained that partnership is not an approach to home-school interactions; rather, it is a guiding philosophy for educators and parents as they attempt to build a meaningful long-term relationship. Specifically, these authors articulated four components necessary for building meaningful partnerships: approach, attitudes, atmosphere, and actions.

Approach is a means of interacting with families that embraces the values of both the home and the school in constructing the educational experiences of children. This is concerned with educators taking an approach with families that communicates their value, unique knowledge and skills, and necessity to be involved in education. When this approach is taken, it is assumed that parents will be involved in the school curriculum and that educators will support parents so that children are successful in the home setting. Attitudes are the perceptions that families and schools have of one another. This component is concerned with identifying strengths in partners, focusing on these strengths, and assuming a joint responsibility for students' development and education. Atmosphere is the overall climate set for families and educators. Finally, the actions component involves strategies for building long-term partnerships. Actions are not merely activities that include families in a passive manner; rather, they are intended to be behaviors that, over time, actively include families in the educational process. Perhaps the most salient feature of this model is its emphasis on the processes required to build relationships over time.

#### **5. CONCLUSION**

Cultural, economic, and political influences have created an educational system that values parental involvement in the education of all students. As parents and educators have adjusted to their evolving relative roles, they have engaged in home-school collaboration as a means of working together toward achieving mutually determined educational goals for all students. As parents, educators, and other professionals (e.g., school psychologists) have searched for a guiding philosophy regarding their approach to interacting with one another, many have adopted a partnership philosophy. Partnerships are long-term relationships based on the premise that coequal power and coequal participation

are prerequisite conditions for meaningful collaborative endeavors. When this philosophy is adopted, parents and educators share resources, ideas, decision-making power, and ownership for mutually determined goals regarding both the home- and school-based educational curricula. Partnerships in education allow parents to be more intricately involved in school-based educational processes as a means of enhancing home-school collaboration. In addition, schools can assist parents with the home curriculum by participating in activities such as conjoint behavioral consultation and solution-oriented family-school collaboration. Partnerships, like all relationships, take time to form and require mutual cultivation to develop and evolve. However, they can result in both meaningful and effective educational and developmental outcomes for children and yield benefits that cannot be achieved when families and schools work in isolation. With the appropriate approach, atmosphere, attitudes, and actions, home-school partnerships can create an essential condition for optimal success.

### See Also the Following Articles

Consultation Processes in Schools ■ School-Community Partnerships

### Further Reading

Carlson, C. I., Hickman, J., & Horton, C. B. (1992). From blame to solutions: Solution-oriented family-school con-

- sultation. In J. S. Christenson, & J. C. Conoley (Eds.), *Home-school collaboration: Enhancing children's academic and social competence* (pp. 193-213). Silver Spring, MD: National Association of School Psychologists.
- Christenson, S. L., & Conoley, J. C. (Eds.). (1992). *Home-school collaboration: Enhancing children's academic and social competence*. Silver Spring, MD: National Association of School Psychologists.
- Christenson, S. L., & Sheridan, S. M. (2001). *Schools and families: Creating essential connections for learning*. New York: Guilford.
- Epstein, J. L., Sanders, M. G., Simon, B. S., Salinas, K. C., Jansorn, N. R., & Van Voorhis, F. L. (2002). *School, family, and community partnerships: Your handbook for action* (2nd ed.). Thousand Oaks, CA: Corwin.
- Esler, A. N., Godber, Y., & Christenson, S. L. (2002). Best practices in supporting home-school collaboration. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 389-412). Bethesda, MD: National Association of School Psychologists.
- Minke, K. M. (2000). Preventing school problems and promoting school success through family-school community collaboration. In K. M. Minke, & G. C. Bear (Eds.), *Preventing school problems—Promoting school success: Strategies and programs that work* (pp. 377-420). Bethesda, MD: National Association of School Psychologists.
- Telzrow, C. F., & Tankersley, M. (Eds.). (2000). *IDEA amendments of 1997: Practice guidelines for school-based teams*. Bethesda, MD: National Association of School Psychologists.
- Welch, M., & Sheridan, S. M. (1995). *Educational partnerships: Serving students at risk*. Fort Worth, TX: Harcourt Brace.



# Homophobia

Douglas C. Kimmel

City College, City University of New York, New York, USA

1. Heterosexist Bias in Western Culture
  2. Definition of Sexual Prejudice
  3. Internalized Homophobia
  4. Social Attitudes, Physiological Arousal, and Homophobia
  5. Gender Differences in Prejudice: Homosexuals, Bisexuals, and Transgendered Persons
  6. Education, Civil Rights, and Primary Prevention of Hate Crimes
- Further Reading

anatomical sex and seek some degree of sex reassignment procedures, whereas others express their gender identity through external self-presentation and behavior.

Homophobia refers to antipathy toward persons who are thought to be gay, lesbian, homosexual, or deviant from gender stereotypes in ways that suggest a same-sex sexual orientation. It is a form of prejudice and may be related to harassment or violence toward persons thought to be homosexual. Most people in cultures that embody heterosexist values, such as Western societies, adopt attitudes that stigmatize same-sex sexual, erotic, or affectional attractions. These attitudes tend to be more prevalent in men than in women and tend to be more potent when focused on male same-sex attractions than on female ones. Frequently, these prejudicial attitudes are internalized and can lower the self-esteem of lesbians, bisexuals, and gay men. They are also associated with hate crimes toward lesbians and gay men, and in many states such crimes are considered violations of civil rights ordinances. Casual use of terms such as “faggot” may stigmatize youth in school and other settings and should not be tolerated any more than are other prejudicial and derogatory terms. Education with regard to civil rights, diversity in gender-related behavior, and tolerance of individual differences is important at all levels of school and may be most effective when it is modeled by peer leaders and authority figures.

## GLOSSARY

**heterosexual bias** The belief that heterosexuality is normal and that everything associated with heterosexuality is superior to anything associated with homosexuality.

**homophobia** Antipathy toward persons who are thought to be gay, lesbian, homosexual, or deviant from gender stereotypes in ways that suggest a same-sex sexual orientation.

**internalized homophobia** The negative attitude a gay man, bisexual, or lesbian holds about himself or herself as a result of growing up and living in a society with a potent heterosexual bias.

**sexual minorities** Lesbians, gay men, bisexuals, and transgendered persons.

**sexual prejudice** The negative attitude related to sexual orientation, especially toward homosexual behavior, people who have a homosexual or bisexual orientation, and the gay, lesbian, bisexual, and transgender communities.

**transgendered persons** Those whose gender identity, expression, or behavior is not the same as their birth sex; some experience their gender as incongruent with their

## 1. HETEROSEXIST BIAS IN WESTERN CULTURE

Intolerance of behavior that differs from stereotypes of masculine and feminine roles is frequent in some societies but relatively absent in others. Likewise, social prohibitions against certain forms of private, consensual sexual behavior are enforced in many, but not all, cultures. The United States continues to reflect a puritanical set of attitudes about gender roles and sexual behavior. These attitudes may be seen clearly in views of same-sex erotic or affectional attractions or relationships, which have been described by some commentators as a threat to traditional social or moral values and family structure.

In the United States, persons who are perceived to have a same-sex sexual orientation can be fired from a job or denied housing, credit, or service in a restaurant or hotel unless the state or local government has an explicit law prohibiting such discrimination. Members of the U.S. armed forces will be expelled, despite their service records or other qualifications, if they reveal their homosexuality—the “don’t ask, don’t tell” policy. In some states, foster parents who are gay or lesbian are not allowed to adopt their foster children—or any other children—due to their homosexuality. At least one private religious university has a policy that requires students and faculty to sign a statement attesting that they are not homosexual to be allowed to study for a degree or work at the school. Likewise, other religious and secular institutions, such as the Boy Scouts of America, discriminate against homosexuals. Moreover, persons thought to be homosexual are targeted for hate crimes, including violent attacks and harassment.

In the United States during the 1960s, a growing sentiment began to emerge that homosexuals were different from heterosexuals only with regard to the sex of the persons toward whom they felt attraction and erotic love. Psychological research by Evelyn Hooker on gay men and lesbians who did not seek psychiatric treatment found no difference in psychological adjustment, and her subsequent work with the Task Force on Homosexuality of the National Institute of Mental Health began to educate professionals about the gay and lesbian subculture. At the same time, the civil rights movement, protests against the war in Vietnam, and the women’s movement were challenging traditional views of social roles and institutions. Lesbians and gay men began to see the pathological view of homosexuality as something imposed on them by outside society rather than as an inherent condition of their inner selves.

Likewise, they were beginning to identify as a subculture that shared characteristics with other minority groups.

During the late 1960s, a heterosexual psychotherapist, George Weinberg, coined the term “homophobia” to refer to individual anti-gay behaviors and attitudes. Herek, in an article titled “The Psychology of Sexual Prejudice,” maintained that Jack Nichols and Lige Clark first used the term in print in their column on gay issues in the May 23, 1969, issue of *Screw Magazine*. One month later in New York City, a police raid on a gay bar called the Stonewall Inn ignited an extended protest against the harassment of gay people and is celebrated around the world as the touchstone of the modern gay, lesbian, bisexual, and transgender movement.

Homophobia is often criticized as being too clinical or pathological a term because it implies a pathological degree of fear (phobia). However, it might apply to a person who is afraid of being looked at or touched by a homosexual or who fears letting a child be in regular contact with a lesbian or gay male teacher. It could also involve hypervigilance or preoccupation with other people’s possible homosexuality. Likewise, it might apply to the fear expressed by officers in the U.S. armed forces that having homosexuals around would damage morale because the men and women would not want to shower or be in a foxhole with a homosexual. Such irrational fears and preoccupation may indicate a true phobia about homosexuals.

The term “heterosexual bias” is frequently a more accurate term to describe the belief that heterosexuality is normal and that everything associated with heterosexuality is superior to anything associated with homosexuality. Heterosexual bias often simply ignores anyone or any situation that is not heterosexual. For example, in psychological research on dating or attraction, participants often are not even asked whether they are heterosexual because they are presumed to be so.

## 2. DEFINITION OF SEXUAL PREJUDICE

Recently, the term “sexual prejudice” has come to be widely adopted in research and theory. Herek, in his seminal article on the topic, defined sexual prejudice as the negative attitude related to sexual orientation, especially toward homosexual behavior, people who have a homosexual or bisexual orientation, and the gay, lesbian, bisexual, and transgender communities. He noted that sexual prejudice, as with any prejudice,

has three components in that it (a) involves evaluation or judgment (an attitude), (b) is directed at a distinct social group, and (c) involves negative feelings (e.g., dislike, hostility). Several research instruments have been designed to measure homophobia as a prejudicial attitude, using questions regarding homosexuals or lesbians and gay men included in a group of questions about attitudes toward various groups and topics.

In research studies with large random representative population samples, sexual prejudice has been found to be associated with other dogmatic attitudes, rigid views of gender roles, conservative religious views, and authoritarianism. Herek found it to be more prevalent among those persons who are older, less educated, and living in rural areas and the southern and midwestern states in the United States. Sexual prejudice was also found to be more prevalent among persons with a fundamentalist religious orientation and among those who attend religious services frequently. In some studies, persons who believe that sexual orientation is the individual's "choice" tend to be less tolerant in their attitudes toward homosexuality than do those who believe that it is beyond the individual's control. Likewise, persons who have had personal contact with lesbians or gay men tend to hold more positive attitudes than do those who have had little or no contact of this kind.

Anti-gay attitudes are, in many settings, the only socially acceptable prejudice that can be openly expressed in the media and in public statements. Fear of homosexuals is often used by right-wing organizations to boost fund-raising and is sometimes used to stigmatize a political opponent.

Society likely pays a price for the effects of sexual prejudice, including the rigid enforcement of gender roles and stereotypes, the stigmatization of long-term same-sex relationships, and the secrecy, deceit, and hypocrisy forced on those in the armed forces and other careers. In contrast, reducing sexual prejudice can enrich society by providing new models of intimate relationships and gender roles for heterosexuals as well as for bisexuals and homosexuals. It may also remove the confusion often expressed between homosexuality and gender role stereotypes such as a lack of "masculinity" or "femininity" and homosexual sexual orientations.

### 3. INTERNALIZED HOMOPHOBIA

A particularly interesting focus of research is on internalized homophobia, which is the negative attitude that a

gay man, bisexual, or lesbian holds about himself or herself as a result of growing up and living in a society with a potent heterosexual bias. One line of research has focused on the effects of living as a sexual minority in society and the consequences that this minority stress may have on lesbian, gay, and bisexual individuals. To study this, Meyer in 1995 used an Internalized Homophobia Scale (IHS) that measures the extent to which gay men have discomfort about their sexual orientation and their efforts to avoid homosexual feelings. Two sample items included the following: "How often have you wished you weren't gay?" and "Have you thought that being gay was a personal shortcoming?" Internalized homophobia seems to be prevalent. Meyer found that 70% of a sample from the gay community expressed some internalized homophobic attitudes, although most had relatively low scores on the scale. In a study of respondents from the lesbian and gay community by Herek and colleagues, lesbians were found to have lower scores on the IHS than were gay men. For both groups, higher scores were associated with less disclosure to heterosexual friends, fewer connections with the lesbian and gay community, and higher scores on a measure of depressive symptoms and demoralization. Gay men with high scores also scored lower on a measure of self-esteem.

Internalized homophobia is a particular risk when young people first recognize their same-sex attractions. Typically, they have learned to fear and hate anything associated with homosexuality, so when they find that they might be a homosexual, they risk turning those negative attitudes in toward themselves by means such as increased incidence of depression, suicidal thoughts and attempts, and substance abuse. Lesbians, gay men, and bisexuals also are at risk for internalized homophobia if they have been victimized by hate crimes directed toward them due to their perceived sexual orientation, thinking that they have in some way deserved being victimized in this way. Although surveys have found relatively low levels of internalized homophobia in lesbians, gay men, and bisexuals over 60 years of age, it is thought that persons are at risk for internalized homophobia throughout their lives.

### 4. SOCIAL ATTITUDES, PHYSIOLOGICAL AROUSAL, AND HOMOPHOBIA

As with all attitudes, sexual prejudice is learned from peers, cultural media, and authority figures. It is likely

that pronounced homophobia has deeper roots than do other prejudices, however, given that it is so directly connected to gender stereotypes and sexual behavior, both of which are powerful themes in Western culture. For example, terms such as “faggot,” “sissy,” and “pussy,” when directed at boys who are not sufficiently competitive or aggressive, may be using homophobic imagery to maintain gender role conformity. The male role is often reinforced by homophobic imagery as it frequently becomes defined by an avoidance of anything “feminine,” including emotions that suggest vulnerability, tenderness, or uncertainty. Likewise, making an athletic woman who prefers short hair and is not interested in dating men to be the butt of jokes and sexual innuendos may be an example of homophobia as harassment. In both examples, homophobia is often used to reinforce gender stereotypes and to stigmatize those who deviate from the so-called traditional male and female roles. In a society that is becoming more gender irrelevant, this use of homophobia is actually an attempt to retard this change and may explain why it is stronger among both less educated and more conservative people.

Homophobia is also often linked with sexual behaviors and acts that are associated with homosexuals, especially male homosexuals. Those sexual acts, when the male is in the “passive” role, are highly stigmatized and are frequently described as ultimate humiliations of a male. Conversely, when performed by a female, even with another female, the same acts (e.g., oral sex, anal sex) may be regarded as erotic or even positively valued. Language is filled with expressions of fear regarding male anal exposure or penetration, so men try to “protect their asses” at all costs. Likewise, anything negative is said to “suck.” Although the term has now mostly lost its sexual connotation, its origins in the denigration associated with oral sex are obvious.

Psychoanalytic perspectives have long suggested that persons who are most virulent in their prejudice about some stigmatized group may be projecting their own inner conflict onto that group, although they are conscious of neither their own conflict nor its projection. For example, a man who feels same-sex sexual attraction, but who cannot accept that in himself and so represses the attraction, might project his own hatred of his same-sex attraction onto someone who he perceives to be homosexual. Expression of hatred toward the homosexual would have the added benefit of deflecting any suspicion of his own same-sex feelings by others who may perceive him as anti-gay. Lesbians, bisexuals, and gay men often must work with, or be

supervised by, persons such as this and are required to hide their pro-gay feelings.

One study provided empirical support for the idea that homophobic men may secretly be erotically excited by same-sex sexual acts. The experiment by researchers at the University of Georgia involved a device that measured change in the circumference of the penis and indicated the extent of erotic arousal as the men watched a series of sexually explicit videotapes. The men were also asked to rate their level of arousal by the tapes. Surprisingly, 54% of the men who scored high on a measure of homophobia before the study were definitely aroused while watching the same-sex male videos, according to the measure of penis erection, compared with 24% of the men who scored lower on homophobia. Despite their penile erections, the more homophobic men did not indicate a subjective erotic arousal by those tapes, rating their arousal as low as did the less homophobic men who showed no significant penile erections. Many of the men responded both subjectively and physiologically to videotapes of same-sex female sex acts and to male–female sex acts; these responses were unrelated to their scores on homophobia. It is possible, of course, that anxiety or other strong emotions—and not erotic stimulation—at the sight of male–male sex videos produced penile erections in the more homophobic men who did not experience subjective erotic arousal.

## **5. GENDER DIFFERENCES IN PREJUDICE: HOMOSEXUALS, BISEXUALS, AND TRANSGENDERED PERSONS**

In general surveys of attitudes within the United States, heterosexual men express more anti-gay attitudes than do heterosexual women. Heterosexual women do not express more negative attitudes toward gay men than they do toward lesbians, whereas heterosexual men typically hold much more negative attitudes toward gay men than they do toward lesbians. A similar pattern was found in attitude research among African Americans, where the men had especially negative attitudes with regard to gay men, believing that their behavior was “unnatural.” It is thought that men who violate the male gender role, which is a highly valued role in Western society, are stigmatized more than are women who violate the female gender role. Moreover, because female gender roles are less rigidly defined

than male gender roles, there may be less concern among men about lesbian women or among women about gender role violations in general.

Prejudice also exists toward bisexuals, that is, persons who feel erotic or affectional attraction to either sex. This prejudice has been termed “monosexism” to reflect the view that either homosexuality or heterosexuality is better than a combination of both orientations. Some homosexuals even believe that bisexuals should be encouraged to admit that they are actually gay, as if bisexuality were not a legitimate orientation. Herek found that attitudes toward bisexuals are more negative than those toward homosexuals. Women rated homosexuals more favorably than they rated bisexuals of either gender, and men rated males more negatively than they did females regardless of whether they were bisexual or homosexual. In general, bisexuals were rated lower on a “feeling thermometer” than were all other groups, except injecting drug users, in the survey questionnaire.

Little research has been done on attitudes toward transgendered persons, that is, those who identify themselves as neither male nor female, those who were previously male or female but have changed or are changing gender, those who have both male and female identities and wardrobes, or those whose bodies have characteristics of both sexes. A poll in 2002 found that a slim majority in the United States believed that it is “all right” to be transgender, but more than one-third believed that it is wrong. Those who believed that being transgender is a matter of the individual’s choice were more likely to see it as a moral issue than were those who believed that transgendered persons were born that way. In some cultures, such persons are regarded as “two-spirit” persons and are given roles in society that fit their unique gender.

## 6. EDUCATION, CIVIL RIGHTS, AND PRIMARY PREVENTION OF HATE CRIMES

Research has shown that personal contact with lesbian or gay persons, such as friends, family members, classmates, and colleagues, is associated with lower degrees of sexual prejudice. Therefore, disclosing one’s sexual orientation to associates can be effective education, especially if a friendship exists before or after the disclosure is made. Likewise, having homosexual peers speak in schools or social group meetings can be effective education. Finally, recognizing that most

families have one or more homosexual members often helps individuals to understand that sexual prejudice may affect everyone’s friendship and family circles.

Sexual minorities, including lesbians, gay men, bisexuals, and transgendered persons, are sometimes a protected class under civil rights laws. Harassment, bullying, threats, and verbal or physical violence are generally illegal and may be the grounds for civil suits. Because school districts have been held responsible for protecting sexual minority students from persecution, they are beginning to take homophobia seriously.

Hate crimes related to sexual prejudice, such as the murder of Matthew Shepard in Wyoming, have drawn attention to the most destructive effects of homophobia. Other forms of victimization related to sexual orientation can also have lasting effects and, in some states, are punished more severely than are non-hate crimes. Primary prevention efforts are needed to make communities safe for sexual minorities and to minimize the cost to society of the discrimination and prejudice that has long stigmatized lesbians, gay men, bisexuals, and transgendered persons.

### See Also the Following Articles

Homosexuality ■ Risk-Taking and Sexual Behavior ■ Sexual Behavior ■ Sexual Behavior and Culture

### Further Reading

- Adams, H. E., Wright, L. W., Jr., & Lohr, B. A. (1996). Is homophobia associated with homosexual arousal? *Journal of Abnormal Psychology, 105*, 440–445.
- Garnets, L. D., Herek, G. M., & Levy, B. (1990). Violence and victimization of lesbians and gay men: Mental health consequences. *Journal of Interpersonal Violence, 5*, 366–383.
- Garnets, L. D., & Kimmel, D. C. (Eds.). (2003). *Psychological perspectives on lesbian, gay, and bisexual experiences* (2nd ed.). New York: Columbia University Press.
- Greene, B. (2000). Beyond heterosexism and across the cultural divide: Developing an inclusive lesbian, gay, and bisexual psychology—A look to the future. In B. Greene, & G. L. Croom (Eds.), *Education, research, and practice in lesbian, gay, bisexual, and transgendered psychology* (pp. 1–45). Thousand Oaks, CA: Sage.
- Herek, G. M. (2000). The psychology of sexual prejudice. *Current Directions in Psychological Science, 9*(1), 19–22.
- Herek, G. M. (2002). Heterosexuals’ attitudes toward bisexual men and women in the United States. *Journal of Sex Research, 39*, 264–274.
- Herek, G. M., & Capitano, J. P. (1996). “Some of my best friends”: Intergroup contact, concealable stigma, and

- heterosexuals' attitudes toward gay men and lesbians. *Personality and Social Psychology Bulletin*, 22, 412–424.
- Herek, G. M., Kimmel, D. C., Amaro, H., & Melton, G. B. (1991). Avoiding heterosexist bias in psychological research. *American Psychologist*, 46, 957–963.
- Kite, M. E., & Whitley, B. E., Jr. (1998). Do heterosexual women and men differ in their attitudes toward homosexuality? A conceptual and methodological analysis. In G. M. Herek (Ed.), *Stigma and sexual orientation: Understanding prejudice against lesbians, gay men, and bisexuals* (pp. 39–61). Thousand Oaks, CA: Sage.
- Meyer, I. (1995). Minority stress and the mental health in gay men. *Journal of Health and Social Behavior*, 7, 9–25.
- Rothblum, E. D. (2000). "Somewhere in Des Moines or San Antonio": Historical perspectives on lesbian, gay, and bisexual mental health. In R. M. Perez, K. A. DeBord, & K. J. Bieschke (Eds.), *Handbook of counseling and psychotherapy with lesbian, gay, and bisexual clients* (pp. 57–79). Washington, DC: American Psychological Association.
- Savin-Williams, R. C. (1999). Matthew Shepard's death: A professional awakening. *Applied Developmental Science*, 3, 150–154.



# Homosexuality

Bertram J. Cohler and Phillip L. Hammack

University of Chicago, Chicago, Illinois, USA

1. Introduction
2. History, Culture, and the Problem of Defining Homosexuality
3. Understanding Same-Sex Desire: The Search for Biological Origins
4. Realization of a Gay or Lesbian Identity and the Life Course
5. Childhood and Adolescence
6. Homosexuality and Society
7. Conclusion  
Further Reading

## GLOSSARY

**coming out** Process of self-acknowledgment and disclosure to others, including friends and family, of attraction to, and desire for a sociosexual relationship with, another person of the same sex.

**gender-stereotyped interests** Childhood play and school and sports activities presumed to be appropriate for own gender; they are most often characterized as “rough-and-tumble” play for boys and as doll play and quiet activities for girls.

**sex role** Recognition and enactment of the socially constructed set of behaviors traditionally considered appropriate to one’s own gender.

**sexual identity** Construction founded on the confluence of social and historical circumstances and personal experiences of a sense at any point across the course of life of membership in that group of persons sometimes called homosexual, gay, lesbian, queer, or spectrum or in that group of persons sometimes called heterosexual or straight.

**sexual lifeway** Constellation of culturally defined expectations regarding the life course for persons self-identifying with a particular sexual orientation such as gay, bisexual, or lesbian.

**sexual orientation** Enduring awareness of sexual attraction to, or desire for, others of the same or opposite sex.

Homosexuality refers to both a socioerotic attraction to others of the same gender and an identity as a member of a community sharing this same-sex socioerotic attraction. Although it is a term preferred by men and women in earlier generation cohorts, younger cohorts view this as a medical term and one that implies opprobrium. Members of cohorts born during the postwar era and coming of age from the 1970s to the present prefer a term such as “gay” or “lesbians and gay men.” Members of the youngest cohort, born during the 1980s, often prefer a term such as “queer” or “spectrum,” referring to an other than normative (heterosexual) identity and beyond simple classification.

## 1. INTRODUCTION

The term “homosexuality” refers to the experience of exclusive or nearly exclusive erotic preference for others of the same sex in fantasy and, characteristically, through realization of sexual intimacy with others of the same sex. It can be conceptualized as desire, behavior, and identity (although it is not always congruent

within one person at any one time and so may represent a source of personal conflict). Homosexual desire is the affective experience of same-sex attraction. Homosexual behavior occurs when members of the same sex engage in sexual activity. Homosexual identity represents the assumption of a self-label in which one acknowledges the primacy of homosexual desire and behavior in his or her self-composition. In contrast to homosexual identity, homosexual orientation reflects the affective experience of primarily same-sex erotic desire beyond conscious control and outside the purview of historical and social construction.

The experience of same-sex sexual contact has occurred across cultures and historical eras. As such, it is difficult to generalize the meaning of homosexuality from one society or one time to another society or time. Even in American society, the meaning of homosexuality has undergone radical transformation and permutations during the past century. The current discussion is focused largely on the experience of men and women in contemporary Western society who self-identify as “homosexual” or, more recently, as “gay,” “lesbian,” or “queer.” This identity assumption in postmodern Western culture represents a particular sexual “lifeway” bound by culture and history. Thus, one’s conception is necessarily limited and rooted primarily in the experience of urban bourgeois gay men and lesbians in North America and Europe because these are the voices represented nearly exclusively in the literature on homosexuality. This article begins with a historical account of homosexuality in the current cultural context, followed by a discussion of the controversies regarding the origin of homosexuality. Next, it reviews theory and research on the gay and lesbian life course, beginning with the “coming out” process. Finally, the article considers four contemporary “problems” pertaining to gay men and lesbians related to the larger social and political ethos: mental health, social stigma and hate crimes, acquired immune deficiency syndrome (AIDS), and marriage.

## **2. HISTORY, CULTURE, AND THE PROBLEM OF DEFINING HOMOSEXUALITY**

In much of the public health and gay and lesbian psychological literature, the emergence of the phrase “men or women who have sex with other men or women” has emerged in response to the phenomenon of consistent homosexual behavior in the absence of

homosexual self-identification. For example, many young adults eschew any sexual identity as a “natural kind,” represented in polarizing labels such as “gay” and “straight.” The term “queer,” as used by these young adults, is meant to defy such essentializing efforts and reflects a new view of sexual desire as less amenable to simplistic bifurcations.

Recent research and theory have highlighted the importance of studying gay and lesbian lives within the context of shared, cohort-specific understandings of sexual identities and lifeways. Following a series of more gradual shifts in the social organization of sexuality during the first half of the 20th century, gay and lesbian life exploded into the public consciousness following the advent of the gay rights movement during the late 1960s and early 1970s. Inspired by and modeled on the civil rights movement, gay liberation was ignited in the aftermath of a 1969 riot in a New York City bar, the Stonewall Inn. Although the gay rights movement, symbolized by the Stonewall Inn riot, forever altered the shape of gay and lesbian life, its impact was largely cohort specific. Not only are different generations defined by different historical moments, but they also react in somewhat distinct ways to the same historical events. Over the subsequent three decades or so, there was a dramatic change in discourse regarding sexuality and sexual identity that has led to recognition of same-sex socioerotic identity as “virtually normal.”

One’s understanding of homosexuality today is highly influenced by the cultural and historical context of the late modern West. In particular, World War II created a context in which individuals with homosexual desires who perhaps felt alone in their experience were able to discover one another. Following the war, they settled in large metropolitan centers and began to form distinct gay and lesbian communities or subcultures. The emergence of a distinct homosexual identity as a clearly definable social category, as well as the growth of the gay and lesbian community, was fueled in large part by the social ethos of the 1960s and its culture of sexual liberation. The resistance of larger social institutions to accepting same-sex intimacy or homosexuality as a legal, legitimate social identity inevitably led to the Stonewall Inn riot, which sparked the beginning of the gay civil rights movement. Within the backdrop of this cultural and historical context, individual lives unfolded, and tantamount to social acceptance of gay men and lesbians has been the search for a causal mechanism—one that would ultimately demonstrate the immutability of homosexual desire.

### 3. UNDERSTANDING SAME-SEX DESIRE: THE SEARCH FOR BIOLOGICAL ORIGINS

Our culture is one in which lives, like stories, are assumed to have a linear organization with beginnings, middles, and ends—or childhood, adolescence, adulthood, and old age. Our society is one in which particular emphasis is placed on beginnings and origins. From learning theory (with an emphasis on primacy and recency of learning, ethology, and the concept of the critical period) and psychoanalysis (with a stress on the nature of the experience of early ties with caregivers), personality theory and developmental study has focused on origins and “causes” of particular personality characteristics. Nowhere is this concern with origins more evident than in the search for the causes of same-sex desire. This concern has been manifest both in the study of presumed biological determinants of homosexuality and in the discussion of the early development of the boy or girl later destined to choose same-sex sexual partners.

The debate in research and theory on homosexuality between the essentialist and constructionist positions is particularly concerned with the question of origins. The more politically controversial constructionist position emerges from social theories of identity as created through practice, suggesting that gay identity is formed within and among individuals as a consequence of shared cultural practice. The essentialist position, in contrast, posits that homosexuality is an innate condition of human experience. The biological line of research on homosexuality is representative of this position. Because of its controversial nature, including the implication of gay men and lesbians as “others” with some congenital defect (believed by some gay activists as fostering the cause of gay and lesbian political rights), the biological account of homosexuality is explored in depth in this section.

#### 3.1. The Four Areas of Biological Study of Homosexuality

The search for biological determinants of same-sex desire began during the 19th century and has continued to the current time with reports in scientific journals. Interest in this has been enhanced by those who believe that if a biological determinant can be found, same-sex desire might be viewed as an inherent and unchangeable attribute, much like hand preferences (right or left) or other differences, and then subject to legal protection.

Significantly, although there have been scattered references in the literature to the biological determinants of homosexuality among women, nearly the entire corpus of biological study focuses on the determinants of homosexuality among men.

The four areas in which biological contributions to the study of the origins of homosexuality have been reported are social biology and evolutionary theory, behavioral genetics and the search for the “gay” gene, hormones and prenatal development, and brain structure and anomalies from “normal” brain anatomy. Sociobiological models maintain that because the gene for homosexuality has not disappeared from the population, this gene must have some adaptive value. Claiming that gay men are more potent than their heterosexual counterparts, then, a heterosexual man who is heterozygous for the gay gene would have a reproductive advantage over a straight counterpart who is not carrying the gay gene. Other sociobiological theories suggest that parents recognize that some offspring are less biologically fit (e.g., lower intelligence, lower social adeptness) than others and so protect these children from challenging life experiences such as parenthood and “make” them into homosexuals. This perspective is virtually silent regarding homosexuality among women.

A second biological model focuses on behavioral genetics. Building on the observation that there is a “family disposition” toward homosexual interests, early reports contrasting monozygotic (MZ) and dizygotic (DZ) twins claimed that means of genetic transmission could be understood from a Mendelian perspective. More recent behavioral genetic studies have assumed polygenetic transmission. In 1991, Bailey and Pillard reported that the rate of concordance of homosexuality among MZ twin pairs is approximately 50%, whereas the rate among biological siblings is 22% and that among adopted brothers is 11%. This research group has also reported concordance rates for homosexuality among women who are MZ and DZ twins approximating those for homosexual men. After reports showing enhanced concordance for homosexuality among MZ twins as contrasted with DZ twins or brothers, Hamer and colleagues in 1993 reported a study of possible genetic linkage following the logic that if a trait is inherited through a single gene, relatives who share the trait will share the gene more often than would be expected by chance. Even if a gene that is responsible for a trait may not be identified, it may be shown to exist by means of transmission of another clearly inherited quality to which the trait is linked. Linkage studies are a common means for the study of inheritance. Following

this logic, Hamer and colleagues studied 40 families, each of which included two gay sons. Focusing on a gene located on the X chromosome that is transmitted from the mother but expressed only in men, the investigators located a marker on the Xq28 region that was concordant for 33 of the 40 pairs of gay brothers. In a large group of presumably straight brothers, the marker was randomly distributed. The chance of this association was less than 1 in 10,000. Hamer and colleagues maintained that this trait of homosexuality was transmitted through the maternal line, with the gene for homosexuality being located on the X chromosome, and confirmed this prediction in the reports by a group of gay men who attributed homosexuality to a greater number of maternal relatives than paternal ones.

A third area of biological study regarding the determinants of homosexuality focuses on the role of hormones, particularly during prenatal sexual development. This tradition presumes that same-sex sexual orientation, particularly among men, is a result of biosocial stress during critical periods of fetal development. Using animal models, prior observation had shown that in the absence of fetal androgens, the external genitalia develop as female, leading to the presumption that masculinization of brain centers also required the presence of these hormones. Particularly important for the development of homosexuality is the development of sex hormone receptors outside of the genital apparatus during the fourth through seventh months of gestation. Appropriate levels of testosterone appear to be necessary for these receptors to develop. The possible impact of disturbed intrauterine testosterone levels, such as might result from unusual stress or other environmental factors, has been a subject of continuing study. Aberrations in the hormonal receptor sites that might be caused by such changes would not be manifest until the time these receptors ordinarily begin to function. For example, males who had partially failed to develop testosterone receptor sites might not show the changes in muscle mass and strength that typically accompany pubertal development. Testosterone and its metabolite dihydrotestosterone (DHT) are necessary for male sexual arousal, so abnormal hormonal receptors might also influence sexual arousal. The impact of the failure to produce testosterone on the development of masculinity has been demonstrated in laboratory studies using rats.

This area of study has played an important role in work summarized in 2002 by Friedman and Downey regarding the emergence of homosexuality in men. Presuming androgen insufficiency that leads to

incomplete masculinization in boys destined to become homosexual, Friedman and Downey maintained both that parents implicitly recognize these boys' feminine nature and react to them as if they were less than masculine and that these boys avoid the "rough-and-tumble" play that is characteristic of typical boys. In fact, Friedman and Downey conflated gender-atypical interests (e.g., in theater or the arts) with femininity and assumed that these boys have some elements of a gender identity disorder that later becomes manifest as same-sex sexual orientation.

A fourth area of study regarding the biological determinants of sexual orientation, primarily homosexuality among men, was stimulated by reports of differences in the first through third interstitial nuclei of the anterior hypothalamus observed on contrasting groups of gay and presumably heterosexual men. These reports suggested that the hypothalamus, long believed to be important in determining sexuality, is sexually dimorphic. This study is closely related to that regarding intrauterine androgen exposure given evidence that the size of anterior nuclei of the hypothalamus in rats depends on prenatal androgens during a critical period of prenatal development, that this area is critical for rat sexual functioning, and that reported differences in the size of these nuclei in the human anterior hypothalamus are consistent with the prenatal stress-androgen hypothesis. This research presumes that gay men must be like women in preferring men as sexual partners because, for example, the third anterior nucleus of the hypothalamus has been shown to be larger in women than in men and so should be larger among gay men, who are like women in their preference for men as sexual partners, than among heterosexual men.

### **3.2. Problems Posed by Biological Study of the Origins of Homosexuality**

Much of the study attempting to show some biological determinants of homosexuality has reflected serious problems related both to the method of study and to the measures used to evaluate sexual orientation. For example, the social evolutionary perspective has been challenged on the ground that this perspective is mired in stereotypes and unwarranted assumptions such as that gay men are more potent and more successful in finding sexual partners than are heterosexual men. In 1999, Stein observed that in behavioral genetics study, there is an "ascertainment bias," possibly leading gay

men with identical twins both to be more likely to volunteer for such studies than gay men with fraternal twins and to overreport concordance for same-sex desire. Stein noted that concordance rates for sexual orientation are considerably lower than had been claimed when issues of volunteering were explored and that they have been inflated by sampling bias. Furthermore, whereas the genetic hypothesis predicts equal rates of homosexuality among DZ twins and brothers, Bailey's research group reported that unrelated adopted siblings were more likely than brothers to report same-gender sexual orientation (11 vs 5%). Finally, because twins were reared together in virtually all studies, it is difficult to disentangle environmental and biological effects. In sum, various problems of method, including the manner in which participants were located, reliance on indirect reports regarding sexual desire, focus on behavior rather than fantasy, measurement of zygosity by self-report, use of possibly unreliable scales for measuring sexual orientation, and very small sample sizes, all compromise the findings from this research.

Further problems are found in studies of the gay gene, including sampling methods, problems in replicating earlier studies, and problems of volunteer bias that are further compounded by the use of mail-in questionnaires and the reports of a variety of sexual behaviors among these volunteering participants. An approach to the study of the biology of homosexuality fails to consider the complex interplay of genetics, hormones, and social context and also blurs the distinction between association and causation. Reviewing the research in the area of hormones and sexual orientation, Stein concluded that there is little evidence showing even an association between expressed sexual identity and presumed hormonal factors. All of these problems are replicated anew in studies of neuroanatomy and sexual orientation.

Stein was particularly critical of the assumption made in much research on sexual orientation that this orientation is of a "natural kind" and that gay men can be readily classified as either "gay" or "straight." Stein systematically examined the assumptions underlying claims that sexual orientation is founded on some essential (biological) attributes. He observed that the dilemma inherent in understanding sexual orientation from this essentialist position is founded in the very presumption of categories (either binary or bipolar) of "natural kinds" of human sexuality, such as those of homosexual or heterosexual, man or woman, and sexual identity or gender identity, rather than recognizing

the many ways in which persons understand or make meaning of their sexuality and the limitations imposed by historically and culturally available constructs to name one's own experience. There may be little merit in trying to fit people into these binary categories. As Chodorow observed in 1992, it is not even understood why men and women become heterosexual.

The politics of gender and sexuality in contemporary society have further contributed to the publication of studies believed to show that sexual categories are a natural kind. The question is not whether biology is involved (biology is involved in all aspects of human development) but rather how biology is involved in particular kinds of sexual desires according to a permissive model in which genetic factors influence aspects of neuroanatomy, which then influence particular preferences; a direct model that constrains or shapes a particular preference; or an indirect model in which genetic factors influence personality, which then interacts with life circumstances to shape sexual orientation. Stein suggested that most findings to date support an indirect model of determination. Stein questioned the assumption that persons who have sex with others of the same gender uniquely possess attributes that need to be explained or that sexual orientation may be regarded as either binary (gay or heterosexual) or bipolar (a continuum marked at one end by gay and at the other end by heterosexual).

#### **4. REALIZATION OF A GAY OR LESBIAN IDENTITY AND THE LIFE COURSE**

The process of coming to self-identify as a homosexual (or, more recently, as a gay or queer) man or woman is often termed "coming out" (of the closet). In 1994, Chauncey observed that this term may be traced back to the 1930s when it was used as a satire for the debutante experience. However, during the postwar period, this term was further refined to mean coming out into the gay life or the gay world and out of hiding (portrayed by the often used term "closet"), including first discovery of same-gender sexual experiences and, more recently, disclosure of same-gender sexual desire to family and community. The terms "gay," "lesbian," and "queer" that are preferred by many younger urban-living men and women experiencing same-sex desire reflect an effort to turn a spoiled identity into a positive one.

Coming out represents more than just the disclosure of one's same-sex desire to others. It is a process of

internal experience in which an individual's entire sense of identity is challenged, reformulated, and crystallized. Either concurrent with or proceeding the process of coming out to self, individuals engage in a cultural resocialization process, usually by attending a coming out group at a community agency or gay social services center, that results in the eventual assumption of a gay, lesbian, or queer identity. Models of the coming out process have typically assumed a developmental character, best exemplified by the stage models, and have been challenged by perspectives that are more reflexive, discursive, and social constructionist. The historical and cultural relativity of the experience of those with same-sex erotic desire questions the value of stage-based developmental models of the coming out process. Stage-based models are concerned with ideal types and ignore the complexity of intersections with other aspects of identity, such as ethnicity, and so should be used only as a very basic guide to conceptualizing the identity development process for gay men and lesbians.

Adopting a symbolic interactionist perspective from sociology, humans behave in accordance with their own perceptions of meaning that are constructed through social interaction. The idea of homosexuality represents a socially constructed category or dimension of sexual experience that resonates with individuals constantly engaged in meaning-making (i.e., symbolic) human interaction. One becomes "homosexual" in this framework through the enactment of a particular role in social interaction. Being homosexual occurs in and through practice. It is not an "essence" located within a particular anatomical or genetic substrate. Rather, it is a meaningful classification of identity formulated, negotiated, and maintained through real human social activity. The social constructionist account of sexual identity has also been promulgated by a number of scholars working with lesbians. Much of this work reveals that, for White middle-aged lesbians and bisexual women of European descent, coming out is a nonlinear process that cannot be mapped onto specific stages.

The psychological and sociological approaches are nicely complemented by anthropological perspectives on gay and lesbian identity development such as that proposed in 1997 by Herdt, who posited that coming out represents a ritual that marks the beginning of a cultural resocialization process in which individuals with same-sex erotic desire learn the history, lore, and customs of a larger gay culture. In sum, coming out represents a process that occurs both within the individual and his or her relationship to others through social interaction and within the larger gay and lesbian

culture. As such, it is useful to consider the gay and lesbian identity development process from psychological, sociological, and anthropological perspectives given that they complement one another in their emphases on the individual, society, and culture.

## 5. CHILDHOOD AND ADOLESCENCE

Research on the childhoods of gay men and lesbians has been based largely on adolescent or adult retrospective reports. A theme that is particularly salient during the childhoods of gay men and lesbians is that of difference. As children, gay men and lesbians often report feelings of being different from peers, although they often do not define this difference in terms of sexual desire. Yet most gay men report very early awareness of same-sex attraction, typically during the elementary school years. Research on childhood experiences of homosexuality are obviously biased by this reliance on retrospective reports, but one can conclude that the affective response to same-sex desire begins during childhood.

Many investigations of lesbian, gay, bisexual, and transgendered (LGBT) adolescents have focused on issues relating to mental health and psychological adjustment. Although some studies of the mental health of gay and lesbian adolescents have suggested that, in contrast to their heterosexual counterparts, LGBT adolescents report greater levels of depression, hopelessness, and suicidal preoccupations, after controlling for environmental factors such as stress and social support, significant differences between the groups disappeared. These findings suggested that changes in resources for LGBT youth have the potential to diminish negative psychological outcomes. Studies have suggested that reactions of parents and perceptions of parental support are particularly important in predicting levels of self-esteem among LGBT youth and adolescents (as well as among LGBT individuals as whole) who are at increased risk for victimization and harassment (i.e., acts of verbal and physical violence).

Acts of victimization and harassment have been associated with a number of problems commonly experienced by LGBT youth, including school-related problems, suicide, running away, substance abuse, prostitution, and conflict with the law (usually for substance abuse, prostitution, running away, and/or truancy). When acknowledging the relationship between these problems and victimization and harassment, it is clear that LGBT youth need support systems. Organizations committed to aiding LGBT youth must also commit

themselves to breaking down the stereotypes and prejudice that result in victimization and harassment.

### 5.1. Adulthood and Aging

Little research has focused explicitly on the lives of lesbians and gay men in their 20s and 30s, due in part to the rapidly changing life course for gay men and lesbians. For example, with ages of coming out decreasing, gay men and lesbians are more closely mirroring their heterosexual counterparts in terms of romantic relationships. Once youth come out during adolescence, they are free to explore intimate relationships with same-sex partners by late adolescence. Early adulthood, then, is often characterized by serial monogamous relationships among gay men and lesbians, just as it is among their heterosexual counterparts.

Most discussions of adult development occur within a framework that tends to assume heterosexuality. Work is a major issue for lesbians and gay men at middle age and is also an area that differentiates them from their heterosexual counterparts. There tends to be more continuity in the professional lives of lesbians than in those of their heterosexual counterparts because lesbians' careers are less often interrupted by child rearing. However, Patterson noted in 2000 that this is likely changing as more gay men and lesbians are raising children. Most lesbians are in committed relationships and working during midlife. However, there is enormous diversity in experience during midlife among both lesbians and gay men, and little systematic research has addressed the question of midlife development among gay men.

At middle age in our society, people begin to look backward in evaluating their own lives rather than looking to the future. The "crisis" that is so often portrayed for men in our society as they reach middle age might be exacerbated for gay men as a result of their perceived loss of physical attractiveness. Sense of a foreshortened life course is common among the cohort of now middle-aged men who are survivors of the AIDS pandemic that peaked during the 1980s and early to mid-1990s and who suffered the loss of so many lovers and friends to AIDS. However, because the developmental trajectory of gay men across the adult years does not necessarily follow that of their heterosexual counterparts, it is difficult to generalize from heterosexual to gay experience of self and others across the second half of life.

The experience of later life (i.e., after 65 years of age) for lesbians and gay men is becoming increasingly investigated as the field of gerontology and aging expands

exponentially (in conjunction with increased longevity). As with gay men and lesbians at middle age, the lives of older gay men and lesbians cannot be extracted from historical context and the view of the larger culture toward homosexuality. It is possible that the unique experience of coming out (a major life transition in which one evaluates the self), together with lifelong experience in dealing with stigma and anti-gay prejudice, leads to crisis competence that fosters the development of personal strengths and, among older gay men and lesbians, further leads to successful adjustment to the challenges posed in our society for older adults. Research with older lesbians and gay men reveals highly successful coping mechanisms in the face of considerable stressors and the expected course of aging.

## 6. HOMOSEXUALITY AND SOCIETY

This final section reviews four contemporary issues involving homosexuality and society. First, the research on homosexuality and mental health is reviewed, detailing briefly the current controversy over "conversion" therapy for gay men and lesbians. Second, the impact of social stigma on personality development is reviewed, with specific attention given to the social problem of hate or bias crime. Third, the relationship between homosexuality and the AIDS pandemic is discussed. Finally, the issue of gay and lesbian marriage, and its implications for the changing history of the gay and lesbian life course, is discussed.

### 6.1. Mental Health

The question of the mental health of lesbians and gay men has been a major focus of inquiry since the elimination of homosexuality as a diagnosable mental illness in 1975. As members of a social group that has been marginalized in contemporary society, gay men and lesbians continue to be at risk for a host of mental health and adjustment difficulties, especially those associated with the coming out process. For example, mental health problems that adolescents experience concurrent with coming out are experienced intrapsychically but are rooted in external sources of stress (e.g., prejudice, victimization) and lack of social support. As attitudes toward homosexuality become increasingly positive, and as institutional support for young gays and lesbians grows (e.g., through gay-straight alliances in high schools), sources of external stress will likely diminish, improving the overall coming out experience. Mental

health problems that are typically experienced during the coming out process include low self-esteem, depression, anxiety, and suicidal ideation.

A major mental health concern for lesbians and gay men involves the experience initially portrayed by Malyon in 1982 as internalized homophobia. Internalized homophobia occurs when a gay man or lesbian possesses negative feelings of self based on the larger social phenomenon of homophobia or anti-gay prejudice. That is, the individual fails to reject the anti-gay attitudes of the dominant ideology and, instead, assumes the veracity of these attitudes. Research has suggested that gay men and lesbians with higher levels of internalized homophobia are more likely to engage in self-destructive behaviors, such as substance abuse, and to experience a whole host of other adjustment difficulties. The experience of internalized homophobia is likely implicated in a host of mental health and adjustment difficulties for gay men and lesbians. In sum, the mental health of the vast majority of gay men and lesbians is comparable to that of heterosexuals. The mental health problems most commonly experienced by gay and lesbian patients center on their struggles to adapt to a heterosexist society.

Although the mental health field has a history of pathologizing homosexuality, the ideology underlying more recent clinical training and practice has been largely affirmative of gay and lesbian identity. Advocates of so-called conversion or reparative therapy have claimed success in changing the behavioral sexual orientation of gay men expressing dissatisfaction with their gay identity, although they have not been able to show that these men have given up their fantasies about same-sex experiences. Drescher's careful review of this controversial therapy in 1998, as well as the report of the Group for the Advancement of Psychiatry in 2000, highlights the problems with the assumptions on which this therapy is founded and also the validity of claims regarding presumed successful outcomes of this conversion psychotherapy. Currently, most of the major professional mental health associations regard practice of reparative therapy as problematic or even unethical.

## 6.2. Social Stigma and Development

A second major issue involving homosexuality and contemporary society is the impact of social stigma on personality, identity, and overall mental health. During the recent history of Western society, attitudes toward lesbians and gay men have largely been characterized by heterosexism and homophobia. In 1993,

Herek portrayed heterosexism as an ideology that denigrates and stigmatizes nonheterosexual lifestyles and privileges heterosexuality. Prizing heterosex reflects a value assumption that it is the only appropriate manifestation of love and sexuality and devalues all that is not heterosexual. Herek reported that heterosexist and anti-gay attitudes appear to be more common among heterosexuals who are more religious, adhere to a more "traditional" ideology of family and gender, and have friends with anti-gay attitudes. Heterosexuals who report more contact with gay men and lesbians tend to endorse fewer anti-gay attitudes than do heterosexuals who have not had extensive contact with gay men and lesbians.

The large-scale stigmatization of lesbians and gay men exerts negative social, economic, and emotional effects on their development. At both the cultural and individual levels, heterosexism is implicated in the increased experience of victimization and anti-gay violence. Gay men and lesbians, especially youth, continue to experience victimization based on their sexual identity. The highly publicized, tragic murder of Matthew Shepard in Wyoming was a poignant reminder of the deleterious effects of cultural heterosexism and individual anti-gay attitudes. Heterosexism assumes a powerful role in both the internal (e.g., feelings of difference) and external (e.g., victimization) aspects of the lives of gay men and lesbians. The elimination of attitudes associated with heterosexism and homophobia has implications for enhancing the optimal development of lesbians and gay men by creating a context in which prejudice and the devaluing of a homosexual identity are minimized.

## 6.3. Homosexuality and the AIDS Pandemic

Too often, homosexuality among men has been conflated with being positive for the virus leading to outbreak of symptoms of AIDS, to such an extent that young men coming of age and experiencing same-gender sexual desire during the subsequent decades became terrified that their attraction meant certain infection and death. The virus responsible for AIDS was isolated and a test for presence of the antibodies responsible for the human immunodeficiency virus (HIV) infection became widely available by 1985. However, with no known treatment, there was some reluctance to get tested. Even after the development of the HIV antibodies test, as men began getting sick and dying, few understood

that unprotected anal sex, which facilitated blood–semen contact, was the primary means of transmission.

It is well known that AIDS is ever less a “gay” disease as many millions of persons around the world have become infected through needle use and other means of transmission but cannot afford the expensive medication available to affluent men and women in the United States. Much of the effort in the mental health field has focused on means of risk reduction through education promoting safer sex. Belief in the efficacy of safer sex leads to safer sex practices. Men who are more comfortable with their sexual orientation as gay, and who are out in the community with a lessened focus on anti-gay prejudice as a stress for them, are also more likely to practice safer sex and, in turn, receive a sense of enhanced personal mastery from such practices.

#### **6.4. “Normal” and Beyond: Gay and Lesbian Relationships, Marriage, and Parenthood**

The gay identity is often a contested one on the part of gay and lesbian people and the larger community alike. On the one hand, findings from recent surveys have shown a decrease in anti-gay bias. On the other hand, efforts within the gay and lesbian community to “fit in” within the larger society have been contested both by the larger society and within the gay community itself. Much of this controversy is apparent in the recent heated debate with the Episcopal church regarding the election of “out” gay priest Gene Robinson as a bishop of the church. However, this contested identity as gay or lesbian is most clearly evident in the area of same-sex partnership. Marriage in the United States symbolizes both social order and compliance with traditional norms. Even though divorce marks the end of one in three marriages, shared beliefs reflect the larger belief that marriage has a special meaning in terms of prevailing concepts of love and romance between husband and wife.

Given the significance of this belief in the special quality of the marital tie, there was wide-scale protest in 1993 when the Hawaii Supreme Court held that same-sex couples shared the right with heterosexual couples to marry under the equal protection clause of the state’s constitution. This decision prompted a firestorm as many states and the federal government passed legislation banning same-sex marriages. In 1996, Congress passed the Defense of Marriage Act, signed into law by President Bill Clinton, that denied legal recognition to same-sex unions at the federal level. The Catholic

Church and most mainline Protestant churches prohibit clergy from performing marriage rites with same-sex couples. At the same time, there is increasing support for legal recognition of same-sex civil unions beyond support of domestic partnership, as was recently approved in Vermont. It appears that the word “marriage” has some symbolic significance reflecting traditional norms ever more important when confronting the reality of social change.

In November 2003, the Massachusetts Supreme Court held that a prohibition on same-sex marriages violated the equal protection clause of the state’s constitution. The court upheld the equality of all persons and noted that marriage confers legal, financial, and social benefits, and also brings stability, to society. The court saw no constitutional reason why such benefits should be denied on the basis of sexual orientation. After the Vermont decision to legally recognize same-sex civil unions came a decision by Canada’s province of Ontario to recognize same-sex marriages (soon to be extended across the country), legal recognition of same-sex marriages in The Netherlands, a move for such legal recognition in Australia, and a move for legal protection for same-sex unions in Scandinavia, there is emerging support for legal recognition of gay marriage. In response to this trend, federal officials in the United States, from the president to a majority of Congress, continue to be opposed to same-sex marriages, and the Massachusetts decision has sparked renewed discussion regarding a U.S. constitutional amendment that would define marriage as a union between a man and a woman.

Posner, a federal appeals court judge and legal scholar, maintained in 1992 that government must not discriminate against classes of citizens. Although he appeared to be in favor of same-sex marriages, he indicated that society was not yet ready for this change and suggested granting legal protection to same-sex couples, such as in granted in Scandinavia, as an interim solution. Underlying these legal decisions there is continuing controversy even within the gay community regarding same-sex marriages. In 1997, these disparate positions were included in Sullivan’s edited collection that includes summaries of important court cases, congressional debate on the Defense of Marriage Act, and passages from the works of both supporters and critics of same-sex marriages. In 1995, Sullivan himself argued in favor of same-sex marriages, noting that extending legal protection would enhance the stability of same-sex relationships. He refuted the position of many opponents to same-sex marriage that

homosexual lifeways are necessarily depraved and the argument that remaining single is a better social outcome for gay men and lesbians than is entering into a legally recognized union that would enhance personal and social stability. In 1999, Warner indicated a concern that such legal recognition might undermine the distinctive culture created by gay men and lesbians and noted that the high divorce rate undermines the argument that heterosexual marriage fosters social stability.

The court in the Massachusetts decision explicitly commented that a ban on same-sex marriages proved to be a disadvantage for children of same-sex unions. Indeed, one area in which legal protection for gay and lesbian unions would be particularly important concerns parenthood and child care. One of the numerous areas in which the social changes of the past decade have been significant concerns birth, custody, and adoption among gay and lesbian couples. Custody decisions in divorce cases have generally favored the wife and mother, even when she has self-defined as lesbian and has sought divorce so as to join her lover. In 2000, a report of the Group for the Advancement of Psychiatry on the problem of bias and mental health among gay men and lesbians reported that several states have ruled that sexual orientation does not by itself imply unfitness of a parent in a contested custody proceeding and that parental unfitness and evidence of harm to the child must be demonstrated. In the past, courts have been reluctant to award custody to the father in divorce cases. Most often, where gay couples have been involved in parenting, it has been through joint custody or visitation rights. Other case law has maintained that sexual orientation should not interfere with visitation rights. Other courts have opposed custody for gay and lesbian parents on the ground of possible problems posed for the children at school and in the community by having two same-sex parents.

The three principal routes to lesbian or gay parenthood are where (a) one member of the couple brings one or more children from a heterosexual marriage, (b) the couple adopts, and (c) one partner (among lesbian couples) becomes pregnant through donor insemination. Evidence reported from a number of studies has not shown that children living with gay or lesbian couples face the problems in adjustment that has been presumed by some courts. Whether considered from the perspective of teachers, parents, or children, findings regarding the adjustment of children of gay or lesbian parents point to relatively few problems. Patterson's report suggested that children of lesbian mothers reported somewhat greater stress than did children of heterosexual mothers,

but they also reported an enhanced sense of well-being and comfort with their own lives. Patterson's review of studies in this area showed that children of gay or lesbian parents have the expected social ties during middle childhood and report few problems at school as a result of having two same-sex parents.

Children of gay fathers have somewhat greater difficulty in accepting their parents' sexual orientation than do children of lesbian mothers, although the nature of children's acceptance depends on their age at the time of parental disclosure. School-aged children have a somewhat easier time than do adolescents in making sense of parental sexual orientation for their own lives. Patterson observed that early adolescence may be a particularly difficult time for children to learn about their parents' alternative sexual orientation. Where problems with having gay parents have been reported, most notable among gay couples with children either adopted or brought from prior marriages, the extent of problems reported by the children is related to the degree of comfort that their same-sex parents have with their own role as gay men and fathers.

Child care and division of labor in doing routine household tasks appear to be more of a problem among gay couples than among lesbian couples. Too often in our society, child care has been considered the work of women and fathers have served as children's primary caregivers less often than have mothers. Gay fathers often try to compensate for any hardship their sexual orientation may pose for their children. They characterize themselves as both warm and nurturing as well as more likely than heterosexual fathers to set limits. Among both gay and lesbian couples, the two partners are more likely to be supportive and involved in child care than is the case among heterosexual couples, where child care is most often seen as primarily the responsibility of the wife and mother. Gay fathers, in particular, have had to work so hard to attain custody or to adopt that they are particularly likely to be involved in parenthood.

Gay fathers and lesbian mothers sometimes report anecdotally that they are concerned that their children should adopt gender-appropriate roles and engage in gender-stereotyped activities that conflate gender and sex roles but provide assurance that their offspring will not adopt a gay or lesbian identity as a reflection of their own still unresolved feelings about their gay or lesbian identity. In the past, gay fathers expressed greater concern than did lesbian mothers regarding the impact of their sexual orientation on their children's development. Although the offspring of gay and lesbian parents are not more likely than the

offspring of heterosexual parents to report a gay or lesbian sexual orientation, their parents are particularly concerned that their offspring adopt gender-stereotyped interests and not show evidence of a same-sex sexual orientation. This somewhat exaggerated concern reflects their awareness of social bias, their own experiences of growing up in the midst of a hostile climate at school and in the community, and their own internalized anti-gay prejudice.

## 7. CONCLUSION

Understood as same-gender sexual orientation and the consequent assumption of a sexual identity as gay or lesbian, homosexuality has assumed a changed significance in postwar society. Formerly a furtive identity that necessarily required “covering” in a conservative heterosexual society, urban-living lesbians and gay men (and, to some extent, even those living in smaller towns) have enjoyed increased social acceptance and diminished prejudice over the past two decades or so. In turn, this enhanced acceptance has led to greater ease among youth disclosing to friends and family their gay or lesbian sexual identity and has permitted gay and lesbian adults to enjoy their lives with friends, family, and children, comparable to the lives of their heterosexual counterparts. Much of what is known about the impact of alternative sexual identity on personal development across the course of life has been based on these generations growing up in the midst of prejudice. Challenges for the next period of study include long-term study of the adjustment of children of gay and lesbian parents as they reach adulthood and the impact of sexual orientation on the adjustment to aging as “baby-boomer” lesbians and gay men reach later life.

### See Also the Following Articles

Gender and Culture ■ Gender Role Development ■ Homophobia ■ Prejudice and Discrimination ■ Risk-Taking and Sexual Behavior ■ Sexual Behavior ■ Sexual Behavior and Culture ■ Stereotypes

### Further Reading

Bailey, J. M., & Pillard, R. (1991). A genetic study of male sexual orientation. *Archives of General Psychiatry*, 48, 1089–1096.  
Cabaj, R. P., & Stein, T. S. (1996). *Textbook of homosexuality and mental health*. Washington, DC: American Psychiatric Press.

Chauncey, G. (1994). *Gay New York: Gender, urban culture, and the making of the gay male world, 1890–1940*. New York: Basic Books.  
Chodorow, N. (1992). Heterosexuality as a compromise formation: Reflections on the psychoanalytic theory of sexual development. *Psychoanalysis and Contemporary Thought*, 15, 267–304.  
Cohler, B. (in press). *Writing desire: Generation and life-story among men who seek sex with other men*. Madison: University of Wisconsin Press.  
Cohler, B. J., & Galatzer-Levy, R. L. (2000). *The course of gay and lesbian lives: Social and psychoanalytic perspectives*. Chicago: University of Chicago Press.  
Drescher, J. (1998). *Psychoanalytic therapy and the gay man*. Hillsdale, NJ: Analytic Press.  
Friedman, R. C., & Downey, J. I. (2002). *Sexual orientation and psychoanalysis*. New York: Columbia University Press.  
Group for the Advancement of Psychiatry. (2000). *Homosexuality and the mental health professions: The impact of bias*. Hillsdale, NJ: Analytic Press.  
Hamer, D. H., Hu, S., Magnusson, V., Hu, N., & Pattatucci, A. (1993). A linkage between DNA markers on the X chromosome and male sexual orientation. *Science*, 261, 321–327.  
Herdt, G. H. (1997). *Same sex, different cultures*. Boulder, CO: Westview.  
Herek, G. M. (1993). The context of anti-gay violence: Notes on cultural and psychological heterosexism. In L. D. Garnets, & D. C. Kimmel (Eds.), *Psychological perspectives on lesbian, gay, and bisexual experiences* (pp. 89–108). New York: Columbia University Press.  
Malyon, A. (1982). Psychotherapeutic implications of internalized homophobia in gay men. *Journal of Homosexuality*, 72, 59–69.  
Murray, S. O. (2000). *Homosexualities*. Chicago: University of Chicago Press.  
Patterson, C. J. (2000). Family relationships of lesbians and gay men. *Journal of Marriage and the Family*, 62, 1052–1069.  
Plummer, K. (1995). *Telling sexual stories: Power, change, and social worlds*. New York: Routledge.  
Posner, R. (1992). *Sex and reason*. Cambridge, MA: Harvard University Press.  
Purcell, D. W. (1998). Current trends in same-sex marriage. In R. P. Cabaj, & D. W. Purcell (Eds.), *On the road to same-sex marriage* (pp. 29–40). San Francisco: Jossey-Bass.  
Savin-Williams, R. C. (2001). *Mom, dad, I'm gay: How families negotiate coming out*. Washington, DC: American Psychological Association.  
Stein, E. (1999). *The mismeasure of desire: The science, theory, and ethics of sexual orientation*. New York: Oxford University Press.  
Sullivan, A. (Ed.). (1997). *Same-sex marriage: Pro and con*. New York: Vintage Books.  
Sullivan, A. (1998). *Love undetectable: Notes on friendship, sex, and survival*. New York: Knopf.  
Warner, M. C. (1999). *The trouble with normal: Sex, politics, and the ethics of queer life*. New York: Free Press.



# Human Rights

Willem Doise

University of Geneva, Geneva, Switzerland

1. Institutional Definition of Human Rights
  2. The Ambiguous Message of Social Psychology
  3. Human Rights as Normative Social Representations
  4. Research Findings
  5. Putting Things in Perspective
- Further Reading

## GLOSSARY

**collective rights** Particular rights based on membership in cultural, religious, linguistic, or national groups. The main function of such rights is to maintain and further links between members of a community as well as to preserve their lifestyle.

**contextualization of rights** Processes that actualize a more or less broadening or restricting definition of rights that are deemed to be relevant in a given situation.

**human rights** Universal rights that all individuals should enjoy regardless of their belonging to different social groups or categories. Individuals are invested with these rights by the mere fact of being a member of the human species. The Universal Declaration of Human Rights, voted by the General Assembly of the United Nations in 1948, is the official document stating these rights.

**normative social representations** Representations that specifically allow humans to evaluate their relations and interactions. All social representations imply a normative aspect because they are beliefs embedded in systems of norms.

**social representations** Organizing principles of symbolic relationships between individuals and groups. They imply common understanding of social reality, organization of

individual positioning toward issues at stake in society, and anchoring of individual variations of positioning in social realities.

Human rights are defined as normative social representations embedded in institutional juridical definitions. Conclusions from psychological studies on matters of rights enforcement are ambiguous. Research findings show that human rights can be studied as normative social representations implying a degree of common understanding across cultures together with organized differences within and between cultures. Important factors in modulating individual positioning in the realm of human rights are experiences of social conflict and injustice, beliefs about the efficiency of various social actors to have rights enforced, and attitudes of liberalism or collectivism.

## 1. INSTITUTIONAL DEFINITION OF HUMAN RIGHTS

The Universal Declaration of Human Rights, adopted on December 10, 1948, by the General Assembly of the United Nations Organization, serves as the main institutional reference for defining human rights (HR). The declaration comprises six groups of articles. The first group of articles (Nos. 1 and 2) enunciates the basic principles of equality, freedom, and dignity; the second group (Nos. 3–11) focuses on rights of the individual,

such as security of the person and equal protection by the law; the third (Nos. 12–17) concerns rights relative to interindividual relations, among which there are freedom of movement and the right to found a family; the fourth (Nos. 18–21) involves civil rights, such as freedom of expression or equal access to public service; the fifth group (Nos. 22–27) deals with economic and social rights, such as social security and the right to rest and leisure; and the sixth (Nos. 28–30) covers societal rights relative to international law and order as well as to duties to the community.

When the declaration was drafted, it was not at all obvious that eventually all groups of articles would be retained by the vote of the general assembly. Fierce controversies opposed members favoring exclusively a declaration of individual judicial rights together with public and civil rights and members more concerned about providing individuals with socioeconomic rights, such as rights to a decent standard of living, education, and health care. This antagonism is still acute and is reflected in the contents of the two additional covenants of 1966 that have defined a more legal status for rights defined in the declaration, one covenant bearing on civil and political rights and the other on economic, social, and cultural rights. These resulted in procedures aimed at guaranteeing respect of specific rights, at least among the countries that signed them. Not all the countries that signed the universal declaration signed both additional pacts. Hence, the universal declaration remains the most universal and comprehensive reference in the realm of human rights.

Half a century after its adoption, it is more than evident that enforcement of the rights stipulated in the declaration is far from being guaranteed in many countries throughout the world. Reasons for this lack of HR respect are numerous and of diverse nature: political, economic, cultural, institutional, and also psychological.

## **2. THE AMBIGUOUS MESSAGE OF SOCIAL PSYCHOLOGY**

Psychological dispositions without doubt influence the way individuals take into consideration the rights of others. In order to study such influences, one could borrow a lot of useful suggestions from studies on prosocial behavior. However, in the present context, human rights enforcement is considered, above all, a matter of institutions warranting rights of individuals. The functioning of institutions involves individuals

in that they participate in specific relationships. Therefore, I privilege here social psychological analyses of such functioning rather than analyses in terms of individual differences.

Lessons of social psychology in relation to human rights are often lessons in skepticism. In research performed by Stanley Milgram in 1974, participants went ahead in delivering what they thought to be dangerous electric shocks to a person refusing to be further shocked or who was already unconscious. In a series of critical replications published by Meeus and Raaijmakers in 1987, it was shown again that for the sake of an experiment on stress, many people prevented an unemployed man from being hired. Of course, unknown to the experimental subjects in these experiments, the unemployed man was an actor.

Another lesson of skepticism is that in order to uphold a basic belief in a just world, many people often denigrate and chastise other people for the simple reason that they suffer. According to Melvin Lerner in his monograph published in 1980, a just world belief, seemingly very widespread in Western societies, would imply that remuneration and punishment, positive sanctions, and negative sanctions are not distributed randomly in this world. If one suffers, there should be a reason, and therefore people who suffer are often denigrated and considered responsible of their own fate. Literature on victimization, as summarized by Janoff-Bulman and Hanson Frieze in 1983, is based on a similar hypothesis. This trend of research can be illustrated by a concrete example: Witnesses of a rape scene, and even the victim of such an aggression, often tend to blame the victim for having an important part of responsibility in the aggression.

Lerner puts forward the idea that the fundamental need for justice could easily turn to discriminatory practices. In order to satisfy one's own needs as well as those of one's next of kin in an appropriate manner, it would be necessary to keep one's distance from a large part of humanity that cannot satisfy its hunger and that lives in inequality and injustice. To interact with such people outside strictly regulated or even discriminatory terms is looked upon as risky behavior endangering the progress of justice at home. It would be precisely those people who subscribe most to the concept of a just world who would best accept institutionalized patterns of discrimination.

Note, however, that the principle of a just world can also be put forward to explain altruistic behavior. It would thus account for feelings of indignation felt when witnessing some blatant injustices and could

even occasionally be at the base of “heroic” attempts made to redress these torts.

Research on submission to authority as well as that on the perverse effects of a belief in a just world are examples of the various trends of research that justify an attitude of skepticism about the often optimistic vision maintained by human rights supporters. Another important point to be added to this rather grim picture is the diffusion of responsibility described by Latane and Darley in 1970. Their studies aim to explain why individuals often do not intervene when confronted with an emergency, such as when somebody is attacked by another individual or falls in a state of unconsciousness on the street. A robust finding is that the presence of other bystanders reduces to an important extent the willingness of individuals to intervene in such a situation. Apparently, people feel less concerned to intervene in a situation of emergency when others can also intervene. Theoretically, one can suppose that this is often the case in situations of human rights violations and that therefore a reaction of “it is not my responsibility to intervene” is very likely to be manifested. Taking the rationale further, individuals confronted with massive and institutionalized HR violations may think that their own role in redressing these violations is utterly insignificant. Fortunately, when broadening the realm of research covered by psychologists in order to include research on the social development of children and adolescents, indications favoring more optimistic conclusions start to emerge.

What characterizes modern thinking is the refusal that traditional bodies such as churches, authorities, or charismatic personalities decide what is right or wrong, what is good or bad. It is this refusal that necessitates the building of a constructivist concept of knowledge and ethics. This conception of modernity oriented the groundbreaking research by Jean Piaget published in 1932 on *The Moral Judgment of the Child*. His aim was to study the transition of morals among children based essentially on the approval of rules decreed by others to more autonomous morals allowing for contractual changes of these rules through discussion and agreement between parties involved. According to Piaget, participation of children in interaction with equals makes such a transition from heteronomous to autonomous morality possible.

To consider that this developmental hypothesis by Piaget has been empirically authenticated would be exaggerated. Nevertheless, it embodies a general concept that has been at the root of other investigations too numerous to list. Only a brief presentation of the

stages of moral development described by Lawrence Kohlberg in 1984 is provided. Indeed, Kohlberg has refined, to a large extent, Piaget’s distinction between the two kinds of morality. He describes no less than six stages in the moral judgment of individuals. It is important to briefly characterize these stages, especially because the final two offer foundations for a constructivist concept of human rights ideas.

The first four stages are not directly related to human rights. The first one is essentially governed by constraints issuing from rules and sanctions: The avoidance of punishment and respect of authority are considered as ends in themselves. During the next stage, an individualistic orientation still predominates, but it already includes an instrumental concept of trading and reciprocity. The respect of the interest of others is used to better guarantee one’s own interests. The third stage represents conformism and seeking of social recognition. Attitudes such as to be well intentioned, to respect others and to keep one’s promises, and to remain faithful in allegiances give rise to morals illustrated by the statement “Do not do unto others what you would not have others do unto you.” The fourth stage is based on the principle of duty—the need to follow laws, rules, and conventions in order to keep the social system functioning well. A majority of adults often seem to reason following principles of the third and fourth stages.

The fifth stage sees the explicit intervention of a relativist attitude by acknowledging that many individual values and opinions are characteristic of given groups or societies and that other social contracts could set a higher value on different attitudes. The problem of the universality of specific values such as freedom or respect for life is explicitly put forward and a great significance is assigned to the consensual definition of a procedure allowing the fixing of rights and values to be respected by all.

The sixth and last stage is hardly encountered elsewhere than in writings of specialists such as philosophers and human rights militants. Universal ethical principles should supersede rules and conventions when they are in conflict. Following Immanuel Kant, people are considered an end in themselves. The moral point of view is the one that all human beings should adopt toward each other as free and autonomous subjects. Reasoning characteristic of this stage should take into consideration the point of view of each person susceptible of being affected by consequences of a decision to be made; such reasoning is based on the hypothesis of reversibility, according to which everyone should be able to put themselves in the others’ positions.

This is not the occasion to start a technical debate on the theoretical value of Kohlberg's stage model in the frame of current developmental psychology. It would be equally wrong to pretend that a person would necessarily fit into only one stage, as well as to pretend that their order of succession in individual development would occur at random. For present purposes, the important point is that an idealized conception of human relationships has also been the object of psychological studies.

In *Minding the Law*, A. G. Amsterdam, an expert in U.S. constitutional law, and Jerome Bruner, whose contributions to the psychology of child development, cognition, education, language, and culture are well known, analyze a series of decisions by the U.S. Supreme Court related to racial discrimination and the equal protection clause of the Fourteenth Amendment. Studying relevant documents and opinions of the Supreme Court in terms of category processes, narratives, rhetorics (conceived as social influence devices), and cultural dynamics, the authors show how important variations occur over time. The cases they comment on more extensively deal with the rendition of a runaway slave (*Prigg v. Pennsylvania*, 1842), racial segregation in railroad transports (*Plessy v. Ferguson*, 1896), racial desegregation in public schools (*Brown v. Board of Education*, 1954), the death sentence of an African American (*McCleskey v. Kemp*, 1987), and restrictions on desegregation in public schools (*Freeman v. Pitts*, 1992, and *Missouri v. Jenkins*, 1995).

The main conclusion of the study on historical variations over more than a century and half in the decisions by the Supreme Court is framed by Amsterdam and Bruner in the following terms: "This trail can be seen as marking the fall, the rise, and the fall again of a brightly egalitarian and progressive American Creed—or as marking the rise, the fall, and the rise again of an equally potent, darkly suspicious American Caution" (p. 16). Indeed, as others before them, the authors conclude that a basic component of American culture is the principles proclaimed in the Declaration of Independence, but they also conclude that the evolution of this culture cannot be understood if one does not admit the intervention of an equally potent principle restraining and limiting these rights for long periods. In discussing the links between these two fundamental aspects of American culture, they echo ideas of Lerner: "So the preservation of the American Creed requires that We (sic) create a category of people who are essentially and immutably different and inadmissible" (p. 264).

Hence, taking into consideration various research traditions, there seems to be a double message: Psychologists deliver lessons of optimism as well as

lessons of pessimism concerning matters related to human rights. According to Piaget and Kohlberg, the ideal of human rights would be, in a certain way, a spontaneous product of democratic interaction patterns. According to Milgram, citizens of democratic states can easily be brought to violate fundamental rights of others, and according to Lerner the proper belief in a just world may be used to justify practices of social discrimination as well as attempts to restore justice. In their contribution to the study of U.S. law and culture, Amsterdam and Bruner conclude that adherence to an optimistic egalitarian credo in a liberal democracy can be accompanied by an equally strong belief that rights of certain groups of people should be severely limited.

At first sight, these messages carry insurmountable contradictions that should resist integration within a same theoretical framework. However, studying human rights as normative social representations can allow for a better understanding of these seemingly conflicting lessons and offer a possibility of integrating them in a common explanatory frame.

### 3. HUMAN RIGHTS AS NORMATIVE SOCIAL REPRESENTATIONS

Multiple relations interconnect members of the human species directly or indirectly. Between the different groups constituting mankind circulate not only genes, viruses, pollutants, epidemics, consumer goods, or money but also ideas and institutions. What inhabitants of one country practice affects to varying degrees the fate of people abroad, their way of farming land, their state of health, their access to all kinds of resources, and the functioning of their institutions. It is difficult to imagine that a group of humans would not be affected to some extent by the way of life of other groups. In this sense, apartheid does not exist, but on the other hand, there is no perception, no definition, nor exhaustive analysis of the numerous ties that bind humans together. Even if globalization is a new term for naming that general interdependence, it does not mean that the phenomenon can be described precisely and fully understood. It is also true that general terms, such as globalization, can be used to conceal the obvious reality of exploitation or of negative interdependence. Such realities, even though they are part of a larger relational field, exercise their own noxious effects in many ways.

Interaction and communication between humans initiate symbolic representations, social norms, and contractual principles that often remain implicit. When we enter into a relationship with other persons, we know that the fate of participants in such a relation will be affected by their interaction, in some measure, within certain limits, at a certain cost. Normative representations exist on what these effects should be. Multiple forms of interdependence exist, characterized by all sorts of differences in status, purpose, interdependency, and formality. Various models of acceptable relationships, prototypes of fair and just situations, and contractual principles govern evaluations of these relationships. Such evaluative criteria have been built in multiple interactions, they are culturally defined, and their application can be guaranteed by institutions. Some of them are culturally specific, whereas others are spread over different cultures.

Human rights are such evaluative principles or normative social representations that could allow, at least at the level of the intention, humans to evaluate and to organize their relations and interactions. For historical (i.e., economical, political, military, religious, and scientific) reasons, Western societies organized relationships within national and cultural boundaries as well as across them. Many encounters with foreign cultures were sources of continuous conflict, and much irremediable harm has been inflicted in the frame of these contacts. Nevertheless, revolutionary ideas about universal human rights that were initially formulated as reactions against the arbitrariness of established powers (at the occasion of the American Declaration of Independence in 1776 and the French Declaration of the Rights of Man and Citizen in 1789) were thereafter progressively invested with new meanings. These meanings were elaborated by diverse and sometimes antagonistic militant movements in the realms of antislavery, antiserfdom, antiracism, and anticolonialism causes; by movements defending rights of workers, women, cultural minorities, aboriginal peoples, wounded soldiers, displaced persons; and by movements furthering right claims of many other social categories. The visions of human rights elaborated by these movements, and often forged at the expense of much effort and struggle, became institutionalized in the Universal Declaration of 1948, which is to be considered according to its preamble

As a common standard of achievement for all peoples and all nations, to the end that every individual and every organ of society, keeping this Declaration constantly in mind, shall strive by teaching and education to promote respect for these rights and freedoms and by progressive measures, national and international

to secure their universal and effective recognition and observance, both among the peoples of Member States themselves and among the peoples of the territories under their jurisdiction.

Human rights are clearly normative social representations, even if they are also important political and judicial realities.

It is generally accepted that in interactions with others, independently from their origin, the principles adopted in the Universal Declaration compel the respect of corporal integrity of these others, their liberty and their dignity, their access to vital resources, and their integration in a societal order guaranteeing protection against arbitrary decisions. Logically, in order to respect commitments taken by public authorities, citizens of many countries should only involve themselves in relationships respecting these principles. However, it is a fact that in numerous situations basic rights of individuals are not respected, as evidenced by economic conditions imposed on Third World countries or when some people, even in First World countries, are deprived from basic health care or from subsistence-level resources.

Here, a huge field of research opens. If human rights are indeed normative social representations, social psychologists—at least those involved in social representation studies—have theoretical and methodological tools at their disposal to study them.

#### 4. RESEARCH FINDINGS

Social representations (SRs) are defined as organizing principles of symbolic relationships between individuals and groups. A first assumption on which this definition is grounded is that various members of a population under study share common views about a given social issue. SRs are generated in systems of communication that necessitate common frames of reference for individuals and groups participating in these communications. Hence, studying human rights in SRs terms always implies a search for common reference systems and for their organization. To what extent does the Universal Declaration offer references that are common to different populations under study?

Mainly, three studies are discussed here because they deal with SRs of HR in a transcultural setting. Two of them are extensively described in the 2002 monograph by Doise.

The first study derived from an interdisciplinary interview study on representations of violations. In a

questionnaire study, students aged 13–20 years in four countries (Costa Rica, France, Italy, and Switzerland) were invited to answer 21 items presenting various situations involving violations or limitations of individual rights. Some of these situations (e.g., racial discrimination, imprisonment without trial or legal assistance, and starvation) can easily be referred to classical definitions contained in the Universal Declaration. Other situations, dealing with the rights of children or with family affairs, are less explicitly related to the declaration. Some situations dealing with economic inequality or health matters (e.g., prohibition of smoking or hospitalization in case of contagious illness) are not covered by official definitions of rights.

The results were clear: Across countries generally the same situations are more often considered to be human rights violations. These findings have since been confirmed by research in the United States and Romania.

In a second transcultural study, the complete text of the declaration was presented to students of 35 countries. Respondents (university students in psychology, law, science, social work, and various other fields from the five continents) were asked to answer questions about personal involvement, agreement and efficacy, as well as governmental efficacy for each of the 30 articles of the declaration. A hierarchical cluster analysis designed to accentuate intraclass homogeneity was performed on indices of the Euclidean distances between articles based on the sums of responses to the eight scales for each article. This analysis resulted in the division of the articles into two main classes, which in turn were divided into subclasses. These subclasses showed an almost complete correspondence with the groups of articles described at the beginning of this article.

A subclass of the first main class aggregates the more social articles, relational rights (group 3, except article 14 about asylum), public liberties (group 4), and articles dealing with economic, social, and cultural rights (group 5). The other subclass aggregates the three articles referring to the more basic individual rights (protection from torture and slavery and right to life; articles 3–5 from group 2).

The second main class can be subdivided into three classes: a subclass of judicial individual rights (remaining articles of group 2) together with the right to asylum (article 14), basic principles (group 1), and the three articles concerning societal order (group 6).

In each national group, respondents show greater adherence to the basic and social rights than to the rest of the rights. These results clearly support the idea of a common organization of responses in various

countries. Such a common organization was also observed by Spini in a follow-up study using other samples of respondents from 16 countries belonging to five different sociogeographical areas.

A third study was carried out by Herrera and Doise in the province of Quebec, Canada. Participants in this study were French-speaking and English-speaking Quebecois and members of the aboriginal Innu nation (Montagnais). Unlike the previous studies, the questionnaire used in this study started with two different open-ended questions asking respondents to mention the human rights they know and also the specific rights claims of respectively the Quebecois, the Quebecers, and the autochthones according to the ethnic group membership of the three samples under study. For instance, for the Innu sample, this question was worded as follows: “According to you, which are the five most important rights of autochthonous people that would do them justice?”

Analysis of answers for the three groups shows that most of the rights that were mentioned as examples of HR could indeed be related to the Universal Declaration of Human Rights. There was also a high convergence between evocations of HR by the three groups: The four groups of rights most often referred to by the three samples were public freedoms and political rights, basic human rights principles, fundamental individual rights, and economic social and cultural rights.

Such a convergence was not observed for the evocations of collective rights (i.e., rights of the Quebecois or of the Aboriginal people also investigated in this study). Even if almost all rights mentioned as collective rights could be related to both the Universal Declaration of Human Rights and the Draft Declaration of Rights of Indigenous People, the rights more often mentioned by respondents of the three groups were not the same. The members of the Innu sample mentioned rights to dispose of their own territory and resources, fundamental collective rights such as the right to autonomy, rights to the preservation of their traditions and language, and rights to self-government and specific indigenous laws more often than other categories of rights. Of the rights contained in the Draft Declaration of Rights of Indigenous Peoples, only the rights concerning language and tradition figured among the rights most often quoted by both other samples.

Findings of this study again showed to what extent ideas about human rights are shared across group boundaries, although collective rights are interpreted in a more diverse way.

Social representation theory does not imply that individuals sharing common references hold the same positions. It is therefore important to investigate differences in individual positioning. Modulators of positioning are beliefs about personal efficacy and the efficacy of institutions (e.g., governments) with respect to human rights. Human rights positions also bear a relationship with value choices. Values as measured by Rokeach's terminal values are considered general to the extent that they supposedly organize symbolic relationships within a social environment. Given the historical origins of human rights ideas in conditions of discrimination and conflict, it is also expected that HR representations will be modulated by individual experiences of injustice and of conflict between social groups and categories.

A typology was constructed of respondents in the study with the 30 articles of the declaration. Four groups were retained: A group of advocates and a group of skeptics had relatively the highest and lowest scores on scales concerning general importance of HR and personal and governmental involvement in enforcing rights. A group of personalists considered that HR concerned them personally but that it was not easy for governments to do something for their enforcement, whereas governmentals considered themselves rather powerless for having HR respected but considered the government as relatively more effective. Strong support for the values of universalism and social harmony was found to be systematically related to HR advocacy. Intense experience of collective injustice, together with less concern for personal happiness, was characteristic of personalists rather than of governmentals. It was also found that skeptics were relatively more numerous in Japan and India, and personalists were more often found in countries with serious human rights problems according to the 1992 assessment published by C. Humana and with human development problems according to the 1996 ratings by the United Nations Development Programme. Governmentals were relatively more numerous in developed countries or countries that changed recently to a democratic regime.

Individual positioning in the study on violations of HR was also investigated. For the respondents in the initial four-country study and the Romanian sample, a cluster analysis extracted three groups of individuals according to the similarity of their responses. The ranking of violations for the two main clusters clearly overlapped the ranking found for the total population. However, these two clusters could be differentiated mainly on situations related to family affairs. Obliging children to forego schooling or to attend mass, parental

child beating, and preventing one's spouse (wife or husband) from going out alone were less clearly perceived as human rights violations by a group of respondents sharing a restricted definition of human rights, as were the obligation for women to veil their faces and the refusal to rent an apartment to a black person or a gypsy family. Respondents adhering to a more extended definition considered all these cases to be violations of HR. On the other hand, responses given by members of a third cluster were clearly atypical, probably reflecting a relative disinterest for the HR cause, at least in the manner they were institutionally defined.

How were individual variations in social positioning toward human rights violations linked to positioning in other societal realms? In order to investigate these relations, five additive scores were computed for each individual on a series of anchoring variables, namely the number of governmental actions considered unacceptable (e.g., death penalty and suppression of elections or of television broadcasting critical of government), the number of queries made by government and business firms considered to be unfair (e.g., concerning health and militancy in unions), the number of official regulations (e.g., regarding vaccination and compulsory military service) interpreted as infringement of individual freedom, the number of individual actions reported as unacceptable (e.g., bribery, hiding an escaped convict, and giving in to blackmail), and the number of agreements with fatalistic assertions (e.g., there will always be rich and poor, good and bad people, and nations that do not respect HR).

These anchoring variables were then used as independent variables in a discriminant function analysis with the three clusters of the typology as dependent variables. The two functions were statistically significant. The first function depicts the specific anchoring of the cluster of adherents to a restricted definition. In comparison to members of the other two clusters, individuals adhering to a conception that adheres more closely to the institutional definition of human rights cultivate a relatively more fatalistic conception of social reality while expressing broader tolerance toward governmental interference and various social regulations. One can clearly identify the origin of a restricted definition of rights: increasing power given to the government and acceptance of various social regulations in the context of a fatalistic conception of social relations. The second function merges an extended definition with refusal of various governmental and managerial infringements, with a higher tolerance for questionable individual deeds and a better

adhesion to various social regulations. In other words, this conception focused on protecting the individual—even while occupying a marginal social position—against authorities while accepting their regulatory role. The location of the atypical or unconventional position on each of these two functions reveals the complete opposition of its anchoring to those described previously.

In the Quebec study by Herrera and Doise, some HR were mentioned as collective rights and vice versa. A factor analysis shows that answers to both questions are interrelated in various ways, but that individual scores on the factors are again related to the way respondents perceive the actual and possible contribution of different social actors in having human and collective rights respected as well as to their experience of discrimination and rights violations.

A nonhierarchical cluster analysis classifies respondents into three groups. The position of members of a first group can be called one of liberalism. Tenants of this position are observed to adhere above all to the basic principles of human rights and to public freedom rights. The position shared by members of a second group corresponds to the cause of sovereignty that privileges mainly cultural and linguistic rights, but this position also implies an awareness about self-government issues and problems related to socioeconomic rights compatible with HR principles. These two profiles were more characteristic of nonaboriginal French-speaking as well as English-speaking participants who seemed equally aware that in their socio-cultural context important issues in defining HR and collective rights are those of liberalism and sovereignty. Self-determination, especially in matters concerning the use of land and resources and self-government, characterizes the position of a third group of respondents. As could be expected, the latter profile was characteristic of the aboriginal sample.

The differences of the SR approach and other social psychological approaches in HR studies are difficult to assess because few psychologists have researched HR. In a 1995 study, Diaz-Veizades *et al.* shared some of the concerns mentioned in this article because they also used the Universal Declaration of Human Rights in order to study the organization of individual positioning as well as anchoring of this positioning. However, they undertook important adaptations of the original document, reformulating the 30 articles of the declaration into 116 more concrete items, such as “If a person does not make enough money to support his or her family adequately, the family should be aided by the

government,” “There are times when people should be kept from expressing their opinion,” “Men and women should have equal rights in a marriage,” and “A person’s home is his or her ‘castle’ and should not be interfered with by others.”

For each item, their respondents (mainly North American college students) expressed their degree of agreement on a 7-point scale. However, the authors report that many items “had very low variance because of high rates of endorsement, so they could not correlate highly with other items. After these low-loading items were deleted, 38 items were reanalyzed using an iterated principle factor analysis.” Clearly, we are confronted here with a logic different from the one adopted in this article. Diaz-Veizades and colleagues favored the study of systematic interindividual variations to the detriment of analyzing the common meaning aspect of SRs. When eliminating the most consensual items of their questionnaire, they could retain four factors for the remaining items that represented only one-third of all items presented to respondents. A first factor, called the social security factor, concerned access or entitlement to an adequate standard of living (e.g., food, housing, and medical care). A second factor, civilian constraint, dealt with the acceptability of limiting individual civil and political rights. The theme tying items together for the third factor was that of equality, evidenced most clearly by items dealing with equal access to basic rights for all individuals regardless of race, gender, or beliefs. Finally, items with highest loading on the fourth factor involved privacy issues. Diaz-Veizades and colleagues also analyzed anchoring patterns in showing, for instance, that respondents adhering more to a civilian constraint conception of HR were also those who obtained higher scores on a nationalism scale and lower ones on internationalism and civil liberties scales; their political preferences were more likely to favor Republicans over Democrats as opposed to the preferences of adherents to a social security conception, who favored Democrats and had higher scores on internationalism.

## 5. PUTTING THINGS IN PERSPECTIVE

An important issue that preoccupies colleagues from anthropology and cultural psychology is the problem of the universality of HR, and some of them envisage the possibility that HR are merely an expression of

Western ethnocentrism. Indeed, ethnocentric use of HR was experimentally evidenced by Moghaddam and Vuksanovic in 1990. In a first study, they asked undergraduate students in Montreal to answer a questionnaire about HR issues. The questionnaire is identical in three conditions, except for the fact that in a first condition the context referred to was Canada, in a second condition it was the Soviet Union, and in a third condition it was the Third World. Typical items were as follows: "All forms of censorship should be done away with in [Canada/Soviet Union/Third World] societies" and "Everyone in [Canada/Soviet Union/Third World] societies should have access to free health care." Overall, support for HR was stronger in the Soviet Union and Third World contexts than in the Canadian context. In a second study, three scenarios were used concerning a television news anchor woman who had been fired from her job, a member of a terrorist group who had been tortured, and a female shoplifter, again presented in three different contexts. As in the first study, support for HR was stronger in the conditions Soviet Union and Third World than in the condition Canada. Such results tend to confirm the opinion of critics who consider HR as a Western export article, considered useful for others but less so for Westerners themselves.

The previous discussion suggests that the HR issue is indeed a highly normative one. Individual stances are embedded in value choices and orient judgments about social reality also in the realm of international relations. In one's own environment, strong convictions about values are not necessarily applied in a consistent way. Clémence and Staerklé explored this issue and their findings are reported in the monograph by Doise. When comparing judgments on rights limitations reported in the realm of privacy, free education, or the death penalty, they found that Genevan participants in their study very strongly adhered to the relevant HR principles, giving average scores higher than 6 on a 7-point scale. However, scores dropped significantly when it was a matter of indicating their degree of acceptance (1, acceptable; 7, unacceptable) of violations of these rights in cases of a search without a warrant of a home by a police officer, refusal of free primary schooling to handicapped children, or the death penalty for minors. Furthermore, when these limitations of individual rights were merely reported as newspaper items ("faits divers"), their degree of rejection was on average 4.94, whereas when they were explicitly reported as extracts of HR reports, the average degree of unacceptability decreased by almost

1 point on the scale to 4.01. It seems as if references to HR for events in Western countries were less pertinent and prevented people from going into a more detailed analysis of the situation. Paradoxically, the absence of an explicit invocation of HR implies that individuals pay greater attention to reprehensible aspects of acts occurring in their own environment and, as a consequence, find them less acceptable. On the contrary, when general HR principles are explicitly evoked, concrete characteristics of events happening in one's own country are not taken into account, and in this way people can moderate their judgments.

The fact that HR are to be considered normative social representations invests them with very strong value connotations. As a consequence, in some situations violations of rights are considered less important when they happen in one's country, whereas HR violations far abroad are strongly condemned.

Without doubt, variations in enforcement of HR exist across countries. However, differences in beliefs related to the actual enforcement of HR in one's own country also exist between inhabitants of Western democratic countries. The principal component of a normative social representation will certainly remain its insufficient realization in many social contexts. Otherwise, it would no longer have reason to exist. It is precisely the awareness that rights are violated that often makes people claim that these rights should be enforced—that their enforcement has to become part of social reality.

### Further Reading

- Aboriginal and Torres Strait Islander Commission (1996). *United Nations Draft Declaration on the Rights of Indigenous Peoples*. Canberra, Australia.
- Amsterdam, A. G., & Bruner, J. (2000). *Minding the law*. London: Harvard University Press.
- Bierhoff, H. W. (2002). *Prosocial behaviour*. New York: Taylor & Francis.
- Black, C. L., Jr. (1997). *A new birth of freedom. Human rights named and unnamed*. New York: Grosset/Putnam.
- Diaz-Veizades, J., Widaman, K. F., Little, T. D., & Gibbs, K. W. (1995). The measurement and structure of human rights attitudes. *Journal of Social Psychology*, 135, 313–328.
- Doise, W. (2002). *Human rights as social representations*. London: Routledge.
- Doise, W., Clémence, A., & Lorenzi-Cioldi, F. (1993). *The quantitative analysis of social representations*. Hemel, Hempstead: Harvester Wheatsheaf.
- Finkel, N., & Moghaddam, F. (2004). *The psychology of rights and duties: Empirical contributions and normative commentaries*. Washington, DC: APA Books.

- Herrera, M., & Doise, W. (2004). Social representations of human and collective rights in three sociocultural groups of Quebec. *New Review of Social Psychology*, 3.
- Humana, C. (1992). *World human rights guide* (3rd ed.). New York: Oxford University Press.
- Janoff-Bulman, R., & Hanson Frieze, I. (1983). A theoretical perspective for understanding reactions to victimization. *Journal of Social Issues*, 39, 1–17.
- Kohlberg, L. (1984). *The psychology of moral development*. San Francisco: Harper & Row.
- Lauren, P. G. (1998). *The evolution of international human rights. Visions seen*. Philadelphia: University of Pennsylvania Press.
- Lerner, M. J. (1980). *The belief in a just world. A fundamental delusion*. New York: Plenum.
- Meeus, W. H. J., & Raaijmakers, Q. A. W. (1987). Administrative obedience as a social phenomenon. *Current Issues in European Social Psychology*, 2, 183–230.
- Moghaddam, F., & Vuksanovic, V. (1990). Attitudes and behavior towards human rights across different contexts: The role of right-wing authoritarianism, political ideology and religiosity. *International Journal of Psychology*, 25, 455–474.
- Renteln, A. D. (1990). *International human rights: Universalism versus relativism*. London: Sage.
- Shweder, R., Minow, M., & Markus, H. R. (2002). *Engaging cultural differences*. New York: Russell Sage Foundation.
- United Nations Development Programme. (1996). *Human development report*. New York: Author.



# Humor and Well-Being

James Rotton

Florida International University, University Park, Florida, USA

1. Multiple Meanings of Humor
  2. Mechanisms: Possible Mediators and Moderators
  3. Methodological Issues
  4. Trait Humor, Distress, and Longevity
  5. Humor and Physical Health
  6. Conclusions
- Further Reading

## GLOSSARY

**humor** A broad and multifaceted concept that is variously used to refer to the quality of communications, the emotions that such communications elicit, and a tendency to respond with amusement in a variety of situations.

**state humor** Emotions (e.g., amusement, mirth) that link stimulus materials (e.g., comedy tapes) and observable responses (e.g., laughter).

**trait humor** A readiness to find or create humor in different situations (i.e., sense of humor).

**well-being** Physical and emotional health.

Articles on humor and health frequently begin by reminding readers, for example, that “a merry heart doeth good like a medicine” (from Proverb 17:22 in the King James Bible). Although few would argue with the first part of this proverb—“A merry heart doeth good”—there is very little agreement about whether or not humor should be likened to medicine. On the one hand, it is easy to come away from movies such as *Patch Adams* and articles in *Reader’s Digest* with the

impression that “laughter is the best medicine” (a phrase that has been used in the titles of several scientific articles). On the other hand, scientists doing research on health and well-being have expressed a considerable amount of skepticism about the medicinal effects of humor. Probably no article has done more to stimulate interest in humor and well-being than one that appeared in the *New England Journal of Medicine* by Norman Cousins, who was then the well-connected editor of the *Saturday Review*. In this brief article, which was subsequently expanded into a best-selling book, Cousins described how he checked himself out of a hospital after being diagnosed with a progressive, and usually fatal, form of rheumatoid disease. Following a self-prescribed regimen of Marx Brothers movies and *Candid Camera* television episodes, Cousins concluded that “10 minutes of genuine belly laughter had an anesthetic effect and would give me at least two hours of pain-free sleep.”

## 1. MULTIPLE MEANINGS OF HUMOR

Humor is a broad and multifaceted concept that is variously used to refer to the quality of communications (e.g., funny or amusing stories), the emotions that such communications elicit (e.g., amusement, mirth), and a tendency to respond with amusement in a variety of situations (i.e., a sense of humor). As a stimulus, humor has been used to describe the

contents of literary material, television shows, movies, and (more recently) videotapes that are supposed to be funny. However, more than one study has reported checks on manipulations indicating that the individuals exposed to supposedly humorous material do not share the experimenters' definitions of "funny" or "amusing." This finding highlights the importance of individual differences. Few things are as idiosyncratic as people's preferences for different types of humor. If everybody agreed about what is funny or humorous, one would never have to endure the agony of seeing a comic "die on stage."

Serious problems are also encountered when one tries to define humor as a response or behavior such as laughing or smiling. No single behavior is a sure sign of amusement or mirth. For example, a subordinate may smile at the boss's joke for reasons that have nothing to do with how funny the joke is. To take another example, few movies provoke as much laughter in a college classroom as do Milgram's *Obedience to Authority*, which shows individuals forced to make stressful decisions. In such situations, laughter communicates anxiety rather than amusement. Furthermore, Provine's observational studies indicate that laughter in natural settings is typically preceded by mundane and pedestrian remarks. Provine could have used the familiar form of his first name, Bob, to illustrate this point. It was not long ago that the mundane greeting "Hi, Bob" (said by every person in a room) provoked laughter, first on *The Bob Newhart Show* and later whenever an individual named Robert entered a room.

Obviously, stimulus-based definitions of humor require that we observe some kind of response, and response-based definitions presume that evoking stimuli possess certain characteristics. These facts have led most theorists and researchers to define humor as an intervening variable. Following the lead of investigators doing research on emotions such as anxiety and anger, scholars frequently draw a distinction between state humor and trait humor. State humor describes emotions (e.g., amusement, mirth) that link stimulus materials (e.g., comedy tapes) and observable responses (e.g., laughter). Trait humor is a readiness to find or create humor in different situations. The latter corresponds to what is usually termed a person's "sense of humor." Investigators doing research on trait humor also draw a distinction between "getting a joke" (i.e., appreciative humor) and "telling a joke" (i.e., productive humor). In 1998, Craik and Ware identified several other types of humor. For example, the Humor Style Questionnaire highlights the well-known distinction between aggressive humor and friendly

humor. Although the questionnaire's originators have been concerned primarily with people's health, research on aggression highlights the importance of drawing such a distinction. It has consistently been found that friendly humor reduces aggression, whereas witticisms (i.e., jokes at the expense of others) and hostile humor usually provoke hostility and/or aggression.

## 2. MECHANISMS: POSSIBLE MEDIATORS AND MODERATORS

At a time when individuals are constantly reminded that good health requires that they engage in activities they would rather avoid, such as exercising and eschewing sweet and fatty foods, it is not surprising that the public has seized on the idea that humor is good "for whatever may ail a person." Those who favor humor might not state things so bluntly (e.g., "Here's something you will like"), but they make a point of noting that there is a fairly strong connection between negative emotions and illness. It has been argued that if things that people do not like cause stress, which is associated with illness, then things that people like should add to health and longevity. Of course, the same argument could be made for sugar, fatty foods, and addictive substances.

Positive affect is usually assessed by having individuals complete self-report measures, although investigators occasionally obtain nonverbal measures such as frequency of laughter and smiling. These responses have been treated as separate dependent variables in research on humor and health. However, the case for mediation would be stronger if investigators performed analyses showing that relationships between humor and well-being vanished or were reduced after investigators statistically controlled for the intervening variable of positive affect. Unfortunately, such analyses have not been performed. In addition, there is no evidence to support the frequently heard claim that endorphins are released when individuals are exposed to humorous material.

Early research tried to identify cardiovascular and endocrinological changes that might be responsible for humor's effects. The physiological effects of laughter are similar to those of brief bouts of exercise—faster breathing, rapid heart rate, elevated skin temperature, and increased blood pressure. Higher levels of circulating catecholamines and cortisol have also been recorded. Early researchers suggested that people benefit from

the relaxation that follows laughter. However, these physiological changes have not been shown to be intervening variables that lead to long-term improvements in health and well-being.

Physiological explanations can be contrasted with cognitive approaches that emphasize appraisal and distraction. It is possible that humor enables individuals to cope with stress by helping them to view events as less threatening (i.e., cognitive appraisal) and by putting their problems into perspective (i.e., cognitive reappraisal). The latter is typified by jokes that begin, "You think you have problems." Another possibility is that the task of processing humor (i.e., getting a joke) distracts attention away from unpleasant events and experiences such as pain.

Finally, some of the effects attributed to humor may be due to the demonstrable effects of social support. In 2000, Provine pointed out that individuals rarely laugh when they are alone. Unfortunately, social support has been introduced as a confound by the common practice of testing individuals in small groups. Research is needed to determine whether humor exerts salubrious effects when individuals are tested alone.

### 3. METHODOLOGICAL ISSUES

The distinction between state humor and trait humor is relevant to the procedures that investigators employ to assess the effects of humor. By and large, researchers interested in the effects of state humor have adopted an experimental approach, whereas those conducting studies of trait humor have employed correlational techniques. Typically, in experiments on state humor, behavioral and physiological measures are collected after some participants have been exposed to an audio or video recording of a comic (nearly always Robin Williams or Bill Cosby). This has been done in both field and laboratory settings. It is difficult to interpret the results from laboratory studies when the dependent variable is pain tolerance because it is likely that most participants (especially undergraduate males) guess that the experimenters want to find out whether they can "grin and bear it." Regardless of whether studies are conducted in laboratory or field settings, researchers must take care to assess and control for their expectations and demand characteristics. So-called placebo effects are a problem because it is hard to disguise manipulations of humor. Because so much has been written about humor's supposedly

beneficial effects, it is likely that most participants guess that the experimenter expects them to respond in a positive manner. This suggests that verbal ratings of stimulus materials, which are the usual checks on manipulation, might not be particularly useful or informative. As Provine pointed out, it is essential that researchers verify their inductions by monitoring behaviors such as smiling and laughter. Another necessity is to include a control group whose members are exposed to equally interesting and arousing but nonhumorous material. This type of control is needed to ensure that effects attributed to humor are not in fact caused by arousal or by distracting attention away from pain and unpleasant events.

In correlational investigations, participants typically complete several measures of humor and emotional states such as anxiety and depression. Two of the most widely used measures in this area are the Situational Humor Response Questionnaire (SHRQ) and the Coping Humor Scale (CHS). Lefcourt and Martin developed these to measure productive humor, but it is not entirely clear what these scales actually measure. Psychometric investigations indicate that these and other humor scales load heavily on a factor that Köhler and Ruch termed cheerfulness. The scales also tap a factor that has been termed "low seriousness" (or frivolity) that appears to accompany the detachment and perspective taking that humor is supposed to promote. Köhler and Ruch also found that correlations between self-report and behavioral measures of humor were low. These findings suggest that results obtained in this area might be due to other personality traits such as optimism, creativity, extraversion, and lack of social sensitivity. This is another way of saying that studies on trait humor lack discriminant validity. In addition, nearly all studies have been cross-sectional. Very few investigators have undertaken longitudinal investigations, which would strengthen claims that humor and well-being are linked in a causal fashion.

Although the distinction between trait humor and state humor is useful for organizing the literature, a few studies have combined the two approaches by measuring individuals' sense of humor before they are exposed to humorous material. It was thought that individuals with a well-developed sense of humor might benefit more from humorous viewing matter than would those with a less developed sense of humor. This hypothesis has not received support, but it is possible that people who have a sense of humor (or a cheerful disposition) go out of their way to expose themselves to material and situations that make them smile and laugh.

#### **4. TRAIT HUMOR, DISTRESS, AND LONGEVITY**

As might be expected, measures of trait humor are correlated with scales that assess positive emotions such as happiness and subjective well-being. Conversely, they are negatively correlated with measures of distress (e.g., anxiety, depression, hostility). However, this pattern may reflect little more than the operation of a more general trait such as cheerfulness. Thus, it is noteworthy that early studies by Lefcourt and Martin found that trait humor may act as a buffer that reduces the negative effects of stressful life events. These investigators concluded that humor is “an antidote to stress” after they had undergraduates complete three sets of self-report measures, including the previously mentioned measures of humor (SHRQ or CHS) and a measure of stressful life events. They used scores on these two measures to predict scores on a third measure (total mood disturbance), which was the sum of four negative mood states (e.g., anxiety, depression) and a negative score for positive mood. They found that some, but not all, types of humor acted as a stress-reducing buffer. In particular, people who scored high on productive humor (i.e., attempts at being funny) were less likely to report negative mood states when they encountered stressful life events than were those who scored low on productive humor. However, this finding received mixed support in subsequent studies. Overall, productive humor does not appear to be any more effective than humor appreciation in helping people to cope with stress. In addition, the effects of trait humor appear to depend on how stress is defined (i.e., stronger effects when adversity is recalled than when it is experienced), on how distress is operationalized (i.e., stronger effects for depression than for anxiety), and on gender (i.e., stronger effects for males than for females). These inconsistencies led Martin and Lefcourt to use words such as “ambiguous” and “uncertain” to describe the buffering effects of trait humor.

Suggestive and sometimes intriguing findings have also been obtained in studies that have examined relationships between trait humor and physical symptoms. Unfortunately, all of the findings are correlational and based on self-report measures. Because investigators have not controlled for negative affectivity, it is possible that people with a sense of humor report fewer symptoms because they are the kind of people who are less likely to appraise events in a negative fashion.

Nonsignificant and sometimes contrary results have been obtained when investigators have tried to relate trait humor to more objective measures such as longevity. One study found that famous humorists and comics, who earned their reputations by producing humor, did not live any longer than individuals who achieved eminence in other fields. Granted, this study can be faulted for failing to draw a distinction between friendly humorists (e.g., Fred Allen, Will Rogers) and comics whose witticisms and insults are associated with negative emotions (e.g., Don Rickles). This criticism does not apply to a longitudinal analysis of mortality rates of gifted individuals who were first tested at 12 years of age as part of Terman’s Life Cycle Study. Friedman and colleagues combined teacher ratings of optimism and sense of humor into a measure of cheerfulness. Contrary to expectations, they found that individuals with a cheerful disposition died at earlier ages than did their more serious counterparts. This unexpected result was explained by suggesting that a cheerful disposition might lead individuals to ignore or discount threats to their health.

#### **5. HUMOR AND PHYSICAL HEALTH**

In addition to longevity, investigators have looked at the effects of humor on blood pressure, immune function, and pain tolerance. It appears that humor does not have long-term effects on cardiovascular measures such as heart rate and blood pressure. However, there is some reason to believe that trait humor acts as a buffer to reduce the systolic blood pressure of females but not that of males. It may be that females use humor to reappraise threatening events (e.g., “Look on the bright side”), whereas males show a preference for aggressive humor.

Mixed results have been obtained in studies that have looked at the effects of humor on the body’s immune system. On the one hand, a few positive results have been obtained in studies that relied on blood assays, but immunological indicators vary from one study to the next—sometimes natural killer (NK) cell activity, sometimes interleukin cell activity, sometimes both, but more often neither. It is difficult to interpret these findings because investigators have not controlled for general emotional arousal and the pyramiding of Type I errors that occurs when a large number of tests of significance are conducted. On the other hand, several studies have found that humor leads to higher levels of secretory immunological A (SIgA), which is the body’s first line

of defense against upper respiratory illnesses. However, nonsignificant results were obtained when investigators looked at the frequency of cold and flu symptoms. Of greater concern, studies on SIgA have lacked adequate controls for general emotional arousal, distraction, and diurnal variations in hormone production and immune function.

Finally, although research on most health variables has been disappointing, it is in some ways fitting that studies on pain provide partial vindication for the claims that Cousins originally made in 1976. Laboratory studies are fairly consistent in finding that individuals have a higher tolerance for pain after watching an amusing videotape than they do after watching one that is merely interesting or distracting. There is also some evidence that humorous materials help individuals to endure orthopedic and arthritic pain in hospital settings. For example, in a field experiment in 1996, Rotton and Shats compared the amount of pain medication that patients requested after watching either comedies or adventure movies following orthopedic surgery. They found that patients who watched comedies requested smaller doses of minor analgesics (e.g., aspirin) than did those who watched adventure movies or who were not given movies to watch. However, this beneficial effect did not extend to major analgesics (e.g., Percodan). Indeed, the results indicated that patients actually asked for more medication when they were not given a choice as to which films they could watch. As the researchers observed, having no choice but to watch comedies that fail in their attempts to be amusing probably elicits negative emotions.

## 6. CONCLUSIONS

Belief in the healing power of humor is so widespread that this article's skeptical tone might seem surprising to some. Those who question humor's medicinal effects run the risk of being labeled curmudgeons. Thus, it should be noted that this article's evaluation echoes conclusions found in books and articles by the most active and frequently cited researchers in this field. It should be cautioned that these conclusions are based on a relatively small number of empirical studies. For example, Martin reported that his recent review of health effects uncovered no more than 45 empirical studies.

It is possible to offer two reasons why very little research has been done to untangle relationships between humor and well-being. First, belief in the healing

power of humor is so widespread that it already seems to be taken as proven. That is, why do research to test a hypothesis that most people already believe is true? Second, one may recall that the concept of humor appears to load onto two factors: cheerfulness and lack of seriousness. Most people do not object to being termed "cheerful," but they find it difficult to have their work dismissed as lacking in seriousness. These two tendencies may explain why publications in this area rarely include footnotes that acknowledge funding sources.

Reviews usually address a paucity of research by suggesting that more is necessary. However, what is really needed is better research—experiments that contain necessary control groups and verifiable measures of humor (e.g., laughter, smiling), studies that rule out general arousal as a rival hypothesis, intervention studies that avoid all of the biases (e.g., expectancies, placebo effects) that self-report measures introduce, investigations whose results are based on a sufficient number of participants, and research that explores the possible mediating role of social support.

Rather than comparing humor to medicine, as has been done so frequently, it might be more appropriate to compare the effects of humor to those that may be produced by vitamin pills and the supplements one finds at health food stores. The government prohibits sellers from claiming that dietary supplements cure anything, but those who buy vitamins believe that they benefit from ingesting them. In the same way, the literature in this area contains intriguing results from unfortunately flawed studies suggesting that individuals may sometimes benefit from a dose of laughter. Stores can sell vitamins because they do not contradict the physician's edict: "First do no harm." The same can be said of humor. There is probably no harm in a regimen that keeps people in stitches—unless, of course, an individual is recovering from an appendectomy.

### See Also the Following Articles

Emotion ■ Emotion and Culture ■ Social Comparison and Subjective Well-Being ■ Stress ■ Traits ■ Well-Being

### Further Reading

- Cousins, N. (1976). Anatomy of an illness (as perceived by the patient). *New England Journal of Medicine*, 293, 1458–1463.
- Craik, K. H., & Ware, A. P. (1998). Humor and personality in everyday life. In W. Ruch (Ed.), *The sense of humor* (pp. 63–94). New York: Guyton de Gruyter.

- Köhler, G., & Ruch, W. (1998). Sources of variance in current sense of humor inventories: How much substance, how much method variance? *Humor*, 9, 305–321.
- Lefcourt, H. M. (2002). Humor. In C. R. Snyder, & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 619–631). New York: Oxford University Press.
- Martin, L. R., Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Criqui, M. H., & Schwartz, J. E. (2002). A life course perspective on childhood cheerfulness and its relation to mortality risk. *Personality and Social Psychology Bulletin*, 28, 1155–1165.
- Martin, R. A. (2003). Sense of humor. In S. J. Lopez, & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and methods* (pp. 313–326). Washington, DC: American Psychological Association.
- Martin, R. A., Puhlik-Doris, P., Larson, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, 37, 48–75.
- Provine, R. R. (2000). *Laughter: A scientific investigation*. New York: Penguin.
- Rotton, J., & Shats, M. (1996). Effects of state humor, expectancies, and choice on postsurgical mood and self-medication: A field experiment. *Journal of Applied Social Psychology*, 26, 1775–1794.



# Hypnosis

John F. Kihlstrom

University of California, Berkeley, California, USA

1. Hypnotizability and Its Correlates
  2. Pain and Analgesia
  3. Other Clinical Applications
  4. Sensorimotor Performance Enhancement
  5. Learning and Memory
  6. Risks of Hypnosis
  7. Conclusion
- Further Reading

## GLOSSARY

**age regression** In hypnosis, an experience in which the hypnotized person returns to a scene or an event experienced in the past.

**analgesia** A reduction in pain.

**anesthesia** A reduction in tactile sensation; in surgery, this is often accompanied by a general loss of consciousness.

**conscious sedation** A surgical technique in which the patient remains conscious during a procedure but receives analgesic drugs to reduce pain as well as sedative drugs to reduce anxiety and induce amnesia for the procedure.

**false alarms** In memory testing, an instance where the person incorrectly identifies an item as having appeared on a previously studied list.

**hemispheric specialization** The idea that the right and left hemispheres of the cerebral cortex are specialized for different psychological functions.

**hyperesthesia** Enhancement of sensory acuity and other perceptual abilities; it is the opposite of anesthesia.

**hypermnnesia** Enhancement of memory; it is the opposite of amnesia.

**hyperpraxia** Enhancement of muscular strength, endurance, and motor ability; it is the opposite of apraxia.

**hypnotizability** Individual differences in response to hypnosis, as measured by standardized psychological tests such as the Stanford Hypnotic Susceptibility Scales.

**openness to experience** A “Big Five” personality trait characterized by richness in fantasy life, aesthetic sensitivity, awareness of inner feelings, need for variety in actions, intellectual curiosity, and liberal values.

**psychodynamic** Broad label for a class of theories of personality and psychotherapy based on Freud’s idea that behaviors are motivated by unconscious conflict.

**psychosomatic** The term applied to medical illnesses and other physical conditions that appear to be caused by mental states such as particular beliefs and attitudes.

**self-fulfilling prophecy** A situation in which a person’s beliefs and expectations lead him or her to behave in such a manner as to make the expectations come true, thereby confirming the person’s initial beliefs.

Hypnosis is a social interaction in which one person responds to suggestions given by another person (the hypnotist) for imaginative experiences involving alterations in perception, memory, and the voluntary control of action. In the classic instance, these suggested experiences are associated with a degree of subjective conviction bordering on delusion and an experience of involuntariness bordering on compulsion. Hypnosis has its origins in the “animal magnetism” promoted by Franz Anton Mesmer, an Austrian physician, for the treatment of medical illnesses.

Mesmer's theories were discredited in 1784 by a French royal commission chaired by Benjamin Franklin, but interest in his technique persisted. Over the years, animal magnetism was transformed into "mesmerism," then "artificial somnambulism," then "neurohypnotism," and finally hypnosis as we know it today. Since the time of William James and Ivan Pavlov, hypnosis has been of interest to psychologists and other scientists and philosophers for the insights it can yield about perception, memory, consciousness, and other basic psychological functions. However, it has also found its way into practical applications of various sorts.

### 1. HYPNOTIZABILITY AND ITS CORRELATES

It is important to understand that not everyone responds positively to hypnotic suggestions. Individual differences in hypnotizability are measured by performance-based tests such as the Stanford Hypnotic Susceptibility Scale and the Harvard Group Scale of Hypnotic Susceptibility. Some researchers and clinicians use alternative scales, such as the Hypnotic Induction Profile, but the Harvard scale is the most popular choice for this purpose and the Stanford scale remains the "gold standard" for measuring hypnotizability. The hypnotizability scales consist of a standardized induction procedure accompanied by a set of representative hypnotic suggestions, the response to which is scored according to objective behavioral criteria. As measured by the scales, most people are at least moderately responsive to hypnotic suggestions; relatively few people are refractory to hypnosis, and relatively few "hypnotic virtuosos" fall within the highest level of responsiveness. Some theorists have argued that hypnotizability can be enhanced by developing positive attitudes, motivations, and expectancies concerning hypnosis. As with any other skilled performance, hypnotic response is probably a matter of both aptitude and attitude. Negative attitudes, motivations, and expectancies can interfere with performance, but positive ones are not by themselves sufficient to create high hypnotizability.

Hypnotizability is not substantially correlated with most other individual differences in ability or personality such as intelligence and extraversion. One exception is absorption, that is, the tendency to have subjective experiences characterized by the full engagement of attention (narrowed or expanded) and blurred

boundaries between self and object. Absorption, in turn, is a facet of the "openness to experience" dimension in the "Big Five" structure of personality. However, even these correlations are too weak to permit hypnotizability to be predicted in advance on the basis of the usual sorts of personality questionnaires.

Although hypnosis is commonly induced by suggestions for relaxation and even sleep, the brain activity in hypnosis more closely resembles that of a person who is awake. The discovery of hemispheric specialization led to the speculation that hypnosis was a "right-brain" function, but it is obvious that response to hypnotic suggestions also requires the verbal functions normally associated with the left cerebral hemisphere. Because the experience of involuntariness is a central feature of hypnosis, some theorists have proposed that the state entails alterations in functioning of the frontal lobes of the brain that are known to be involved in executive control. However, a better understanding of the neural substrates of hypnosis awaits studies of neurological patients with focalized brain lesions as well as brain-imaging studies (e.g., positron emission tomography [PET], functional magnetic resonance imaging [fMRI]) of normal individuals.

### 2. PAIN AND ANALGESIA

By far, the most successful clinical application of hypnosis involves suggestions for analgesia to relieve pain. As early as the 1840s, John Elliotson and James Esdaile each reported that they had performed a large number of surgeries painlessly with artificial somnambulism as the only anesthetic agent. However, beginning in 1846, the introduction of chemical anesthetics, such as ether and chloroform, quickly supplanted hypnosis. Nevertheless, when chemical analgesics or anesthetics were unavailable or contraindicated, physicians still sometimes turned to hypnosis. Between 1955 and 1974, more than two dozen cases in which hypnosis had been used as the sole analgesic or anesthetic agent in surgery were published. Of even greater importance was the publication, during the early 1960s, of extensive case series documenting the successful use of hypnosis in obstetrics and in the treatment of cancer pain. Other clinical studies found that hypnosis can effectively relieve pain in dental patients. One comprehensive review of the clinical literature estimated that approximately half of medical, dental, and surgical patients could obtain significant pain relief through hypnosis alone. More recent reviews confirm that even when patients are unselected for

hypnotizability, hypnotic suggestions can have quite substantial effects on both acute and chronic pain. Hypnosis may be especially useful in cases of chronic pain, where chemical analgesics such as morphine pose risks of tolerance and addiction, and in childbirth, where the active participation of the mother may be desirable.

In addition to replacing chemical analgesics when circumstances warrant, hypnosis can serve as a useful adjunct to more conventional medical treatment of pain such as “conscious sedation” for outpatient surgery. Careful clinical studies by at least three independent groups of investigators have shown that the addition of hypnosis reduces both pain and anxiety. Patients request, and receive, less pain medication than do controls. Hypnosis is associated with a decrease in interruptions and adverse events during surgery. The procedure takes less time, and patients recover faster, when patients receive both hypnosis and medication. Although hypnosis can rarely substitute for the chemical analgesia and anesthesia that is the standard of care in modern medicine, it is a cost-effective adjunctive treatment (a complement if not an alternative) that improves the quality of patient care.

In addition to this increasing body of clinical research, a large amount of laboratory research has shed light on the mechanisms by which hypnotic suggestions have their effects. Hypnotic analgesia is not mediated by mere relaxation or by the release of endogenous opiates (endorphins). It is not merely a placebo effect (although there is a placebo component in hypnotic analgesia, just as there is in chemical analgesics). Hypnosis can reduce both sensory pain and suffering; in fact, suggestions specifically targeting one or the other of these components of pain have selective effects on activity in the somatosensory and anterior cingulate regions of the brain. Psychologically speaking, hypnotic analgesia seems to go beyond self-distraction, stress inoculation, cognitive reinterpretation, and tension management. According to one prominent theory, hypnotic analgesia involves a dissociation, or a division of consciousness, that effectively reduces patients’ awareness of pain.

### 3. OTHER CLINICAL APPLICATIONS

Hypnotic suggestion can have “psychosomatic” effects on aspects of bodily functioning other than pain, and these can also be useful in the practical sense. For example, it has frequently been observed that hypnotic suggestions can reduce bleeding from wounds and speed the healing of burns. However, these effects have not yet been subject to rigorous clinical and experimental study.

Several well-controlled laboratory and clinical studies have shown that hypnotic suggestion can affect allergic responses, asthma, and the remission of warts. Such successes have led some practitioners to offer hypnosis in the treatment of cancer, but there is no evidence that hypnosis affects tumors themselves as opposed to patients’ quality of life.

Hypnosis has also been used in psychotherapy. Some psychodynamic theorists, working in the tradition of Sigmund Freud, consider hypnosis to be a form of “adaptive regression” or “regression in the service of the ego,” and they use it to promote relaxation, enhance imagery, and generally loosen the flow of free associations. However, there is little evidence from controlled outcome studies that “hypnoanalysis” or “hypnotherapy” is more effective than nonhypnotic forms of the same treatment. In contrast, a recent review found substantial treatment gains when hypnosis was used adjunctively in cognitive-behavioral therapy for problems such as obesity, insomnia, anxiety, phobia, and hypertension. In view of the demands of managed medical and mental health care, it is important for practitioners who use hypnosis to document quantitatively the clinical (and economic) benefits of doing so.

An important but unresolved issue is the role played by individual differences in the clinical effectiveness of hypnosis. In the laboratory, as in the clinic, a genuine effect of hypnosis should be correlated with hypnotizability. Unfortunately, clinical practitioners are often reluctant to assess hypnotizability in their patients and clients due to a concern that low scores might reduce motivation for treatment. This danger is probably exaggerated. On the contrary, assessment of hypnotizability by clinicians who are contemplating the therapeutic use of hypnosis would seem to be no different, in principle, from assessment of allergic responses before prescribing an antibiotic. In both cases, the legitimate goal is to determine what treatment is appropriate for what patient.

It should be noted that clinicians sometimes use hypnosis in nonhypnotic ways, and these practices tend to support the hypothesis that whatever effects they achieve through hypnosis are related to its placebo component. For example, there is nothing particularly “hypnotic” about having a patient in a smoking cessation treatment rehearse therapeutic injunctions not to smoke and other coping strategies while hypnotized. It is possible that many clinical benefits of hypnosis are mediated by placebo-like motivational and expectational processes, that is, by the “ceremony” surrounding hypnosis rather than by hypnosis per se.

#### 4. SENSORIMOTOR PERFORMANCE ENHANCEMENT

Much of the popular interest in hypnosis stems from claims that it permits people to transcend their normal voluntary capacities, yielding increased muscular power, endurance, resistance to fatigue, and sensory acuity as well as other special abilities. However, controlled research has largely failed to find evidence that hypnosis can enhance these and other aspects of human performance. Many early studies, which seemed to yield positive results for hypnosis, possessed serious methodological flaws such as the failure to collect adequate baseline information. For example, individuals may “hold back” on nonhypnotic trials so as to satisfy experimenters’ implicit expectations that they show “improved” performance during hypnosis. Of course, in some situations, benefits may be achieved indirectly through suggestions for reduced awareness of pain. For example, hypnotized persons may experience themselves as stronger, but this is not the same as a direct enhancement of human performance capacities *per se*.

In general, it appears that hypnotic suggestions for increased muscular strength, endurance, sensory acuity, and/or learning do not exceed what can be accomplished by appropriately motivated persons outside of hypnosis. Nevertheless, the folklore surrounding hypnosis may lead hypnotized persons to achieve performance enhancements through a kind of self-fulfilling prophecy. For example, instead of giving suggestions for increased muscle strength, one intriguing study asked hypnotizable persons to imagine themselves becoming stronger. These individuals actually outperformed controls who were merely exhorted to maximize their performance. If hypnosis does enhance human performance, it may be by virtue of hypnotic alterations in individuals’ self-concepts rather than by virtue of their skeletal musculatures.

Another highly intriguing line of research contends that hypnotic suggestions may enhance visual acuity in individuals with myopia (although not in individuals with normal vision). The research is especially intriguing because the investigators took special care to rule out “peripheral” factors, such as accommodation and other structural changes in the eyes, that might account for the effect. If hypnotic suggestions do in fact improve visual acuity in people with impaired vision (and this claim still must be confirmed by additional research), the effects appear to be mediated centrally, perhaps involving cortical centers associated with vision or visual attention.

It is sometimes claimed that hypnosis enhances “parapsychological” abilities such as clairvoyance, telepathy, and telekinesis. However, there is no good evidence that this is the case, not least because there is no good evidence of parapsychological effects in the first place.

#### 5. LEARNING AND MEMORY

Similar conclusions apply to learning capacity. Many studies of hypnotically enhanced learning suffer from the same methodological defects as do the aforementioned studies of muscular performance and sensory acuity. One well-known line of research attempted to enhance learning by giving individuals hypnotic suggestions for time distortion. By extending the “subjective” interval of time devoted to studying the material, these investigators hoped to enable these individuals to learn more over short intervals of time. However, the promising results of early experiments have largely failed to be confirmed by more tightly controlled follow-up studies. Hypnotized persons may well experience time as flowing more slowly, or more rapidly, than it actually is, but this distortion in perception does not have positive consequences for learning.

A special case of performance enhancement pertains to hypnotic suggestions for improvements in memory or “hypnotic hypermnnesia.” Hypermnnesia suggestions are sometimes employed in forensic situations, to help forgetful witnesses and victims remember, and in therapeutic situations, to help patients remember childhood sexual abuse and other traumatic personal experiences. Sometimes, suggestions for enhanced memory are accompanied by suggestions for age regression in which hypnotized persons are asked to return to a previous period in their lives. Although the clinical literature contains a number of reports claiming success, most of these are anecdotal in nature and fail to obtain independent corroboration of the memories that emerge. There is no clinical evidence that hypnotically enhanced memories are reliable, that is, that they can be taken as true without independent confirmation.

In fact, evidence from carefully controlled laboratory studies shows quite the opposite. Hypnosis does not appear to increase accurate recollection over and above what can be achieved under nonhypnotic circumstances. To make things worse, hypnosis appears to increase the likelihood of “false alarms” on recognition tests and to inflate individuals’ confidence in their memories, independent of their accuracy. Moreover, by virtue of their

enhanced responsiveness to suggestion, hypnotized persons may be especially vulnerable to leading questions and other suggestive influences that might bias or distort their memories. Hypnotized persons may well believe that they are recovering forgotten memories, but that does not make the memories valid representations of the historical past. Any memories recovered by means of hypnosis must be subject to independent corroboration before they are taken seriously.

Similar considerations apply to hypnotic age regression. Although age-regressed individuals may experience themselves as children and may behave in a distinctly childlike manner, there is no evidence that they actually undergo either abolition of characteristically adult modes of mental functioning or reinstatement of childlike modes of mental functioning. Nor do age-regressed individuals experience the revivification of forgotten memories of childhood.

Given what we know about the unreliability of hypnotic hypermnnesia, the use of hypnosis to recover memories in clinical and forensic situations must be judged to be highly risky and not recommended. In fact, many legal jurisdictions severely limit the introduction of memories recovered through hypnosis due to a concern that such evidence might be tainted. The Federal Bureau of Investigation (FBI) has published a set of guidelines for those who wish to use hypnosis forensically, and similar precautions should be employed in the clinic. Although field studies have sometimes claimed that hypnosis can enhance memory in a powerful way, these anecdotal reports have not been duplicated under laboratory conditions.

## 6. RISKS OF HYPNOSIS

Hypnosis poses few risks to persons who are hypnotized. Because it requires the active participation of the hypnotized persons, people cannot be hypnotized unknowingly or against their will. Nor can even deeply hypnotized persons be led, by virtue of suggestion, to do something that they would not do under other circumstances. In experimental situations, the risks to individuals are usually limited to drowsiness or mild headache—no different in kind or severity from those that crop up when students take exams or listen to lectures. In most instances, these transient problems can be rapidly traced to some specific expectation on the part of the hypnotized persons, perhaps derived from inadvertent comments made by a stage performer or

television commentator seen recently. Occasionally, a hypnotized person actually dozes off during the procedure or rests his or her head in an uncomfortable position, leading to drowsiness or neckache on arousal. In any event, a few minutes of supportive interaction are usually sufficient to deal with the problem effectively. However, problems may be compounded when inexperienced experimenters find it difficult to cope with a minor problem and communicate their lack of competence and self-confidence to the hypnotized person, who may then proceed to make matters even worse. Problems may also occur in stage hypnosis, where the entertainment context has special effects on experience and behavior over and above the hypnotic context.

Untoward consequences are more likely to arise when hypnosis is employed during therapeutic, as opposed to experimental, procedures. A manipulative therapist who simply tries to “suggest away” symptoms may encounter problems that stem from his or her own therapeutic errors. When hypnosis is used to recover ostensibly forgotten memories, whether in therapeutic or forensic situations, the memories of the hypnotized person may be contaminated with information that is objectively false—with potentially serious negative consequences for other people. When hypnosis is used to control pain in athletic situations, competitors may ignore important information about the location and severity of bodily damage. In other cases of performance enhancement, it is possible that individuals will harm themselves through overexertion. Practitioners should not attempt to treat with hypnosis any condition that they are not qualified to treat without hypnosis.

## 7. CONCLUSION

Claims of hypnotically induced performance enhancement are nearly as old as hypnosis itself. In George DuMaurier’s novel *Trilby*, published in 1895, Svengali transforms the eponymous heroine into an accomplished singer. And reviewing the hypnosis literature nearly half a century ago, F. L. Marcuse referred to “The Generation of Hypers,” that is, the notion that hypnosis could produce hyperpraxia, hyperesthesia, and hypermnnesia as well as other special effects. Aside from the ability of hypnotic suggestion to control pain—a genuine phenomenon of substantial practical importance—most claims for hypnotic performance enhancement are dubious at best. Hypnosis is fascinating to observe, is interesting from a theoretical standpoint, and has many actual and potential

applications as an adjunctive treatment in medicine and psychotherapy, but it will not make us better people.

### Acknowledgments

The point of view represented in this article is based on research supported by Grant MH-35856 from the National Institute of Mental Health.

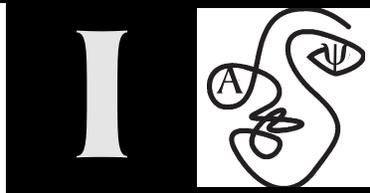
### See Also the Following Articles

Learning ■ Pain Management ■ Psychoneuroimmunology  
■ Psychotherapy in Older Adults

### Further Reading

Bowers, K. S. (1976). *Hypnosis for the seriously curious*. Pacific Grove, CA: Brooks/Cole.  
Fromm, E., & Nash, M. R. (Eds.). (1992). *Contemporary hypnosis research*. New York: Guilford.

Gauld, A. (1992). *A history of hypnotism*. Cambridge, UK: Cambridge University Press.  
Hilgard, E. R. (1965). *Hypnotic susceptibility*. New York: Harcourt, Brace, & World.  
Hilgard, E. R., & Hilgard, J. R. (1975). *Hypnosis in the relief of pain*. Los Altos, CA: Kaufman.  
Kihlstrom, J. F., & Barnhardt, T. M. (1993). The self-regulation of memory: For better and for worse, with and without hypnosis. In D. Wegner, & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 88–125). Englewood Cliffs, NJ: Prentice Hall.  
Kihlstrom, J. F., & Eich, E. (1994). Altering states of consciousness. In D. Druckman, & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 207–248). Washington, DC: National Academy Press.  
Kirsch, I., Montgomery, G., & Sapirstein, G. (1995). Hypnosis as an adjunct to cognitive-behavioral psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 63, 214–220.  
Nash, M. R. (2001, July). The truth and the hype of hypnosis. *Scientific American*, pp. 47–55.  
Patterson, D. R., & Jensen, M. P. (2003). Hypnosis and clinical pain. *Psychological Bulletin*, 129, 495–521.



# Ideological Orientation and Values

Hans De Witte

Catholic University of Leuven, Leuven, Belgium

1. Ideology and Values: Definitions
  2. Values and Value Orientations in Political Psychology
  3. On the Dimensionality of (Conservative) Ideology
  4. Political Consequences of Ideology and Value Orientations
- Further Reading

## GLOSSARY

**attitude** A relatively enduring organization of beliefs around a concrete object or situation, predisposing people to respond in some preferential manner.

**competition** Value orientation that combines the acceptance of inequality with a focus on the individual. Central values are commitment to achieving wealth, social recognition, and social power.

**conformity (to external authorities)** Values stressing obedience to external norms and the need to follow the dictates of authority.

**ideology** A coherent set of attitudes and values, shared by a social group, that motivates their actions.

**materialist** Values expressing the need for economic and financial security, societal stability, personal safety, and law and order.

**post-materialist** Values referring to the need for political freedom and participation, self-actualization, personal relationships, creativity, and care for the environment.

**self-direction** Values emphasizing autonomy and the personal responsibility of the individual.

**(socio-)cultural conservatism versus progressiveness** Also labeled “sociocultural libertarianism–authoritarianism.” Ideological opposition related to the old cleavage between

the Church and the State, typical of pre-industrial sources of political conflict. Conservatism emphasizes traditional norms (e.g., traditional gender roles and strict child rearing practices). Progressiveness puts traditional norms into perspective, stressing, e.g., women’s emancipation and the autonomy of children, and supporting the liberalization of abortion.

**(socio-)economic conservatism versus progressiveness** Also labeled “socio-economic left-right.” Ideological opposition related to socioeconomic cleavages in industrial society. A progressive stand rejects social inequality as unfair and argues for government intervention and stronger trade unions in order to reduce economic inequality. Conservatives oppose government intervention and (militant) trade unions, and emphasize private initiative and individual competition as essential for proper economic functioning.

**universalism** Value orientation that combines equality and focus on others into a universalistic, altruistic value pattern. Equality, social justice, and true friendship are key values in this orientation.

**values** Abstract, fundamental goals. They are evaluative standards with normative power. Values are hierarchically ordered in a coherent pattern.

This article discusses the content and relevance of ideology and values for political psychology. Even though certain scholars advocated “the end of ideology” in the past, recent events suggest that ideology still plays an important role in contemporary society, in shaping and affecting the life of everyday people and of specific, politically active groups. The

terrorist attacks of September 11, 2001 are a powerful example of the impact of an extreme ideology on innocent citizens. This event also demonstrates that radical ideologies can motivate activists to perform extreme acts, by sacrificing their life for a “higher cause.” Less drastic examples of the relevance of ideology can be found in the rise of new political parties on the left (e.g., green parties) and the right (e.g., extreme right-wing parties), which attracted more voters than before during parts of the last decade.

In this article, first, the concepts of ideology and values are defined. Next, influential theories on values and value orientations in political psychology are reviewed, starting with the less complex ones, which will be integrated in more recent multidimensional theories at the end. The third section discusses the dimensionality of conservatism, and relates it to the values section, and to the specific groups that are holding this ideology. Fourth, the political consequences of ideology and value orientations are outlined. The article concludes with some crucial references for further reading.

## **1. IDEOLOGY AND VALUES: DEFINITIONS**

The term ideology is used in various disciplines such as psychology, sociology, and political science. Consequently, it has many meanings. In psychological research, however, consensus exists regarding three core aspects of an ideology: it is a coherent set of attitudes and values, it is shared by and therefore typical of a given social group, and it motivates their actions.

Ideology is not a mixed bag of attitudes. Rather, it is a logical and (relatively) coherent world view and “portrayal of man” that provides an understanding of reality: it offers an interpretation and an explanation of how the world works and of the role and motives of the people in that world. This world view and vision of mankind are founded on certain values. As a consequence, an ideology does not only describe “what is” (first aspect), but also “what ought to be.” So, ideologies are not purely descriptive, they are also prescriptive or normative: they indicate what should be achieved and changed. They describe the desired goal and direct or motivate action (third aspect). This mobilizing function, or incentive to act, is linked to the collective nature of an ideology: ideologies refer to a set of attitudes and values that are shared by a larger

group (second aspect). They reflect the interests of a specific social group and offer identification possibilities to the members of these groups. By accepting the collective ideology, they can show that they too belong. These shared viewpoints can be used to make a distinction between those who belong (“us”) and those to whom this does not apply (“them”). An ideology also describes the group goals that must be achieved in order to secure the group’s interests or to improve the position of the group and its members.

Attitudes and values are part of an ideology. This means that these concepts must also be defined. A wide variety of definitions is used for the concept of “values” in the literature. There is, however, consensus on a number of basic features. Values are fundamental goals. They are evaluative standards with normative power: they control behaviors, thoughts, and feelings. Values are abstract, not directly observable constructs. They are not isolated from each other, but form a coherent pattern that is hierarchically ordered. Finally, values also have a certain continuity and stability. Most recent definitions of values broadly embrace these aspects of the concept. Schwartz (1994), for example, defines a value as “a belief pertaining to desirable end states or modes of conduct, that transcends specific situations, guides selection or evaluation of behavior, people and events, and is ordered by importance relative to other values to form a system of value priorities” (p. 20).

Rokeach, who can be considered one of the founding fathers of value research, emphasized the distinction between values and attitudes. He defined attitudes as a relatively enduring organization of beliefs around an object or situation, predisposing one to respond in some preferential manner. Attitudes are thus more concrete than values, because they refer to a specific object or specific situation. Values have a more abstract and fundamental character. A set of items that relates to the evaluation of income differences and to government intervention in the economy can therefore be considered as a set of attitudes. Because values are more fundamental, they are also more limited in number. An individual can subscribe to a large number of attitudes, but only to a limited number of values. Attitudes thus form “concretizations” of values, because a single fundamental value can be translated into a large number of attitudes with regard to actual topics. Finally, Rokeach also assumes that values occupy a more central position than attitudes within one’s personality and cognitive system. As a consequence, they are determinants of attitudes and behaviors.

## 2. VALUES AND VALUE ORIENTATIONS IN POLITICAL PSYCHOLOGY

Research into values and value orientations is of course not restricted to political psychology. In this field, numerous contributions of scholars in other fields of the social sciences and psychology are used and applied to the political sphere. The following sections discuss some of the most influential, starting with the most limited ones, focusing on just a selection of values. These will be integrated in more recent multi-dimensional accounts later on.

### 2.1. Kohn: Conformity versus Self-Direction and Its Relationship with Social Class

Kohn and his co-workers performed important research on the relationship between values and the position of individuals in the social structure. Their research shows that the middle class is oriented toward the value of self-direction. Respondents from the middle class find it more desirable to behave on the basis of their own insights and emphasize self-direction, autonomy, and the personal responsibility of the individual. In short, they value thinking for oneself. This value orientation has important consequences for their behavior in a variety of life domains. One of these domains is the upbringing of children, in which they value aspects such as consideration, curiosity, and self-control. Blue-collar workers, on the other hand, emphasize conformity to external authorities: they stress obedience to external norms, and the need to follow the dictates of authority. Examples of values stressed in the upbringing of children are obedience, good manners, and neatness. These scholars also demonstrate that self-direction and conformity (to external authority) are opposite poles of a single continuum.

Kohn and his co-workers stress that the difference in attitudes and values between blue-collar workers and white-collar workers is caused by the extent to which their work requires “occupational self-direction.” This variable refers to autonomy in the workplace and is composed of three components: the complexity of the work, the amount of supervision, and the (non-)repetitive character of the job. The extent to which someone can work autonomously depends on his or her occupational position. Employees in higher professions can work more autonomously: their work is more varied,

less supervised, and more complex, so that they must make more decisions on their own. Through longitudinal research, Kohn and his co-workers demonstrated that one’s capacity to work autonomously influences the values one develops. Work that requires autonomy leads to an orientation toward self-direction. When the work does not allow autonomy, then one has to adapt to the (work) environment by emphasizing conformity. The logic behind this link is that obeying external authorities forms the best adjustment to a situation in which one has little control of the course of events.

### 2.2. Inglehart: Materialism versus Post-Materialism

Inglehart’s post-materialism thesis is an example of a theory that advocates the social and political relevance of a new set of values in society. According to Inglehart, advanced Western societies experience a shift from materialist to post-materialist values. Material values express the need for economic and financial security, societal stability, personal safety, and law and order. Post-materialist values refer to the need for political freedom and participation, self-actualization, personal relationships, creativity, and care for the environment. In modern societies, a “silent revolution” is taking place, in which the dominance of materialist values is being replaced by post-materialist values. This value change is caused by the continuous replacement of old generations by new ones. Crucial in this respect is the degree to which fundamental material needs have been met during the formative period of individuals of a given generation. The post-materialism thesis, therefore, combines two hypotheses: a scarcity hypothesis and a socialization hypothesis. The scarcity hypothesis states that individual priorities reflect their socioeconomic environment: the greatest subjective value is placed on what is in relative short supply. The socialization hypothesis assumes that one’s values reflect the conditions during one’s pre-adult formative years and are relatively stable over the individual life course. Per Inglehart (1990), “Taken together, these two hypotheses imply that, as a result of the historically unprecedented prosperity and the absence of war that has prevailed in Western countries since 1945, younger birth cohorts place less emphasis on economic and physical security than do older groups, who have experienced a much greater degree of economic insecurity and that, conversely, the younger birth cohorts tend to give a higher priority to nonmaterial needs,

such as a sense of community and the quality of life” (p. 56). Inglehart’s post-materialism thesis is based on the well-known need hierarchy of Maslow, suggesting that material needs must be fulfilled before immaterial or “higher” (growth) needs such as self-actualization become important. As a consequence, support for post-materialist values comes from post-war generations, the highly educated, and parts of the middle class. By contrast, pre-war generations, the less-educated, and members of the old middle class and the working class are more likely to support materialist values.

### **2.3. Rokeach and Braithwaite on the Political Importance of Values**

Directly relevant to the domain of political psychology are the contributions of Rokeach and Braithwaite. Rokeach developed the two-value model of political ideology. He assumed that political ideologies differ from each other in terms of the social values freedom (independence, free choice) and equality (brotherhood, equal opportunity for all): the combination of these values enables us to describe and contrast all ideologies. According to Rokeach, socialism emphasizes both freedom and equality, whereas fascist ideology values neither freedom nor equality. Capitalism and communism both emphasize one of these core values more than the other: capitalism values freedom at the expense of equality, and communism values equality at the expense of freedom. Support for this model was found through a values content analysis of the writings of crucial authors advocating these different ideologies. Analyses of the values of supporters of different ideologies (such as voters or candidates of political parties), however, are less consistent with Rokeach’s theory. This research generally shows rather important differences in the importance attributed to equality, with leftist activists showing a greater emphasis on equality than right-wing or conservative supporters. This is in line with Rokeach’s theory. The value of freedom does not discriminate these different ideological groups, however, which refutes his claim.

These findings inspired Braithwaite to elaborate Rokeach’s two-value model. She reduced a large number of values to two broader value orientations. The first is called “international harmony and equality,” and covers values such as participation, cooperation, peace, and equal opportunities for all. The second refers to “national strength and order,” and covers values such as national security, economic

advancement, order, and security. Whereas the value orientation international harmony and equality captures Rokeach’s notion of equality, national strength and order refers to a concept different than freedom. Braithwaite also shows that both value orientations are highly relevant for political attitudes and behaviors. Those favoring international harmony and equality emphasized progressive political views, expressed a preference for a leftist political party, and were more inclined toward political activity. The opposite holds for supporters of national strength and order: they preferred conservative views and a right-wing political party and were less inclined to engage in political activity.

### **2.4. Schwartz and Triandis: More Encompassing Value Orientations**

Next to the previous contributions, two broad, more encompassing theories on value orientations have been developed, integrating and expanding previous research in the area. Building on the work of Rokeach, Schwartz proposed a universal theory of the content and structure of values. This theory has been tested in more than 60 countries, and is proven to be cross-culturally stable. Rokeach’s value survey comprised 36 values, which were presented in two groups: 18 terminal values and 18 instrumental values. Respondents were asked to rank the values within each list according to their own preference. Schwartz increased this list to 56 values in his value inventory, and asked respondents to score each value on a 9-point scale. So, in addition to the increase in the number of values (again split up into two series), he also changed Rokeach’s method of scoring from “ranking” to “rating,” enabling the data to be analyzed with more sophisticated statistical methods. Schwartz divided the 56 values in his survey into ten motivational value types, according to the type of motivational goal they express. [Table I](#) summarizes these value types and their content.

An inspection of [Table I](#) shows that Schwartz’s ten value types cover all the different value orientations discussed in the preceding section. Kohn’s dichotomy between self-direction and conformity is covered by identical labels (with identical content), whereas Rokeach’s core value freedom is part of self-direction. Universalism is closely related to Inglehart’s post-materialism and Braithwaite’s international harmony and equality, and includes Rokeach’s core value of

TABLE I  
Schwartz's Ten Motivational Value Types

Value type	Definition	Exemplary values
<i>Power</i>	Social status and prestige, control, or dominance over people and resources	Social power, authority, wealth
<i>Achievement</i>	Personal success through demonstrating competence according to social standards	Successful, capable, ambitious
<i>Hedonism</i>	Pleasure and sensuous gratification for oneself	Pleasure, enjoying life
<i>Stimulation</i>	Excitement, novelty, and challenge in life	Daring, varied life, exciting life
<i>Self-direction</i>	Independent thought and action—choosing, creating, exploring	Creativity, curious, freedom
<i>Universalism</i>	Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature	Broad-minded, social justice, equality, protecting the environment
<i>Benevolence</i>	Preservation and enhancement of the welfare of people with whom one is in frequent personal contact	Helpful, honest, forgiving
<i>Tradition</i>	Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide	Humble, devout, accepting my portion in life
<i>Conformity</i>	Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms	Polite, obedient, honoring parents and elders
<i>Security</i>	Safety, harmony, and stability of society, of relationships, and of self	National security, social order, clean

Source: Schwartz (1994).

equality. Finally, security covers both Inglehart's materialism and Braithwaite's national strength and order. Schwartz, however, distinguishes more value types than these already suggested ones, thus enlarging the scope and possibilities of values research.

Schwartz represents his structure and value types by means of multidimensional scaling. This technique represents the values as points in a multidimensional space. The spatial distance between the points represents the relationship between the values. The more the ratings of a value correlate, the closer the values are located in the graphical representation. For Schwartz's universal values theory, this technique implies that the separate values of his motivational types must lie in each others' proximity and must be distinguishable as a separate field or region in the graphical representation. In contrast, values and value types that express opposing motivations should be discriminated clearly from one another. Fig. 1 shows the value structure obtained by Schwartz by means of such an analysis.

Schwartz claims that the oppositions in Fig. 1 can be classified into two bipolar dimensions. He describes the first dimension as openness to change, as opposed to conservation. This dimension contrasts value types such as stimulation, and self-direction (openness to

change) with value types such as security, conformity, and tradition (conservation). The second dimension is formed by the opposition between self-transcendence and self-enhancement. Value types such as universalism and benevolence are opposed to value types that refer to power and achievement. Schwartz's subsequent research shows that his model and the two abstract dimensions are (very nearly) universal.

Schwartz's use of both higher order dimensions fits in with a research school that classifies the structure of values on the basis of two dimensions, giving rise to four value orientations. The key scholar in this school is Triandis. The first dimension distinguished by Triandis refers to the opposition between individualism (focus on the self) and collectivism (focus on others). The value types tradition, benevolence, and conformity correspond to collectivism. Values such as hedonism and stimulation refer to individualism. The opposition between "vertical" and "horizontal" thinking constitutes the second dimension. Horizontalism means that people consider each other as equal. Universalism typifies this orientation. Vertical thinking accepts forms of inequality, hierarchy, and authority (cf. the value type power). Triandis' two fundamental dimensions allow a more unambiguous interpretation

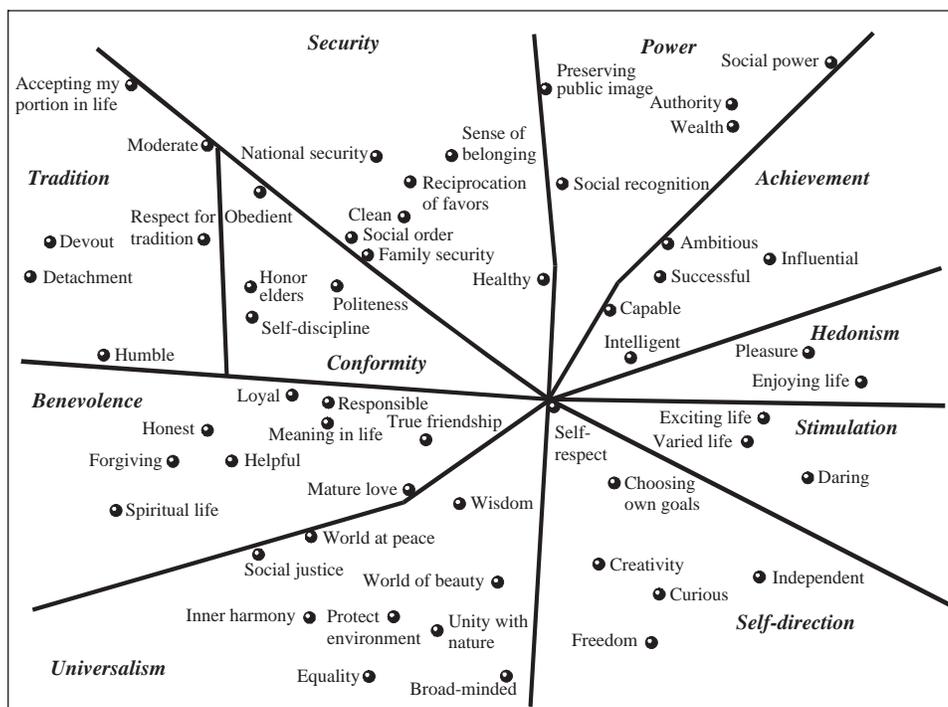


FIGURE 1 Visual representation of Schwartz's value structure. From Schwartz (1994).

than the two abstract value dimensions advanced by Schwartz. Combining the views of both scholars leads to four value orientations:

1. Horizontal collectivism or universalism combines the idea of equality and the focus on others to form a universalistic, altruistic value pattern. Equality, social justice, and true friendship are key values in this orientation. This value orientation appears in the bottom left in Fig. 1.

2. Horizontal individualism or self-direction combines the idea of equality with a focus on the individual. This leads to an emphasis on autonomy (with values such as choosing one's own goals and freedom), with peaks such as stimulation (daring and exciting life) and (slightly less central) hedonistic values (such as pleasure and enjoying life). This value orientation is located at the bottom right in Fig. 1.

3. Vertical individualism or competition combines the acceptance of inequality (the existence of a hierarchically ordered society) with a focus on the individual. The central idea is that the individual can achieve the "top" in society by adopting a competitive attitude. Central values are the commitment to achieving wealth, social recognition, and social power. This value orientation is located at the top right in Fig. 1.

4. Vertical collectivism or conformity, finally, combines a focus on others with the acceptance of a hierarchy. Supporters of this view are prepared to submit themselves to the collective. Key values are obedience, politeness and respect for tradition. It is located at the top left in Fig. 1.

### 3. ON THE DIMENSIONALITY OF (CONSERVATIVE) IDEOLOGY

Values are the building stones of ideologies. The exact content of contemporary ideology, however, has been the subject of large-scale debates between scientists of different disciplines. Even though important differences in position exist, many scholars agree that (social and political) ideology can be captured under the heading of conservatism. The definition of conservatism, however, again proves difficult. Some, like Wilson (1979), proposed a one-dimensional concept, caused by "a generalized susceptibility to feeling threat or anxiety in the face of uncertainty" (p. 65). More prominent are multidimensional accounts. Van Deth and Scarbrough distinguished three important broad

value orientations (or ideologies), which cover important aspects of advanced industrial society as well as the legacies of industrial and pre-industrial societies. Inglehart's postmaterialism (as opposed to materialism) was selected as crucial for advanced (post-)industrial society. The distinction between left and right regarding economic issues was selected as a central value orientation reflecting the class conflicts typical of industrial society. Religious–secular value orientations were their third choice, referring to pre-industrial sources of political conflict.

### 3.1. Two Dimensions of Conservatism

Reviewing the field nevertheless makes clear that distinguishing a two-dimensional concept of ideology offers many advantages. This option is, first of all, suggested by several important scientists from different disciplines, such as Eysenck in psychology, and Lipset in political sciences. More importantly, these authors arrived at a rather similar structure, which corresponds closely to the thorough conceptualization of conservatism developed by the Dutch scholar Middendorp. On the basis of historical and theoretical considerations and a large body of empirical research, Middendorp distinguished two dimensions within conservatism and revealed their underlying values: freedom versus equality. The first was labeled socio-economic left-right and the second socio-cultural libertarianism-authoritarianism. In accordance with the literature in this field, these dimensions can be relabeled as economic conservatism (versus progressiveness) and sociocultural conservatism (versus progressiveness). Both dimensions are relatively stable over time, and—more surprisingly—are unrelated: respondents can combine a progressive view on one dimension with a conservative view on the other. Research also suggests that this two-dimensional conceptualization is rather universal.

The opposition between progressive and conservative attitudes in the socioeconomic domain points to socioeconomic cleavages in industrial society: conflicts between labor and capital. Middendorp showed empirically that the economic opposition between liberalism (in its European meaning of the concept) and socialism lies at the basis of this dimension: those who express progressive attitudes reject social inequality as unfair and unacceptable and argue for government intervention in the economy and for a more active trade union policy in order to reduce economic inequality.

Conservatives express opposition to government intervention in the economy and against (militant) trade unions. They equally reject the claim that social differences are caused by inequality of opportunity, and emphasize private initiative and competition between individuals as essential for proper economic functioning.

The opposition between conservative and progressive on a sociocultural level relates to societal norms and traditions with respect to individual private lifestyle and family life. Ethical topics (such as sexuality and bioethics) and more social themes (such as free speech and tolerance toward minorities) belong to this field. This ideological cleavage relates to Van Deth and Scarbrough's religious versus secular value orientation, typical of pre-industrial sources of political conflict. Cultural conservatism thus can be considered as a relic of the old cleavage between the Church and the State. Historically, the non-religious community in particular has opposed the norms and values dictated by the Church, whereas the latter promoted traditional norms and conformity to traditional institutions. In a conservative sociocultural view, traditional norms in society are emphasized: adherents plead to maintain existing social norms and traditions ("traditionalism"). They prefer the traditional division of roles between men and women and support severe child-rearing practices in which strictness and obedience are stressed. They also defend the traditional work ethic, which emphasizes the moral duty to work. Those who take a progressive stand in the sociocultural field generally hold opposite attitudes. Traditional norms and attitudes are put into perspective. They stress the emancipation of women and the independence of children, support the liberalization of abortion and euthanasia, and emphasize other reasons for work than strict duty-related ones (e.g., self-actualization).

### 3.2. Relationship with Values and Value Orientations

These two broad ideological dimensions are of course related to values. Middendorp argued that Rokeach's values freedom and equality are at the core of both conservatism dimensions. On the socioeconomic level, progressive thought emphasizes equality: unequal opportunities and an unequal distribution of welfare and well-being are rejected. In conservative thought, the emphasis on the socioeconomic level is on the value of (individual) freedom: private initiative and competition among individuals should not be limited. On the sociocultural level, this emphasis is

reversed. The progressive way of thinking emphasizes the freedom of the individual to arrange life according to one's own insight (self-direction). In the conservative way of thinking, such individual freedom is rejected and equality is stressed. Conservatives on the socio-cultural level emphasize that everyone should adhere to the traditional norms and values in society (conformity and traditionalism). So, individuals who emphasize the basic value of individual freedom adhere to opposing views on both ideological dimensions. On the economic level, they show conservative preferences, whereas on the sociocultural level they choose progressive viewpoints. The reverse is true for those emphasizing equality. They combine a progressive view on economic issues with a conservative one regarding sociocultural issues.

Both ideological conservatism dimensions are related not only to separate values, but also to larger value dimensions. Sociocultural conservatism is related to Kohn's distinction between self-direction and conformity. Those who emphasize conformity to external authorities find it equally important to obey the norms and traditions in society (traditionalism). These individuals stress external standards for their behavior: "others" have to tell us what to do. These others can be superiors in the hierarchy (vertical conformity), as well as equals (horizontal conformity). Individuals who take a progressive position on sociocultural issues defend freedom of thought and conscience. They emphasize the right to organize their lives according to their own insights, even if it means deviation from current norms. This view is closely related to self-direction, since the norm for one's own behavior is sought with oneself (internal standards for behavior).

Economic attitudes are related to value orientations as well. The value opposition between universalism and competition correlates with economic conservatism: economic progressiveness shows a positive correlation with the value orientation universalism and a negative correlation with the value orientation competition. The reverse is true for economic conservatism (positive correlation with competition and negative correlation with universalism).

### **3.3. View on Human Nature and Society, and Relationship with Social Groups**

Earlier, ideology was defined as a (relatively) cohesive world view and portrayal of humans, reflecting the interests of a specific social group. How does our

treatment of conservatism relate to these aspects? Based on an historical analysis, Middendorp showed that the ideology of conservatism combines a rather negative view on human nature with a rather positive view on the nature of society. In conservatism, one emphasizes that people are inclined to egoism rather than to altruism. As a consequence, human beings have to be restricted and disciplined by society. Societal institutions, on the other hand, incorporate valuable traditions and wisdom, which enable human beings to behave in a civilized way. This logic, however, is an elite construction. Much knowledge of political and historical trends and a high level of conceptual and cognitive sophistication are needed to understand its logic. As a consequence, research shows that only members of the political elite (e.g., members of parliament) are capable of integrating both conservatism dimensions into one single dimension, ranging from conservatism to progressiveness.

This is not the case within the normal population. "Ordinary" citizens distinguish the two forms of conservatism discussed previously. Both forms of conservatism are related to specific groups, reflecting their interests. Low-income groups are prompted toward an economic progressive position, since this view promotes their interests. Groups with a higher level of education are prompted to stress progressive views regarding sociocultural issues, since these views enable them to organize their lives according to their own insights. Because cultural conservatism is linked to the cleavage between the Church and the State, religious respondents tend to stress this form of conservatism to a larger degree than do non-religious people.

## **4. POLITICAL CONSEQUENCES OF IDEOLOGY AND VALUE ORIENTATIONS**

Ideology and values have important consequences for political behaviors of citizens. Before one can act at the political level, however, one must feel motivated to do so. Political interest and political efficacy play an important role in translating ideology and values into political action. Comparative research shows post-materialism and (to a lesser extent) cultural progressiveness to be associated with political interest. A similar finding applies to political efficacy: the self-image that one is an effective and influential participant in politics. Efficacious citizens see themselves

as comprehending, controlling, and mastering their political environments. People holding post-materialist values perceive themselves to be more efficacious and subjectively competent. This of course strengthens their wish to participate in the political process. Also, cultural progressiveness enhances people's sense of civic competence and political efficacy. However, the net impact of value orientations on political interest and political efficacy is modest, and other variables, such as the level of education, play a more direct and important role.

Given the results mentioned above, it comes as no surprise that post-materialism (and, to a lesser extent, cultural progressiveness) is found to influence different forms of unconventional protest activity, such as signing petitions, attending lawful demonstrations, and participating in boycotts or occupations of buildings. The motivating force of ideology is not limited to these unconventional forms of activity. Research shows that trade union participation is also linked to ideology and values. Economic progressiveness in particular seems relevant in this context, because it is associated with participation that requires a high level of effort, such as becoming a union activist or participation in industrial action (such as strikes). However, even though informal political activities became important in many countries, voting is still the most frequent mode of political participation. Interestingly, research shows that post-materialist values have little impact on electoral turnout levels.

The impact of ideology and values is much more substantial when one analyzes party choice. Here, ideological orientations play a rather important role in distinguishing the electorates of different parties. In analyzing the Dutch vote, Middendorp plotted all parties according to the mean score of cultural and economic conservatism of their voters. This results in an ordering of all parties on the economic dimension (the "left-right" vote) and on the cultural one (the "libertarian-authoritarian" vote). He concluded that the left-right vote is mostly explained by economic conservatism, social class, and income. The libertarian-authoritarian vote was mainly explained by cultural conservatism and religious involvement. A more encompassing analysis of different countries shows that these findings can be generalized across Europe. Socialist and communist parties (e.g., "leftist" parties) have the most economic progressive electorates. The electorates of the (economic conservative, "rightist") liberal parties are the most rightist in each country in continental Europe. Christian democratic voters tend to be centrists, leaning toward the center right of

economic conservatism. Green parties and left-wing socialist parties have the most post-materialist electorates in most countries. Conservative and Christian democratic electorates in most countries are predominantly materialist. In this comparative research, cultural conservatism was not included, so no conclusions regarding this ideology could be reached. Other research, however, suggests that this ideological dimension too is relevant in determining the political vote.

Interestingly, different scholars find that ideological dimensions explain more variance than background characteristics do. The discriminating power of variables such as social class and religion tends to decline somewhat over time, whereas the impact of values on voting intentions seems rather stable in most countries or increases slightly when compared to a more distant past. Some consider this finding an indication of the "ideologization" of the vote: ideology seems to determine the vote to a larger extent in recent elections than in prior ones. Other scholars suggest that instead we are witnessing the rise of issue voting. According to issue salience theory, voting is determined by those value orientations and issues that dominate the political agenda and party conflicts in given political systems and at a specific time. Because these issues are variable, the vote becomes more unpredictable than before, which explains the higher proportion of volatile voters, changing their party preference in between elections. Issues, however, are not unrelated to ideology or value orientations, but are embedded in more encompassing ideological orientations. This means that ideology is still relevant in issue salience theory, even though particular political circumstances might single out specific aspects of this ideology. A good example of this trend is the rise of extreme-right wing parties in Europe. Research consistently shows that these parties appeal to the negative attitude of citizens towards foreigners. Extreme right-wing parties have been successful in putting their anti-immigrant stands on the political agenda, and successfully persuaded the electorate that they will deal with this topic once elected. Negative attitudes toward foreigners are not an isolated domain, however, but are firmly rooted in cultural conservatism. The rise of extreme-right wing parties in Europe thus suggests that at least a part of this kind of conservatism grew in importance as predictor of voting behavior, compared to the past.

### See Also the Following Articles

Political Psychology, Overview ■ Values and Culture

### Further Reading

- Borg, S. (1995). Electoral participation. In J. W. Van Deth, & E. Scarbrough (Eds.), *The impact of values* (pp. 441–460). Oxford: Oxford University Press.
- Braithwaite, V. (1994). Beyond Rokeach's equality-freedom model: Two-dimensional values in a one-dimensional world. *Journal of Social Issues*, 50, 67–94.
- De Witte, H. (1996). Are trade union members (still) motivated by ideology? A review of the importance of ideological factors as determinants of trade union participation in (the Flemish part of) Belgium. In P. Pasture, J. Verberckmoes, & H. De Witte (Eds.), *The lost perspective? Trade unions between ideology and social action in the New Europe. Volume 2. Significance of ideology in European trade unionism* (pp. 275–304). Aldershot: Avebury.
- De Witte, H., & Scheepers, P. (Eds.). (1999). *Ideology in the low countries. Trends, models and lacunae*. Assen: Van Gorcum.
- Evans, G., Heath, A., & Lalljee, M. (1996). Measuring left-right and libertarianism-authoritarian values in the British electorate. *British Journal of Sociology*, 47, 93–112.
- Eysenck, H. J. (1972). *Psychology is about people*. London: Penguin Press.
- Gabriel, O. W. (1995). Political efficacy and trust. In J. W. Van Deth, & E. Scarbrough (Eds.), *The impact of values* (pp. 357–389). Oxford: Oxford University Press.
- Gabriel, O. W., & Van Deth, J. W. (1995). Political interest. In J. W. Van Deth, & E. Scarbrough (Eds.), *The impact of values* (pp. 390–411). Oxford: Oxford University Press.
- Grunberg, G., & Schweisguth, E. (1993). Social libertarianism and economic liberalism. In D. Boy, & N. Mayer (Eds.), *The French voter decides* (pp. 45–64). Ann Arbor, MI: The University of Michigan Press.
- Gundelach, P. (1995). Grass-roots activity. In J. W. Van Deth, & E. Scarbrough (Eds.), *The impact of values* (pp. 412–440). Oxford: Oxford University Press.
- Inglehart, R. (1977). *The silent revolution. Changing values and political styles among Western publics*. Princeton: Princeton University Press.
- Inglehart, R. (1990). *Culture shift in advanced industrial society*. Princeton, NJ: Princeton University Press.
- Inglehart, R. (1997). *Modernization and postmodernization*. Princeton, NJ: Princeton University Press.
- Johnson, S., & Tamney, J. (2001). Social traditionalism and economic conservatism: two conservative political ideologies in the United States. *The Journal of Social Psychology*, 141, 233–243.
- Knutsen, O. (1995). Party choice. In J. W. Van Deth, & E. Scarbrough (Eds.), *The impact of values* (pp. 461–491). Oxford: Oxford University Press.
- Kohn, M. L. (1977). *Class and conformity: a study in values. With a Reassessment (2nd ed.)*. Chicago, IL: The University of Chicago Press.
- Kohn, M. L. (1995). Social structure and personality through time and space. In P. Moen, G. Elder, et al. (Eds.), *Examining lives in context: Perspectives on the ecology of human development* (pp. 141–168). Washington, DC: American Psychological Association.
- Kohn, M. L., & Schooler, C. (1983). *Work and personality: An inquiry into the impact of social stratification*. Norwood, NJ: Ablex Publishing Corporation.
- Lipset, S. M. (1981). *Political man, the social basis of politics*. Baltimore, MA: Johns Hopkins University Press.
- Lubbers, M., Gijsberts, M., & Scheepers, P. (2002). Extreme right-wing voting in Western Europe. *European Journal of Political Research*, 41, 345–378.
- Middendorp, C. (1978). *Progressiveness and conservatism. The fundamental dimensions of ideological controversy and their relationship to social class*. The Hague/Paris/New York: Mouton.
- Middendorp, C. (1991). *Ideology in Dutch politics: The democratic system reconsidered 1970–1985*. Assen: Van Gorcum.
- Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.
- Rokeach, M. (1979). *Understanding human values: Individual and societal*. London: Collier MacMillan.
- Scheepers, P., Lammers, J., & Peters, J. (1994). Religious and class voting in The Netherlands 1990–1991: A review of recent contributions tested. *The Netherlands' Journal of Social Sciences*, 30, 5–24.
- Schwartz, S. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology*, Vol. 25 (pp. 1–65). New York: Academic Press.
- Schwartz, S. (1994). Are there universal aspects in the structure and content of human values? *Journal of Social Issues*, 50, 199–45.
- Triandis, H. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
- Van Deth, J. W., & Scarbrough, E. (Eds.). (1995). *The impact of values*. Oxford: Oxford University Press.
- Van Gyes, G., & De Witte, H. (1999). Conservatism and values. A confrontation of Middendorp with Schwartz and Triandis. In H. De Witte, & P. Scheepers (Eds.), *Ideology in the low countries. Trends, models and lacunae* (pp. 121–150). Assen: Van Gorcum.
- Wilson, G. (1979). *Manual for the Wilson-Patterson Attitude Inventory*. Windsor: NFER.



# Indecision, Vocational

Paul J. Hartung

Northeastern Ohio Universities College of Medicine, Rootstown, Ohio, USA

1. Perspectives on Indecision
2. A Brief Historical Synopsis: From Plato to Parsons to the Present
3. Vocational Indecision: The Positivist Perspective
4. Vocational Indecision: The Phenomenological Perspective
5. Future Directions  
Further Reading

## GLOSSARY

**indecisiveness** A global problem in making life decisions, trait anxiety associated with decision making, and chronic difficulty in resolving the decisional process.

**logical positivism** A philosophy of science that advances use of the scientific method to produce knowledge about what are believed to be discoverable facts in the world; it emphasizes objectivity and empirical data.

**phenomenology** A philosophy of science that views knowledge and meaning as derived from within the individual interpreter of experiences, events, and observations; it emphasizes subjectivity and intuitive meaning.

**vocational indecision** The inability to express an educational or occupational choice when asked to do so; it involves a delay in bringing closure to the career choice process, that is, a wavering, pause, or hesitation in vocational development and an openness to alternative career pathways.

Vocational indecision remains a core and complex construct in contemporary career psychology, linked intimately to the career choice and decision-making

process. It also persists among the most common concerns presented in career counseling and has served as the focus of a vast literature of empirical research and conceptual analysis. Historically, vocational indecision has denoted an inability to express an educational or occupational choice when asked to do so and a delay in bringing closure to the career choice process. Contemporary definitions further characterize indecision as a wavering, pause, or hesitation in vocational development and an openness to alternative career pathways. A synthesis of traditional and alternative interpretations enriches the meaning and the management of vocational indecision.

## 1. PERSPECTIVES ON INDECISION

A scheme useful for both elaborating on and dealing with the complexity of vocational indecision is to consolidate two principal perspectives on the construct that have emerged during the course of comprehending and intervening to address it. One perspective perceives indecision as an objective circumstance, whereas the other perspective views indecision as a subjective experience. An encapsulated history of career decision making, and indecision as a component of the decision-making process, provides a broader context for discussing these two perspectives in turn.

## **2. A BRIEF HISTORICAL SYNOPSIS: FROM PLATO TO PARSONS TO THE PRESENT**

Inquiry into career decision making in the Western world dates back to the Greek philosopher Plato, who assisted individuals with their vocational decisions by proffering wisdom, advice, and recommendations about fitting work. In the East, teachings in the Tao Te Ching have long advised individuals to do what they enjoy in work. Before the dawn of the early 20th-century vocational guidance movement, other people (e.g., Lysander Richards) used quasi-scientific methods of graphology, physiognomy, palmistry, and phrenology to identify suitable work for people.

With the establishment in 1909 of the Vocation Bureau in Boston by Frank Parsons and his colleagues came an approach to understanding and alleviating vocational indecision grounded in logical analysis and scientific inquiry. By advocating the use of the scientific method, Parsons championed the positivist perspective on vocational indecision that predominated throughout the 1900s and continues to hold great sway in vocational psychology and career counseling today. An alternative perspective on vocational indecision, initiated by D. E. Super and his thematic extrapolation method at midcentury, advanced by D. V. Tiedeman and his colleagues during the 1960s, and propelled by prevailing constructivist and narrative approaches to career decision making, uses phenomenology to comprehend the individual experiences of decision makers. The positivist and phenomenological perspectives both contribute uniquely to comprehending and coping with vocational indecision.

### **3. VOCATIONAL INDECISION: THE POSITIVIST PERSPECTIVE**

The traditional perspective on vocational indecision derives from logical positivism and views indecision as an objective phenomenon. It prizes the use of psychometric scales to yield scores that indicate degree and type of vocational indecision. Three distinct phases of this perspective spanning the past century have transformed indecision from a simple dichotomous concept into a complex multidimensional one.

### **3.1. A Dichotomy**

Parsons initially defined indecision for vocational psychology as a dichotomy consisting of two groups. One group contained individuals decided about their educational or vocational plans, whereas the other group contained persons undecided about their educational or vocational plans. This rudimentary scheme (a) framed indecision as a problem, (b) sparked 40 years of empirical inquiry into personality and situational differences between members of these two groups, (c) characterized undecided persons as deficient if not pathological, and (d) engendered counseling efforts to treat the personality defect underlying the indecision. With the advent of the developmental perspective on careers during the 1950s, some theorists began to advance a view of indecision as a normal component of human development from which a career-decided state would emerge. This view preceded contemporary perspectives on indecision as a healthy state of adaptive uncertainty, openness to alternative career possibilities, and flexibility in coping with change and chance events.

### **3.2. A Unidimensional Continuum**

Research based on a dichotomous concept of indecision produced mixed and inconsistent findings, with many studies supporting differences between decided and undecided individuals and many other studies indicating no differences between these groups. Two approaches, one psychometric and the other theoretical, were taken to address these inconsistencies. The psychometric approach sought to better measure career indecision and career decision-making status. The theoretical approach endeavored to better conceptualize what career indecision means.

Taking a psychometric approach, researchers during the 1970s turned to a view of indecision as a unidimensional continuum ranging from undecided to decided. Psychometric scales, such as the Vocational Decision-Making Difficulties Scale and the Career Decision Scale, were subsequently constructed to index positions on this continuum. So conceived, indecision reflected a continuous process operationally defined by scale scores indicating level of decidedness and difficulties that thwart choice.

### **3.3. A Multidimensional Construct**

Advancing the psychometric approach to indecision, the objective perspective has evolved to the current

view of indecision as a multidimensional concept with distinct and measurable subtypes. Rather than perceiving all undecided persons as undecided for the same reasons, the multidimensional perspective recognizes heterogeneous undecided subgroups. Researchers consequently turned to either revising existing scales or developing new measures to assess degree of and reasons for indecision, which would then guide intervention. These measures now constitute two generations of instrument development. First-generation measures operate on a unidimensional definition of indecision and view groups of individuals with equivalent overall scale scores as homogenous. Second-generation measures construe indecision as multidimensional and seek to identify subgroups of undecided individuals and provide different interventions for each type.

### **3.4. Indecision Versus Indecisiveness**

During the 1960s, a theoretical approach to resolving inconsistent research findings about differences between decided and undecided groups centered on distinguishing indecision from indecisiveness. Leona Tyler proposed that indecisiveness reflects a global problem in making life decisions, trait anxiety associated with decision making, and chronic difficulty in resolving the decisional process. Indecision, in contrast, reflects a specific problem to make an occupational choice, a state of situational anxiety about decision making, and a developmental delay in resolving the decisional process. This view on indecision has been propelled by subsequent theorizing about indecision as a natural component of career development, as an opportunity to explore, and as open-mindedness to random events and opportunities in shaping career-related decisions. The intuitively reasonable distinction between indecision and indecisiveness as differentially related to anxiety has met with little empirical support.

### **4. VOCATIONAL INDECISION: THE PHENOMENOLOGICAL PERSPECTIVE**

Phenomenology offers an alternative to the positivist perspective on vocational indecision. Situated in constructivist and narrative approaches to career that have emerged more prominently during the past decade or

so, the phenomenological view conceives of indecision as a subjective experience. Consequently, rather than using psychometric inventories and objective scores to measure level and type of indecision, the subjective perspective prizes the use of narratives and stories to reveal life themes, patterns, projects, and goals that guide movement within the choice process. From the phenomenological perspective, indecision reflects a wavering or pause in vocational development and an openness to alternative career pathways. Individuals can use indecision to contemplate their directions in life, construct possible futures, and convert their identities into truer self-representations.

The subjective perspective on indecision emphasizes the personal meaning ascribed to the lived experience of being undecided. Exploring undecided individuals' subjective, qualitative, and personal realities leads to comprehending their unique experiences of indecision, preoccupations and intentions, and possibilities for constructing their future lives. By eliciting narratives, reviewing early recollections and family stories, and situating work in the context of people's lives, phenomenologically based approaches involve a collaborative endeavor to imbue indecision with personal meaning. Vocational decision making becomes a process of untangling a present filled with uncertainty and doubt to clarify and envision a future path that will provide purpose, direction, fulfillment, and satisfaction.

### **5. FUTURE DIRECTIONS**

Vocational indecision has long been a focus of theory, research, and practice in vocational psychology. Decades of scholarship have examined indecision both as an objective circumstance, with distinct measurable subtypes, and as a subjective experience, wholly unique to the individual who experiences it. Future efforts to meld positivist and phenomenological perspectives on indecision will likely yield a more comprehensive approach to comprehending the career decision-making process and to dealing with vocational indecision as a problem that can interrupt that process. Emerging perspectives on vocational indecision that argue against decidedness as the optimal vocational outcome in favor of blending planfulness, flexibility, and more active engagement with random opportunities for shaping future decisions will also expand possibilities and potential for human development through work and career.

### See Also the Following Articles

Decision Making ■ Holland's Theory (Vocational Personality Types)

### Further Reading

- Cochran, L. (1997). *Career counseling: A narrative approach*. Thousand Oaks, CA: Sage.
- Gati, I., & Asher, I. (2001). The PIC model for career decision making: Prescreening, in-depth exploration, and choice. In F. T. L. Leong, & A. Barak (Eds.), *Contemporary models in vocational psychology: A volume in honor of Samuel H. Osipow* (pp. 7–54). Mahwah, NJ: Lawrence Erlbaum.
- Gelatt, H. (1989). Positive uncertainty: A new decision-making framework for counseling. *Journal of Counseling Psychology*, 36, 252–256.
- Grites, T. J. (1981). Being “undecided” might be the best decision they could make. *The School Counselor*, 29, 41–46.
- Hartung, P. J., & Blustein, D. L. (2002). Reason, intuition, and social justice: Elaborating on Parsons's career decision-making model. *Journal of Counseling and Development*, 80, 41–47.
- Krieshok, T. S. (1998). An anti-introspectivist view of career decision making. *Career Development Quarterly*, 46, 210–229.
- Krieshok, T. S. (2001). How the decision-making literature might inform career center practice. *Journal of Career Development*, 27, 207–216.
- Krumboltz, J. D. (1992). The wisdom of indecision. *Journal of Vocational Behavior*, 41, 239–244.
- Krumboltz, J. D. (1998). Serendipity is not serendipitous. *Journal of Counseling Psychology*, 45, 390–392.
- Mitchell, K. E., Levin, A. S., & Krumboltz, J. D. (1999). Planned happenstance: Constructing unexpected career opportunities. *Journal of Counseling and Development*, 77, 115–124.
- Phillips, S. D. (1997). Toward an expanded definition of adaptive decision making. *Career Development Quarterly*, 45, 275–287.
- Phillips, S. D., & Paziienza, N. J. (1988). History and theory of the assessment of career development and decision making. In W. B. Walsh, & S. H. Osipow (Eds.), *Career decision making* (pp. 1–31). Hillsdale, NJ: Lawrence Erlbaum.
- Savickas, M. L. (1995). Constructivist counseling for career indecision. *Career Development Quarterly*, 4, 363–373.
- Savickas, M. L. (2000). Assessing career decision making. In C. E. Watkins, & V. Campbell (Eds.), *Testing and assessment in counseling practice* (2nd ed., pp. 429–477). Mahwah, NJ: Lawrence Erlbaum.
- Slaney, R. B. (1988). The assessment of career decision making. In W. B. Walsh, & S. H. Osipow (Eds.), *Career decision making* (pp. 33–76). Hillsdale, NJ: Lawrence Erlbaum.
- Tinsley, H. E. (Ed.). (1992). Career decision making and career indecision [special issue]. *Journal of Vocational Behavior*, 41(3).



# Indigenous Psychologies

Uichol Kim

Chung-Ang University, Seoul, Korea

Young-Shin Park

Inha University, Incheon, Korea

1. Historical Background
  2. What Is Indigenous Psychology?
  3. Two Approaches in Indigenous Psychology
  4. How Is Research Done?
  5. An Example: Educational Attainment
  6. Future Directions
- Further Reading

*phenomenological knowledge* Internal thoughts, feelings, and motives of a person.

*practical validity* Theories that can help one to understand, predict, and manage the world.

*transactional model* A scientific model in which phenomenological knowledge is systematically linked with the situation or event, on the one hand, and with behavior or performance, on the other.

## GLOSSARY

*amae* A Japanese term referring to the act of asking for and receiving special favors in close relationships.

*chong* A Korean term meaning deep affection and attachment for a person, place, or thing.

*extrinsic motivation* A factor outside an individual (e.g., parental influence, social support, social reward) that can motivate a person into action.

*guanxi* A Chinese term meaning relationship.

*indigenization from within* Developing theories, concepts, and methods internally, with indigenous information being considered as a primary source of knowledge.

*indigenization from without* Taking existing psychological theories, concepts, and methods and modifying them to fit the local cultural context.

*indigenous psychology* According to Kim and Berry, “the scientific study of human behavior [or mind] that is native, that is not transported from other regions, and that is designed for its people.”

*intrinsic motivation* A factor within an individual (e.g., personal interest, preference, ability) that can motivate a person into action.

Indigenous psychology represents the cultural psychology tradition espoused by Wundt that examines human action by linking people’s intentions, skills, and goals with the situation or event, on the one hand, and with behavior or performance, on the other. It advocates studying people in their natural contexts, using concepts that are meaningful, and investigating the skills that people develop to manage their lives. The goal of indigenous psychology is to create more rigorous scientific theories that can be verified and applied in the natural setting. The phenomenal educational achievement of East Asian students is provided as an example to highlight this approach.

## 1. HISTORICAL BACKGROUND

Wundt is considered to be the “father of modern psychology.” In establishing psychology as an independent scientific discipline, he recognized two traditions in science: the natural sciences tradition and the

cultural sciences tradition. Wundt was instrumental in establishing the experimental method to study basic physiological, perceptual, and cognitive processes. However, he recognized the limitations of experimental method and emphasized the need to develop cultural psychology. He observed that thinking is heavily conditioned by language and culture. He regarded cultural psychology as a more important aspect of psychological research. General psychology represents the natural sciences approach, and indigenous psychology represents the cultural science tradition.

When psychology became established in North America, psychologists adopted the natural science approach, and eliminated the cultural approach, in search of abstract universal laws of human behavior. Behaviorism emerged as the dominant paradigm, and the subject matter of psychology became tailored to fit into the narrow definition of natural science. Early psychologists attempted to eliminate subjective aspects (e.g., consciousness, intentions, meaning) and to conduct controlled experiments in laboratory settings. Psychologists assumed that the results of laboratory studies could be generalized to explain complex human behavior in the natural setting.

Indigenous psychologies evolved around the world as a reaction against unjustified claims of universality of psychological theory. Although existing psychological theories and concepts are assumed to be objective, value free, and universal, in reality they are deeply enmeshed with Euro-American values that champion rational, liberal, individualistic ideals. Indigenous psychology represents an alternative scientific paradigm in which the human qualities (e.g., meaning, goals, creativity) and culture have been integrated as central elements of research design.

## **2. WHAT IS INDIGENOUS PSYCHOLOGY?**

In 1993, Kim and Berry defined indigenous psychology as “the scientific study of human behavior [or mind] that is native, that is not transported from other regions, and that is designed for its people.” Indigenous psychology advocates examining knowledge, skills, and beliefs that people have about themselves and studying these aspects in people’s natural contexts. It represents a bottom-up approach in which the goal of psychology is to understand how people function in their natural contexts and to examine how people interact with their natural and

human worlds. Indigenous psychology recognizes that human psychology is complex, dynamic, and generative, and it recognizes the importance of agency, meaning, intentions, goals, and context in understanding behavior. Theories, concepts, and methods are developed to correspond with psychological phenomena. The goal is to create a more rigorous, systematic, universal science that can be verified both theoretically and empirically. In addition, there are seven facets of indigenous psychology that can be identified.

First, indigenous psychology emphasizes contextualized understanding rooted in a particular setting (e.g., ecological, political, historical, cultural). It emphasizes the discovery and use of natural taxonomies in search of regularities, general principles, and universal laws. It examines how people view themselves, relate to others, and manage their environment.

Second, contrary to popular misconception, indigenous psychologies are not studies of native peoples, ethnic groups, or people living in distant lands. Indigenous research has often been equated with anthropological analysis of “exotic” people living in distant lands. Indigenous psychology is needed for all cultural, native, and ethnic groups, including economically developed countries such as Canada, France, and the United States.

Third, acceptance of indigenous psychology does not affirm or preclude the use of a particular method. Indigenous psychology is part of the scientific tradition where an important aspect of the scientific endeavor is the discovery of appropriate methods for the phenomenon under investigation. Scientists should not, and cannot, be bound to a particular method. The use of multiple methods is recommended to increase one’s confidence that a particular finding is valid and not an artifact of the research method. Results from multiple methods should be integrated to provide a more comprehensive and robust understanding of psychological phenomena.

Fourth, it has been assumed that insiders have a better understanding of indigenous phenomena and that outsiders can have only a limited understanding. Although a person who has been born and raised in a particular community may have insights and understanding of indigenous phenomena, this might not be true in all instances. As Wirth observed in 1946, “The most important thing . . . that we can know about a person is what he [or she] takes for granted, and the most elemental and important facts about a society are those things that are seldom debated and generally regarded as settled.” An outsider, with an external point of view, can call attention to the fact that what is assumed to be natural is

actually cultural. Although an outsider may have a superficial understanding of indigenous phenomena found in other cultures, he or she may point out peculiarities, inconsistencies, and “blind spots” that insiders may have overlooked. Both internal and external points of view are necessary in providing a comprehensive and integrated understanding of psychological phenomena.

Fifth, many indigenous psychologists search philosophical and religious texts for explanations of indigenous phenomena. Too often, they use philosophical treatises (e.g., Confucian Classics) or religious texts (e.g., Koran, Vedas) as explanations of psychological phenomena. However, these texts were developed for philosophical and religious purposes several thousand years ago and not for psychological research in the 21st century. To use these texts, one must first translate them into psychological concepts and empirically verify how they influence the ways in which people think, feel, and behave. One cannot assume that because a person is Chinese, he or she will automatically follow the Confucian way, nor can one assume that Hindu Dharma can explain the behavior of an individual just because he or she is an Indian.

Sixth, as with other scientific traditions, one of the goals of indigenous psychology is the discovery of universal facts, principles, and laws that could explain human diversity. However, psychological universals must be theoretically and empirically verified rather than assumed *a priori*.

Seventh, human consciousness, agency, meaning, and goals are considered to be central constructs in explaining human behavior (or mind). The separation of the subjective and objective worlds that is important in the natural sciences is not possible in the human world because people are both the subject and the object of investigation. Researchers must recognize that research is a value-laden enterprise in search of a probabilistic understanding of human action rather than a search for objective and mechanical knowledge. Research topics and stimuli must be meaningful and contextualized rather than arbitrarily selected, manipulated, and controlled.

### 3. TWO APPROACHES IN INDIGENOUS PSYCHOLOGY

In 1993, Enriquez identified two approaches in indigenous psychology: indigenization from without and indigenization from within. Indigenization from without involves taking existing psychological theories, concepts, and methods and modifying them to fit the

local cultural context. The approaches advocated by cultural and cross-cultural psychologists are examples of indigenization from without. In this approach, rather than assuming that a particular theory is universal *a priori*, researchers modify and adapt psychological theories and integrate them with the local cultural knowledge. Those aspects that can be verified across cultures are retained as possible cultural universals. Existing theories in cognitive, developmental, social, and organizational psychology have been modified and extended by indigenous research.

In the indigenization from within approach, theories, concepts, and methods are developed internally and indigenous information is considered to be a primary source of knowledge. For example, one of the core values and assumptions that psychologists from East Asia question is the emphasis on individualism. In East Asia, the relatedness of one person to another person is considered to be fundamental. In East Asia, the word for human being can be literally translated as “human between.” In other words, it is what happens between individuals that makes them human. In Chinese culture, the concept of *guanxi* (literally “relationship”) occupies center stage, defining interactions in family, school, companies, and society. In Japan, the concept of *amae* (literally “the act of asking and receiving special favors in close relationships”) helps to define and maintain close interpersonal relationships. In Korea, the concept of *chong* (literally “deep affection and attachment for a person, place, or thing”) is a central emotion that binds family, friends, and colleagues together. Recently, both cross-sectional and longitudinal studies yielded results that confirm the importance of close relationships that are highly reliable, valid, and applicable in Korea.

### 4. HOW IS RESEARCH DONE?

Indigenous psychology advocates development and use of methods that are appropriate for psychological phenomena. It recognizes the existence of two types of knowledge: objective third-person knowledge (i.e., analytical, semantic, and declarative knowledge) and subjective first-person knowledge (i.e., phenomenological, episodic, and procedural knowledge). Analytical, semantic, and declarative knowledge represents information based on objective and impartial analysis. Phenomenological, episodic, and procedural knowledge represents experiential knowledge that people have about themselves and ways in which to deal with their physical and human worlds.

In natural sciences, the goal is to obtain objective knowledge through rigorous experimentation, observation, and mathematical analysis. In natural sciences, phenomenological knowledge (e.g., the thoughts, feelings, and motives of a tree, lion, or star) represents irrelevant research questions. For example, if a lion kills a deer, one assumes that the behavior is caused by, and explainable in terms of, instinct. However, if a woman kills a man, one does not simply attribute the killing to her instinct. One seeks additional information such as motive, agency, and responsibility. The same act (i.e., a woman killing a man) has different meanings and consequences based on one's knowledge of the person's motive and agency. If a woman killed her husband to collect an inheritance, it is premeditated murder. If a woman found her husband in bed with another woman and killed him out of rage, it can be classified as manslaughter. If a woman killed a man while she was driving drunk, it can be classified as involuntary manslaughter. If a woman killed a man while resisting being raped, it can be classified as self-defense. If a 5-year-old girl kills her friend while playing with her father's gun, it can be classified as gross negligence on the part of the father. Based on one's assessment of the person's intentions and responsibility, one assigns different outcomes, ranging from capital punishment (in the case of premeditated murder), to 10 years imprisonment (in the case of manslaughter), to no punishment (in the case of self-defense).

In human sciences, unlike in physical and biological sciences, it is essential to know the intentions, goals, and motives of people. This information can be obtained through self-reports. In the transactional model espoused by both Bandura and Kim, the subjective human qualities (e.g., agency, intentions, meaning, goals) can be systematically linked with the situation or event, on the one hand, and with behavior or performance, on the other.

In research, phenomenological, episodic, and procedural knowledge needs to be translated into analytical knowledge. However, many people do not possess the analytical skills in describing how they performed certain actions. For example, adult native English speakers can freely express their thoughts in English (i.e., procedural knowledge), but they might not know the grammatical syntax or structure of the spoken words (i.e., semantic knowledge). In other words, they know how to produce the sentences, but they might lack the ability to describe them analytically. This is the case because people are taught to use the language to communicate with others. Grammatical knowledge is taught as a part

of formal education in school. Conversely, Koreans might know the grammar of English, but they might have difficulty in speaking English because it is a different type of knowledge. It is the role of the researcher to analyze the grammar of psychological phenomena and to verify it to ensure that the analysis is correct. In human life, both experiential knowledge (e.g., a football player describing his experiences in playing a game) and analytical knowledge (e.g., a sport commentator providing a play-by-play analysis) are useful types of information that need to be integrated (e.g., a coach planning strategy for the next game). Indigenous psychological analysis has been conducted in the areas of achievement, conception of self, parent-child relationships, school violence, stress, occupational safety, trust, organizational behavior, and political culture. An example of educational achievement of East Asian students, focusing on Korean results, is provided in the next section.

## 5. AN EXAMPLE: EDUCATIONAL ATTAINMENT

In international comparisons of academic literacy of high school students (e.g., Third International Mathematics and Science Study [TIMSS], Program for International Student Assessment [PISA]), East Asian students are the top achievers in mathematics and science. Although the U.S. government spends more money per student to educate its children and adolescents, American schools have smaller classes, and many psychologists provide their expertise in developing U.S. educational programs, American students perform far below their East Asian counterparts. In the TIMSS, American students were ranked 19th in mathematics and 18th in science; in the PISA, they were ranked 15th in reading literacy, 19th in mathematical literacy, and 14th in scientific literacy. In contrast, East Asian students were ranked at the top in mathematics and science in both the TIMSS and PISA and were ranked near the top in reading literacy (Korea was ranked 6th and Japan was ranked 8th). These results baffle many psychologists because the results directly contradict existing psychological theories.

More than a century ago, psychological testing and assessment was initiated by Galton to measure innate ability. American psychologists (e.g., Terman at Stanford University, Yerkes at Harvard University, Garrett at Columbia University) continued the tradition. Terman developed the Stanford-Binet intelligence quotient (IQ)

test to measure inborn intelligence and to document individual, ethnic, racial, and sex differences. The results of the test were used to shape U.S. national policies (e.g., forced sterilization of “feebleminded” individuals, segregation of races, restricting Asian immigration). In 1924, Asians were considered a “genetically inferior” race, and the passage of the National Origins Act barred Asians from immigrating to the United States until 1965. Also, the IQ test scores were used to justify racial and sex segregation.

In 1977, Kamin stated that when the first version of the Stanford–Binet test was administered and published in 1916, one set of results contradicted the expectations in that girls of all ages outscored boys by an average of 2 to 4%. Terman unilaterally deleted, revised, or added new items so that this difference disappeared, and subsequently boys did better than girls. Thus, the IQ test was developed and revised not on a strict scientific basis but rather to fit the preconceptions at the time.

In 1960, Korea had one of the lowest literacy and educational achievement levels in the world. Currently, Korean students receive one of the highest scores in mathematics, science, and reading literacy. These dramatic reversals in sex and national differences cannot be attributed to genetic basis but can be attributed to social, economic, and cultural factors. Existing psychological theories argue that innate born differences (e.g., IQ, personality), ability attribution, intrinsic motivation, and self-esteem are important factors in predicting educational achievement. However, a series of studies conducted by Stevenson and colleagues and by Kim and colleagues challenged the basic assumptions of these theories.

The success of East Asian students can be traced to the strong sense of relatedness; social support provided by parents; belief in effort, hard work, and self-discipline; and coordinated efforts of parents, teachers, and governments. First, students, parents, teachers, and administrators agree that educational success is the most important goal for students, and they work together in achieving this outcome. Second, East Asians believe in effort, discipline, and persistence as a means to their goals. People believe that if they work hard and discipline themselves, they can accomplish their desired objectives. Third, the sacrifice and support provided by parents are essential ingredients in their children’s educational success. Emotional support in the form of encouragement, praise, security, and understanding, rather than expert advice from professionals, is reported as an important factor for their success. Fourth, East Asians believe that ability can be acquired and personality can be polished through persistent effort rather than believing that these characteristics are innate.

When self-concept in mathematics and science was assessed, Korean students were ranked 32nd and 21st, respectively, whereas U.S. students were ranked 1st and 4th, respectively (TIMSS). Korean students did not believe that they were good in mathematics and believed that they needed to work hard to do well, and this belief and persistent effort led to the attainment of high scores. In contrast, American students believed that they were good in math and did not feel the need to study hard because expending effort implied a lack of natural ability. As a result, they did not do well on these tests. In terms of motivation, when Korean students were asked about their reasons for studying mathematics, 85% indicated that attending a desired university was among their reasons, 62% cited pleasing their parents, and 44% cited obtaining a desired job. In other words, their reasons for studying mathematics were largely extrinsic motivation.

Finally, using the indigenous psychology approach, a 6-year longitudinal study was conducted to examine the factors that influence academic achievement of Korean students. The role of parents emerged as the most important factor, followed by the role of teachers and the role of friends. The current authors found, in a structural equation model, that influence from parents (e.g., social support, achievement pressure, expectations) had a direct effect on students’ achievement motivation, efficacy beliefs, and studying time and, in turn, had both a direct and an indirect effect on academic grades. Moreover, a feeling of indebtedness and respect toward their parents increased students’ achievement motivation and efficacy beliefs and, in turn, increased the amount of time they studied and improved their academic grades. The socioeconomic status of the parents affected how much support they provided to their children. These results point to the role of external factors (e.g., socioeconomic status of parents, the role of significant others) and the belief in effort as being responsible for the phenomenal academic achievement of Korean students. Individual factors (e.g., self-esteem, intelligence, personality, intrinsic motivation, ability attribution) had very little impact on academic achievement.

## 6. FUTURE DIRECTIONS

Although science can provide the most accurate understanding of the world, it can also blind and limit one’s understanding. Researchers start with an idea, model, theory, or method to uncover yet another secret of life. Researchers’ preconceptions can both aid and limit

scientific discovery. As noted previously, psychologists' emphasis and focus on individualistic and intrinsic factors limited their understanding of external factors that are critical to students' academic achievement. A number of external impositions have limited the development of psychology, and indigenous psychology represents being freed of these impositions.

First, psychologists used the natural sciences model to study humans. In the rush to become a respected branch of science, early psychologists tailored the psychological phenomena to fit the natural sciences paradigm. Psychologists were able to achieve a modest degree of methodological sophistication, but central human qualities such as intentions, meaning, and agency have been omitted to distort one's understanding of psychological phenomena. Indigenous psychology advocates the development and use of appropriate methods and concepts to study psychological phenomena.

The second imposition is the assumption of the universality of psychological theories. With very little development, testing, and data, psychological theories are assumed to be universal. This assumption is particularly problematic because most theories were developed in the United States and tested mainly on university students in laboratory settings. In other words, theories that were tested on less than 1% of the total population have been assumed to be universal. Indigenous psychology emphasizes the documentation and analysis of psychological variations before conclusions can be made about their universality.

Third, expert or professional knowledge has been imposed on the lay public. Psychological theories and concepts have been used to explain and modify behavior. In most cases, the predictive value of these theories is very low compared with that of the natural sciences. Psychologists have largely failed to describe psychological phenomena from the inside, that is, from the experiencing persons. Instead, psychologists have dissected the world into abstract concepts and theories that are removed from the "real world" in which people live.

Indigenous psychology advocates being freed of these external impositions and obtaining the phenomenological information as insiders as well as through impartial third-person analysis. Indigenous psychology advocates linking knowledge from humanities (which focus on human experience and creativity) with knowledge from social sciences (which focus on analysis and verification). Most attention has been focused on internal or external validity and not on practical validity. In other words, do the theories help to understand, predict, and manage human behavior? It

is known that most psychological theories (e.g., behaviorist, Freudian, Piagetian, humanist) do not withstand the test of empirical scrutiny. The main reason is that these theories have been developed using limited information (e.g., studying laboratory rats, interviewing too few patients, observing one's children), and they have not developed a process for systematic documentation, understanding, and verification of psychological phenomena.

The people who understand psychological phenomena well might not be psychologists. In fact, perhaps the greatest psychologist was Shakespeare. He was not an analyst like Freud or Piaget, and he did not conduct experiments like Skinner, but Shakespeare was able to capture human drama both on paper and on stage. His dramas have been performed for more than four centuries in many cultures and are loved throughout various parts of the world. In a similar vein, the greatest therapist might have been Beethoven or Mozart, whose music is able to soothe frazzled nerves and the frustrations of daily life. Disney could be considered as the most notable developmental psychologist because he was able to capture the hearts and minds of the young as well as the "young at heart." We might not think of these people as psychologists, but they captured and reproduced human psychology on stage, film, and paper. It is the role of indigenous psychologists to provide an understanding that is both analytical and phenomenological—that is scientific, practical, and relevant to people's daily lives.

### See Also the Following Articles

Educational Achievement and Culture ■ Intrinsic Motivation and Self-Determination ■ Motivation and Culture

### Further Reading

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (1992). *Cross-cultural psychology: Research and applications*. Cambridge, MA: Cambridge University Press.
- Chorover, S. L. (1980). *From genesis to genocide: The meaning of human nature and the power of behavior control*. Cambridge, MA: MIT Press.
- Enriquez, V. G. (1993). Developing a Filipino psychology. In U. Kim, & J. W. Berry (Eds.), *Indigenous psychologies: Research and experience in cultural context* (pp. 152–169). Newbury Park, CA: Sage.
- Kamin, L. J. (1977). *The science and politics of IQ*. New York: Penguin Books.

- Kim, U. (2001). Culture, science, and indigenous psychologies: An integrated analysis. In D. Matsumoto (Ed.), *Handbook of culture and psychology* (pp. 51–76). Oxford, UK: Oxford University Press.
- Kim, U., & Berry, J. W. (1993). *Indigenous psychologies: Experience and research in cultural context*. Newbury Park, CA: Sage.
- Kim, U., Helgesen, G., & Ahn, B. M. (2002). Democracy, trust, and political efficacy: Comparative analysis of Danish and Korean political culture. *Applied Psychology: An International Review*, 51, 317–352.
- Kim, U., & Park, Y. S. (2003). An indigenous analysis of success attribution: Comparison of Korean students and adults. In K. S. Yang, K. K. Hwang, P. Pedersen, & I. Daibo (Eds.), *Progress in Asian social psychology: Conceptual and empirical contributions* (pp. 171–195). New York: Praeger.
- Kim, U., Park, Y. S., & Park, D. H. (1999). The Korean indigenous psychology approach: Theoretical considerations and empirical applications. *Applied Psychology: An International Review*, 45, 55–73.
- Park, Y. S., & Kim, U. (2000). The impact of changing parent-child relationships on adolescents' functioning: Comparison of primary, junior high, senior high, and university students. *Korean Journal of Educational Research*, 38(2), 109–146.
- Park, Y. S., & Kim, U. (2003). Self-efficacy, achievement motivation, and academic achievement of Korean students: Indigenous psychological analysis. *Korean Journal of Educational Psychology*, 17(1), 37–54.
- Sinha, D. (1997). Indigenizing psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology*, Vol. 1: *Theory and method* (pp. 131–169). Boston: Allyn & Bacon.
- Stevenson, H., Azuma, H., & Hakuta, K. (Eds.). (1986). *Child development and education in Japan*. New York: Freeman.
- Stevenson, H., & Lee, S. Y. (1990). Context of achievement: A study of American, Chinese, and Japanese children. *Monographs of the Society for Research in Child Development*, 55, 1–75.
- Wirth, L. (1946). Preface. In K. Mannheim, *Ideology and utopia: An introduction to sociology of knowledge*. New York: Harcourt, Brace.



# Industrial/Organizational Psychology across Cultures

Zeynep Aycan

Koç University, Istanbul, Turkey

1. Introduction
2. The Role of Culture in Industrial and Organizational Psychology
3. Cross-Cultural Approaches to Work Motivation
4. Cross-Cultural Variations in Human Resource Management Practices
5. Managing Cultural Diversity
6. Culture and Work Attitudes  
Further Reading

## GLOSSARY

**culture** Prevailing and shared values, norms, assumptions, belief systems, and behavioral patterns that differentiate one human group from another.

**human resource management practices** The policies and practices involved in carrying out the human resource aspect of a management position, including recruiting, screening, training, rewarding, and appraising.

**motivation** Set of processes that arouse, direct, and maintain human behavior toward attaining some goal.

**organizational diversity** Workforce variety with respect to gender, culture, national origin, handicap, age, and religion.

**work attitudes and behavior** Clusters of feelings, beliefs, and behavioral intentions toward any aspect of work or work settings.

Industrial and organizational psychology (I–O) is one of the major applied fields of psychology that is concerned with behavior in work situations. Because of globalization, there is an increasing interest in the influence of culture on work behavior among scientists and practitioners. In this article, the cross-cultural perspective in I–O psychology is discussed in four main areas. First, the influence of culture on work motivation is examined. Second, human resource management practices in the areas of staffing and performance management are presented. Third, research on managing cultural diversity is reviewed with a focus on expatriate management, women’s career development, and effective functioning of multicultural teams. Finally, cross-cultural research on work attitudes and behavior is summarized in the areas of job satisfaction, organizational commitment, organizational citizenship behavior, and organizational justice.

## 1. INTRODUCTION

Industrial and organizational psychology (I–O) was defined by Blum and Naylor as “the application or extension of psychological facts and principles to the problems concerning human beings operating within the context of business and industry.” During recent years, the interest in cross-cultural perspectives of I–O psychology has

increased, due mainly to changing economic and social conditions such as workforce diversity, availability of telecommunications across cultures, and globalization and global competition. This article provides an overview of the cross-cultural I–O psychology literature in the areas of motivation, workforce diversity, human resource management (HRM), and work attitudes and behavior. Although leadership is an area that has attracted the most cross-cultural I–O research attention, it is not included in this article because a full treatment of the topic is available elsewhere in this encyclopedia.

## **2. THE ROLE OF CULTURE IN INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY**

Organizations are complex systems that operate under the influence of multiple environmental forces that are both internal (e.g., size, type of work, industry and production, type of workforce, technology, stage of development) and external (e.g., political, legal, educational, institutional) to the organization. The challenge of cross-cultural I–O psychology is to disentangle the impact of culture (i.e., sociocultural context) from other internal and external environmental forces.

Since the early 1960s, comparative studies of organizations have focused on noncultural factors influencing organizational structure and practices. Among the most popular noncultural approaches is the “contingency” approach, within which there were four main streams. The first, referred to as the “logic of industrialization,” asserts that industrialization has a homogenizing effect on organizations around the world, irrespective of the political, economic, and cultural contexts. The second thesis, generally referred to as “technological implications,” suggests that technological advancement and automation leads to a transformation of social relations and attitudes at work (e.g., more control over work schedule and work processes, increased emphasis on developing social networks). The “culture-free contingency theory of organizations” approach emphasizes the role of contextual elements such as size of the organization, industry, and dependence on other organizations. The final thesis emphasizes the role of strategic development of organizations, according to which organizations are transformed from small, less hierarchical, and domestic structures to large, complex, professional, and international structures. This transformation has implications for practices such as planning, diversification, and role differentiation.

Another type of noncultural approach stems from the political–economy perspective. In this perspective, organizations in the same sociopolitical systems (e.g., capitalism vs socialism) are assumed to have similar characteristics, especially with respect to organizational objectives, control strategies, and degree of centralization and decision making. The final noncultural approach, namely the “societal effect approach,” takes into account the social context in which organizations operate, with specific emphasis on the educational system, the system of industrial relations, and the role of the state.

Critics of the noncultural approaches are concerned with the deterministic orientation of these approaches as well as their underestimation of the role of culture in explaining organizational phenomena. Some scholars take an interactionist perspective, suggesting that culture influences some aspects of organizational practices more than it does others. For instance, whereas organizational contingencies influence the “formal” characteristics of organizations (e.g., centralization, specialization, span of control), cultural variables influence the “interpersonal” aspects (e.g., power and authority structure, delegation, consultation, communication patterns) or the “organizational processes” (i.e., the way in which organizations function). Others assert that culture has a moderating effect on organizations; that is, even though the contingencies help to determine the organizational structure, culturally driven preferences influence the exercise of choice among alternative structures. For instance, the effect of industrialization on organizations is not homogenous in every country (e.g., the case of Japan). Similarly, within the capitalist system, there is wide variety among organizational and managerial practices. Researchers argue that similarities in socioeconomic systems resulted in similar organizational objectives, but the ways in which these objectives were materialized differed depending on the cultural contexts. The preceding discussion, therefore, underlines the important role that culture plays in understanding organizational phenomena. However, culture’s effect is more salient in some aspects of organizations than in others. The following sections present the state of affairs in those areas of I–O psychology where the impact of culture is significant.

## **3. CROSS-CULTURAL APPROACHES TO WORK MOTIVATION**

Work motivation is one of the most popular research topics in cross-cultural I–O psychology. The “content

theories” of motivation address the question of what motivates employees, whereas the “process theories” attempt to answer the question of how employees are motivated. Studies conducted in both paradigms clearly demonstrate the influence of culture on motivation.

### **3.1. Cross-Cultural Differences in What Motivates Employees at Work: Content Theories**

Cross-cultural research on motivation between the 1970s and the 1990s focused mainly on the comparison of needs and values. The content theories of motivation assert that goal-directed behavior is initiated to satisfy unfulfilled needs, and values are considered the cognitive representation and transformation of needs. Early cross-cultural research tested the validity of various need taxonomies, including those of Maslow, Alderfer, and McClelland. Employees in economically developed and developing countries were found to be motivated by different incentives due to variations in needs. For example, malnutrition was the main cause of low motivation and productivity in Nigeria, whereas job security was most needed, and hence most motivating, in India and Guatemala. Similarly, satisfaction of physiological and safety needs were more salient for cane cutters, whereas affection and esteem needs were most salient for fishermen.

In subsequent studies, it was shown that in collectivist societies, “self-actualization” (i.e., the highest order need) is replaced by “service to the in-group” or actualizing the in-group’s potential (i.e., “collectualization”) as the highest ideal to motivate workers. Also in collectivist cultures, “achievement” has collectivist overtones such as respect for others, fulfilling collective goals, and glorifying established virtues. The conclusion of the cross-cultural research on need-based theories of motivation is that the structure of needs may be universal (e.g., lower order needs such as safety, higher order needs such as esteem) but that the relative strength or the contents of the standards vary by culture.

The other strand of research under the content theories emphasizes the importance of “values” as a motivating force. A major research project was the Meaning of Work (MOW) studies by the MOW International Research Team that was carried out in the United States, Japan, and six Western European countries. The MOW team was primarily interested in the extent to which work is important in people’s lives in various cultural contexts. Their assumption was that

work is more motivating if it is central to one’s life. Results showed that work was most central in Japan and least central in West Germany (with the United States in the middle). More recent studies replicated these findings; Japanese employees are motivated most by the task itself, whereas their U.S. counterparts are motivated mostly by affiliation and recognition.

There are other large-scale cross-cultural studies on the motivating value of work. One was conducted by the Work Importance Study (WIS) group in 11 diverse countries, including Croatia, South Africa, and Portugal. The findings showed significant cross-cultural differences with respect to the extent to which work is a central life value compared with family, community, and leisure. Another study, called the Work Socialization of Youth (WOSY) project, examined the patterns of meanings of work from a longitudinal perspective in 7 West European countries and Israel. The results demonstrated that the meaning of work was changing across cultures and over time. Finally, the Chinese Culture Connection team demonstrated the salience of the “Confucian work ethic” emphasizing moral duty and obligation to the collectivity as being the primary motivating factors.

### **3.2. Cross-Cultural Differences in How Employees Are Motivated at Work: Process Theories**

Process theories deal with the ways in which employees are motivated at work. There are four major process theories of motivation: goal-setting theory, job enrichment theory, equity theory, and expectancy theory. Among the four, goal-setting theory of Locke and Latham has received the most attention from cross-cultural researchers. The theory asserts that the existence of goals is motivating and that goals with particular characteristics (e.g., challenging, specific, accepted) are more motivating than others. Participation and joint decision making are considered to be a key factor in motivation. Participation enhances motivation in that it provides greater autonomy and control over work outcomes and reduces anxiety and uncertainties. The Industrial Democracy in Europe (IDE) International Research Group compared the participative decision-making practices in 12 European countries in 1981 and repeated the study 10 years later in 10 of the original 12 countries as well as in Japan and Poland. This was a major undertaking to show that participation had different meanings and practices across cultures. These

studies show that in the most collectivist cultural contexts, group participation is more motivating and effective in improving work outcomes than is individual participation. Individual and face-to-face involvement in decision making (i.e., the North American model) is, in fact, not desired in all cultural contexts. In cultures that Hofstede identified as high in power distance and uncertainty avoidance, participation may even result in demotivation because subordinates seek guidance from their superiors to reduce uncertainties. Research suggests that for participative goal setting to be motivating in developing countries, organizations should enhance employees' self-efficacy beliefs through empowerment, coaching, and mentoring.

The job enrichment theory of Hackman and Oldham suggests that workers are motivated to the extent that they experience meaningfulness in their jobs. This can be achieved through more autonomy and discretion in one's job, feedback on performance, opportunity to use multiple skills (i.e., skill variety), identification of the contribution of one's job to the overall work context (i.e., task identity), and knowledge of the significant contribution of one's job to others' lives (i.e., task significance). Jobs that are designed to have high autonomy, feedback, skill variety, task identity, and task significance are referred to as "enriched jobs." This theory is based on individualist assumptions of employees' desire for independence, freedom, and challenge. Thus, it is argued that job enrichment might not be as effective in countries with power hierarchies and avoidance of uncertainties. Increasing discretion and responsibility elevate anxiety and uncertainty, and worker autonomy defies the role of the authority. In more collectivist European countries, job design takes the form of "autonomous work groups" in which job enrichment principles are implemented at the group level (e.g., team autonomy, team-based feedback). However, this system is reported to have limitations in developing countries due to uncertainty avoidance and resource scarcity that is endemic to the majority of developing countries.

The equity theory of Adams purports that employee motivation is achieved by invoking the perception of fairness. One way in which to increase fairness is to establish an equitable relationship between "inputs" and "outputs" among workers. Inputs represent everything that employees perceive as their contribution to the job (e.g., training, skills, time, effort, loyalty, commitment to the organization), whereas outputs represent everything that employees perceive as received in return for their contribution (e.g., pay, promotion, praise, recognition and awards). Equity theory has not attracted much

cross-cultural research attention, but the available evidence suggests that what constitutes inputs and outputs varies across cultures. In collectivist cultures, loyalty, commitment, seniority, and tenure are more important inputs for promotion and rewards than is job performance. Also, in cultures where there are status hierarchies, the social class or caste to which one belongs is considered an input. Similarly, there are cultural variations in the perception of outcomes. Challenge and autonomy in the job and salary increases are considered to be valued outcomes in individualist cultures, whereas praise from supervisors can be an important outcome in collectivist cultures. There are ways in which to restore equity such as changing inputs (e.g., working less), changing the comparison group, changing the outcomes (e.g., asking for a salary increase), and leaving the organization. There is not enough research to indicate cross-cultural variations in the ways in which to restore equity. However, there is a large body of cross-cultural research in the area of organizational justice. Reward allocation based on an equity principle is not endorsed in all cultural contexts.

The final process theory of motivation is the expectancy theory of Vroom. According to the theory, employees are motivated to the extent that their expectations are met in the following ways. First, if they exert enough effort, their job performance will be at the desired level. Second, if they perform at the desired level, it will lead to some outcomes. Third, the outcomes will have high valence or attractiveness. Although there is no systematic research testing the validity of the theory across cultures, it seems to have more applicability in individualist cultures than in collectivist ones. In individualist cultures, employees are motivated if there is a high likelihood of meeting their own expectations, whereas in collectivist cultures, motivation is related to the probability of meeting the expectations of "significant others" (e.g., family, organization, superiors). "Face saving" is also an important motivator in collectivist cultures. Another cross-cultural variation may exist in people's beliefs in the controllability of events. In fatalistic cultures, employees do not believe that control of events is in their hands; that is, they have an "external locus of control." Thus, expectancy theory might not be useful in such cultural contexts.

### **3.3. The Culture-Based Model of Work Motivation**

The first theoretical model explaining how culture influences motivational practices and their outcomes was

proposed by Erez and Earley in 1993. According to the cultural self-representation model, motivation is related to the employee's self-concept. Cultural context determines how one perceives oneself. To the extent that motivational practices are congruent with culturally based self-construals (e.g., the way in which the self is defined and perceived), there is a high likelihood of performance and satisfaction. For example, if in teamwork employees receive rewards based on their individual contributions to the team, this is congruent with the cultural context where the self is perceived as an independent entity (i.e., individualism), whereas group reward is more motivating in the cultural contexts in which employees have interdependent self-construals (i.e., collectivism). The congruence results in three self-derived motives: self-enhancement (maintaining a positive cognitive and affective state about the self), self-efficacy (perceiving oneself as competent and efficacious), and self-consistency (experiencing coherence and continuity). For employees who have interdependent self-construals, individual-based rewards in teamwork lower the positive self-representation and hence motivation. In general, employees are motivated through organizational practices that maintain and improve their self-enhancement, self-efficacy, and self-consistency.

#### **4. CROSS-CULTURAL VARIATIONS IN HUMAN RESOURCE MANAGEMENT PRACTICES**

With increasing globalization, staffing and performance management have become the keys to success of multinational firms. In the HRM field, these are the two topics that have received more cross-cultural research attention than have other areas of HRM (e.g., human resource planning, job analysis, compensation, reward management).

##### **4.1. Staffing: Recruitment and Selection**

Culture influences recruitment and selection in many ways such as in attitudes toward selection and testing, the purpose that staffing serves, and the perceived fairness and appropriateness of criteria and methods that are used in the process. The first theme of cross-cultural variation is the attitudes toward selection and testing. In the North American context, selection is a one-way process whereby applicants are tested to

predict their future work behavior. However, in the European context, selection is a mutual agreement and negotiation between the organization and the candidate. In some European countries such as Italy, France, Sweden, and Portugal, the issue of testing has a negative connotation because it is perceived as invasion of privacy, violating individuals' rights to control their own careers and creating a barrier to the holistic representation of oneself.

The main purpose of recruitment and selection also varies across cultures. In North America, strategic human resource planning is geared toward recruiting the right number of people with the best qualifications to do the job. In contrast, in India and Eastern Europe, organizations hire more employees than are needed so as to combat poverty and unemployment. In such cultural contexts, the strategic goal of organizations is to meet societal needs as well as economic ones.

The criteria for selection should be perceived as relevant to the job and as predictive of future performance. In the United States, some of the most commonly used criteria for selection include education, cognitive skills, job experiences, and personality traits. There is a paucity of cross-cultural research on the appropriateness of such selection criteria in other cultural contexts. However, the few existing studies cast doubt on the appropriateness of criteria such as education. It is argued that knowledge and skills acquired by formal education might not be the best predictors of performance in countries where education either is not available to everyone or is inadequate in preparing individuals for the workforce. A few studies suggest that criteria for selection are more interpersonal than individual in collectivist cultures. For example, in collectivist cultures, social selection criteria are used—team members' favorable opinions about the candidate in Japan; agreeableness, good interpersonal relations, and trustworthiness in Islamic Arab countries; belonging to the same in-group as the manager (e.g., the same family or homeland) in India; and positive attitudes toward family life in Latin America.

Once criteria are established, the next step involves determining the method of recruitment and selection. Fairness is related to the candidates' perception of the appropriateness of recruitment channels, the questions that are asked on application forms, and job previews (i.e., a brief and a realistic review of the job context and the work environment). Word-of-mouth is a common method of announcing a job opening in cultures where the hiring of in-group members is favored. Some questions on the job application forms (e.g., marital status, religious and family backgrounds, physical appearance)

could be perceived as offensive in some cultures more than in others. In job previews, a detailed description of a specific job position will not be a requirement of an ideal recruitment process in cultures where “generalists” are preferred over “specialists” (e.g., Japan).

The methods of selection used in various countries have attracted some cross-cultural research attention, but countries included in these studies have been limited mainly to those in North America and Northwestern Europe. In these studies, biodata was found to be among the least common methods of selection, whereas the interview was among the most common. In a majority of countries, references or recommendations are used for different reasons and to varying degrees. For instance, in the United Kingdom, the United States, and Australia, recommendations are used only as a final check, whereas they are relied on heavily in Southeastern European countries and in India.

In the North American literature, studies on the predictive validity of various selection methods have increased in number over recent years. However, the social validity (i.e., perceived fairness and acceptability of a particular practice) is an issue that is overlooked in the cross-cultural literature. For instance, in one study, French job applicants perceived written ability tests to be less impersonal and perceived personality tests to be more offensive in violating privacy than did American applicants.

This brief review of the literature shows that there are cross-cultural variations in various aspects of the recruitment and selection process.

#### **4.2. Performance Management: Criteria, Evaluation, and Development**

Criterion development and measurement has been an issue of ongoing debate among I-O psychologists. There is neither a uniform description of “performance” nor consensus on the dimensions of it. The criterion problem is exacerbated at the cross-cultural level because both the definition and structure of “good performance” are culture bound. According to some of the recent North American theoretical formulations, the most salient dimensions of performance are job-specific knowledge and proficiency, communication competence, effort, mentorship/supervision, quality of output, productivity, interpersonal competence/teamwork facilitation, and administrative competence. Cross-cultural replications of these dimensions have yet to be done.

What constitutes good performance in various cultural contexts? Performance-orientated cultures place more emphasis on the end result rather than on the process. As such, performance criteria are objective and measurable. In collectivist and high-power distance cultures, employee loyalty is valued more than productivity. Outcomes are important, but social and relational criteria, which are more subjective, are weighted heavily in evaluating employees. Such criteria include good human nature, harmony in interpersonal relations, trustworthiness, respectful attitude, loyalty and deference toward superiors, effort and willingness to work, awareness of duties and obligations, gratitude, organizational citizenship, conformity, and contribution to team maintenance.

Unlike individualist cultures, collectivist cultures downplay individual differences. Therefore, the primary purpose of performance evaluation is not differentiating among individuals but rather justifying decisions on compensation and promotion. Some authors purport that appraisal is used to control and instill loyalty. Formal evaluations are conducted less frequently (because the primary purpose is not to provide feedback), and their timing coincides with salary adjustment and promotion periods. In high-power distance cultures, performance evaluations may also serve the purpose of reinforcing the authority structure. In such cultural contexts, performance is usually evaluated by superiors (i.e., immediate supervisors, managers, and/or several levels up). Because performance appraisal is a top-down unilateral process, evaluation of superiors is rare in high-power distance cultures. Therefore, the 360-degree performance appraisal method (i.e., self-evaluation plus evaluation by superiors, subordinates, peers, and customers) that is popular in the United States is not appropriate in high-power distance cultures. In collectivist cultures, the 360-degree evaluation may disturb group harmony due to constant monitoring of the behavior of one’s colleagues. The accuracy of such evaluations is also jeopardized because in-group favoritism and loyalty prevent assignment of low performance ratings to in-group colleagues and superiors.

Rating errors are barriers to obtaining reliable performance evaluations. The most pervasive ones include attribution and leniency errors. The accuracy of the process depends partly on the correctness of attribution of success and failure. Cross-cultural variations in attribution are well documented. For instance, Asians usually attribute their success to luck and help received from others, whereas they usually attribute failure to their lack of ability and inadequate personal effort. In a collectivist cultural context, those who

engage in self-effacement gain acceptance by peers, whereas those who take pride in individual achievements receive disapproval. Modesty in self-evaluations leads to negative leniency bias, that is, assigning a lower evaluation to oneself than is deserved.

The final topic is related to the communication of performance feedback. Culture has a bearing on the way in which feedback is given and received. Regardless of the cultural context, performance feedback is not easy to give or receive, especially when the feedback is of a negative nature. However, in some cultures where the distinction between life and work space is blurred, feedback on one's job performance is perceived as equivalent to a rejection of the person, who then experiences loss of face. On the other hand, positive feedback on performance is also not common in collectivist cultures. Positive feedback on individual performance could disturb group harmony because it may induce jealousy and resentment among those who did not receive such feedback. Also, in collectivist cultures, positive feedback is expected to come from the outside. If a manager praises his or her own employees, it could be perceived as inappropriate and self-serving.

If neither negative nor positive feedback is appreciated, this implies that no feedback could be the most preferred condition in collectivist cultures. There is at least one empirical study that supports this view, showing that Japanese and Chinese employees do not take any initiative in seeking feedback on individual performance. This is explained by three factors. First, as mentioned previously, there is avoidance of negative or positive feedback. Second, seeking feedback on individual performance, rather than on group performance, is perceived as "self-centeredness." Finally, communication is indirect and subtle in collectivist cultures. Performance-related information is usually embedded in contextual cues rather than revealed in a direct and confrontational manner.

The final phase in performance management concerns the improvement of performance. The most frequently administered practices used to improve performance include performance-based rewards, training, and coaching/mentoring. Are rewards contingent on performance? The answer depends on the cultural context. It was found that in collectivist and fatalistic cultures, there is a weak performance–reward contingency due to a number of factors. First, in fatalistic cultures, managers believe that employees are not malleable; therefore, managers do not see any point in rewarding employees as a means of changing work behavior. Second, individualist cultures use the norm of equity in allocating

rewards. Therefore, there is one-to-one correspondence between performance and reward in individualist cultures, whereas reward allocation in collectivist cultures depends on contingencies other than performance. Finally, in collectivist cultures, rewarding the group as a whole is preferred over rewarding individual members. As mentioned previously, individual gratification may result in disturbed group harmony and resentment.

The second practice used to enhance performance is training. One of the first questions to ask is whether or not training needs are determined on the basis of performance outcomes. This seems to be the case in individualist and performance-oriented cultures. In low-performance-oriented and high-power distance contexts, decisions about who will participate in trainings are based on criteria other than job performance. Employees who maintain good relations with higher management are sent to desirable training programs (i.e., training overseas or at resorts) as a reward for their loyalty. Because training is given as a favor or reward, trainees provide highly favorable feedback on trainings and introduce a bias to the evaluation of training effectiveness.

The next practice in performance development is coaching and mentoring, which involves a personalized subordinate–superior relationship to enhance performance. Mentors are senior and more experienced people who provide guidance to employees. The role of a mentor includes modeling correct behavior, setting clear performance standards and goals with employees, coaching, monitoring and giving constructive feedback on performance, and making necessary adjustments to tasks and work environments to improve performance. Mentors work with and for employees. High power distance is not compatible with mentorship. In high-power distance cultures, superior–subordinate relationships are not personalized and managers are not usually participative in goal setting. However, mentorship is more likely when high power distance is combined with paternalism. In paternalistic cultures, superior–subordinate relationships are personalized and managers guide employees in professional as well as personal matters.

### **4.3. The Culture-Based Model of Human Resource Management**

Kanungo and colleagues proposed and tested a theoretical model explicitly linking culture to HRM practices.

The model of culture fit asserts that variations in HRM practices are due to the internal work culture, which in turn is influenced by both the sociocultural context and the enterprise environment. The internal work culture is composed of managerial beliefs and assumptions about two fundamental organizational elements: the task and the employees. Managerial assumptions pertaining to the task deal with the nature of the task and how it should be accomplished best, whereas those pertaining to the employees deal with employee nature and behavior.

Organizations design and implement HRM practices based on the prevailing managerial assumptions about the nature of both the task and the employees. These assumptions, in turn, are influenced by two major environmental forces. First, the enterprise environment includes the ownership status (private sector vs public sector), industry, market competitiveness, and resource availability. Second, employee-related assumptions are influenced by the characteristics of the sociocultural environment.

The model of culture fit was tested in various cultural contexts, including Canada, the United States, Romania, Germany, Israel, India, Pakistan, China, Turkey, and Russia, with the following key findings. In fatalistic cultures, managers assume that it is not possible to change employee nature and behavior. Accordingly, they do not provide feedback or try to improve performance through rewards. In cultures where there are power hierarchies, managers do not hold the assumption that employees are proactive; rather, they assume that employees are reactive (i.e., expecting close guidance and supervision). As a result, managers do not give autonomy and discretion to employees in their job assignments. In collectivist cultures, managers assume that organizations and employees should fulfill obligations to one another. Based on this assumption, they provide empowering supervision both to fulfill their obligations toward workers and to enable workers to fulfill their obligations toward others in the organization. The model of culture fit is an important initial step toward understanding how, and to what extent, culture influences HRM practices. Cross-cultural and multilevel testing of the model is continuing.

## **5. MANAGING CULTURAL DIVERSITY**

This section examines cultural diversity under three categories: expatriate management, women's career advancement, and multicultural teams.

### **5.1. Expatriate Management**

Expatriates are employees of business or government organizations who are sent by their organizations to related units in countries other than their own for pre-designated temporary time periods of usually more than 6 months and less than 5 years. In global competition, successful expatriation not only gives multinational companies a competitive edge, it also contributes substantially to professional growth of expatriates. Despite the importance attached to international assignments, failure to complete the full cycle of such an assignment and failure to work effectively on the job remain major problems. Criteria for effective expatriation include both adjustment to the new cultural context and high job performance. Difficulties in adjustment and performance result in high turnover rates and, therefore, failure to receive a return on investment. Stimulated by the observed problems in expatriation and the costs associated with it, this topic has emerged as one of the most popular topics for cross-cultural researchers.

Studies have identified a number of key factors influencing expatriates' cross-cultural adjustment and performance. The individual factors include those related to expatriates and their families. There is some research suggesting that personality characteristics, such as openness to new experiences, agreeableness, and tolerance for ambiguity, predict better adjustment outcomes. Competence in interpersonal relationships and willingness to learn about a new culture are factors that facilitate both adjustment and performance. Recent research also suggests that those individuals who have had cross-cultural interactions during their childhoods stand a better chance of succeeding in international assignments. Expatriate-related factors also include family adjustment issues. Expatriates whose families have adjustment difficulties are more likely to leave their assignments prematurely.

Organization-related factors include the preparatory stage and the HRM practices of the company. The majority of expatriates who fail in their assignments have been selected solely on the basis on technical and managerial competencies. Their adjustment potential and interpersonal competencies are usually not assessed. Similarly, many organizations fail to provide cross-cultural and country-specific training prior to departure. These problems are exacerbated by the uncertainties related to the assignment terms and job specifications. Difficulty in finding positions in the parent company on repatriation also lowers job motivation and performance after the employees return. Furthermore, the receiving

organization (i.e., the local unit) is usually ill prepared to manage the cultural diversity resulting from expatriates' arrival. Local units that offer training on multicultural team management provide better support and assistance to the incoming expatriates, thereby facilitating their adjustment and performance. Finally, cultural factors influence expatriation outcomes. In assignments to countries with large cultural distance (i.e., dissimilarity between the culture of origin and the culture of the host country), expatriates experience more difficulties.

## 5.2. Women's Career Development

According to the reports of international organizations (e.g., United Nations' *Human Development Reports*, *World Bank Reports*, International Labor Organization reports), women receive less pay, less occupational status, and less career advancement all around the world. Although women share similar experiences at work and in family, cross-cultural differences contribute to the magnitude of discrimination and conflict between work and family. The literature canvasses various approaches to explain women's difficulty in gaining equal access to career development opportunities. First, there is compelling evidence suggesting the effect of gender role stereotypes. In all cultures, women's primary responsibility is toward the family. In some cultures, women feel the freedom to choose between pursuing a career and raising a family; however, in the majority of countries, women do not have such a choice. Cross-cultural differences exist with respect to norms in childbearing. Career women may opt to have children in some economically developed countries (e.g., France), whereas this is not an option for women in many traditional cultures of developing countries. Moreover, in many collectivist cultures, women's family responsibilities include taking care of extended family members, in-laws, and aging parents, adding to women's total workload.

A second difficulty stems from the attitudes toward working women. In the majority of countries in the world, organizations hold negative attitudes toward women's career development. A common belief is that women are cognitively and psychologically ill prepared to become leaders in the business world. This belief is derived from socialization values and practices. In many traditional cultures, women are socialized to be shy, unassertive, dependent, passive, and obedient to the men in their lives. For example, according to the traditional beliefs in China, it is a virtue for a woman to have no leadership ability. Women themselves internalize

these norms and develop unfavorable attitudes toward their career development.

Women are also cut off from male networks. As a result, they miss opportunities that would help them to advance in their careers. For instance, in Japan and Taiwan, evening gatherings in pubs or clubs are very common. However, because women's participation in such social gatherings is considered inappropriate, they are inadvertently cut off from the critical social (and hence informational) networks. Another negative attitude stems from the concern about women's and men's changing status in the family. In traditional societies, if a woman works, it is perceived to be the failure of the husband to provide for the family. It is also feared that women's economic independence will decrease men's power in the family.

Finally, women receive insufficient support, especially with respect to child care, from their families and the organizations for which they work. One of the most important sources of support is the spouse. In traditional societies, a man's helping out his wife with her family duties is uncommon. Husbands who do housework are stigmatized as being "woman-like," "submissive," "passive," or simply "weak." In such cases, the values and behaviors of husbands, especially those from high educational backgrounds, oscillate between traditional and modern. Support is also rare in organizations. Family-friendly work schedules and day care facilities are rare in many countries. Although managers are more tolerant and understanding of women's family responsibilities in paternalistic and collectivist cultures, such tolerance does not indicate support for women's career advancement.

## 5.3. Multicultural Teams

Teamwork is one of the key factors for organizational success. With increasing globalization, effective management of multicultural teams has become a critical issue. Research in this area is still developing, and conclusive findings are few. There are mixed findings regarding the performance of culturally heterogeneous teams compared with that of homogenous ones. There are four factors that influence the functioning of multicultural teams. First, differences in core values and norms regarding how members should behave and interact with one another lower the performance at initial stages. For example, members from a collectivist culture value interpersonal harmony more than they do task accomplishment. As such, they avoid confrontations and giving negative performance feedback.

There is at least one study indicating that social loafing (i.e., the tendency to withhold effort when performing a group task) occurs less in teams in collectivist cultures than in teams in individualist cultures. In collectivist cultures, team members work hard to achieve the group goals, and members feel obligated to contribute to the team. Members of high-power distance cultures prefer teams with leaders who guide them closely, whereas members of low-power distance cultures may function better in autonomous or leaderless teams. Differences in norms, values, and preferences are potential barriers to effective functioning in multicultural teams.

The second factor that influences team performance is the perceived cultural distance of the team members. Recent research found a curvilinear (i.e., inverted U-shaped) relationship between the perceived cultural distance and team performance in such a way that culturally homogenous and highly heterogeneous teams outperform moderately heterogeneous teams. Formation of subgroups is more likely to occur in moderately heterogeneous teams than in highly diverse teams. This lowers the team coherence and performance.

The third factor is time. Performance in diverse teams is a function of time. Heterogeneous teams are found to outperform homogenous teams over time. It takes more time for diverse teams to build trust and shared values and norms. Initial inefficiencies (i.e., process losses) due to inadequate communication and feedback may be overcome in time.

Finally, task structure affects performance of multicultural teams. In jobs that require an in-depth understanding of a particular cultural context (e.g., launching a new product in a new cultural context), cultural knowledge is task related and, hence, increases team performance. Multicultural teams perform better at tasks requiring idea generation and creativity than they do at production tasks requiring little discretion. Diverse teams function better under structured and noncompetitive task environments. Superordinate goals (i.e., goals that group members cannot reach without the help of the other group members) also facilitate team performance in multicultural teams.

Teams working in virtual environments (e.g., the Internet, videoconferencing) experience further difficulties in adjusting to uncertainties arising from technological problems. Virtual teamwork is more difficult for members who prefer face-to-face interactions (e.g., members of collectivist cultures) as well as contextual cues in communication (e.g., members of high-context cultures).

## 6. CULTURE AND WORK ATTITUDES

Work attitudes focus on how employees feel about their jobs and the organization. Exploring the cultural factors influencing work attitudes is one of the most popular research topics in cross-cultural I–O psychology because work attitudes are related to various work outcomes, including performance, productivity, turnover, and absenteeism. This section discusses the following cross-cultural differences in work attitudes: job satisfaction, organizational commitment (OC) and organizational citizenship behavior (OCB), and organizational justice.

### 6.1. Job Satisfaction

Work conditions (e.g., interpersonal relationships, pay, benefits, developmental opportunities) that are satisfying in one cultural context might not be satisfying in another cultural context. Research evidence suggests that workers in individualist cultures report more job satisfaction than do those in collectivist cultures. The higher level of job satisfaction in individualist countries is attributed to the fact that workers enjoy more individual freedom and rights and are more likely to fulfill their self-actualization needs. It has also been shown that people living in individualist countries report more life satisfaction, and this has a “spillover” effect in that it increases job satisfaction. However, recent research suggests that collectivism, rather than individualism, is positively correlated with job satisfaction. First, it is argued that employees “put up with” unpleasant work conditions because they believe that their work group members also suffer from the same conditions. The sense of sharing of this “common destiny” reduces dissatisfaction. Second, positive interpersonal relationships at work (e.g., collaboration, nonconfrontation) create a pleasant work atmosphere and enable workers to better cope with unpleasant conditions.

These seemingly conflicting findings are reconciled by research that identified contextual factors influencing job satisfaction. For instance, intrinsic job characteristics (e.g., opportunity for development, use of multiple skills, challenging job, autonomy) are more satisfying to employees in individualist cultures, whereas extrinsic job characteristics (e.g., pay, benefits, sense of security) and interpersonal aspects of the job are more satisfying to those in collectivist cultures. This variation is meaningful because job satisfaction is closely related to the

fulfillment of needs. In collectivist countries, people need to satisfy primarily their basic and interpersonal needs. As such, job characteristics such as challenge and autonomy do not yield job satisfaction unless salient needs are fulfilled. Furthermore, intrinsic job characteristics are positively associated with job satisfaction in countries with high social security (e.g., welfare system, national wealth) and small power distance, whereas this relationship does not exist in countries with low social security and large power distance. Employees in high-power distance cultures do not expect to participate in decision making, receive performance feedback, or be autonomous in their jobs. Lack of such intrinsic job characteristics does not result in dissatisfaction.

## 6.2. Organizational Commitment and Organizational Citizenship Behavior

OC is conceptualized as the extent to which employees feel a sense of allegiance toward their employers. The growing body of literature on OC suggests that it is an important organizational attitude predicting performance indicators (e.g., absenteeism, turnover intention), employee reactions to work, and employee well-being at work. In 1990, Allen and Meyer proposed a model of OC that consists of three components: affective commitment (i.e., employees' emotional attachment to and identification with the organization), continuance commitment (i.e., the perceived necessity to stay with the organization due to the high cost of leaving it), and normative commitment (i.e., the belief that one is obligated to stay loyal and committed to the employer).

Cross-cultural studies of OC have focused on two issues: structure of the construct (i.e., whether or not the three-factor model replicates across cultures) and correlates of OC. There is mixed support for the generality of the factor structure of the North American OC measures, and this casts doubt on the "universal" meaning of OC in various cultural contexts. The psychometric properties of these measures did not completely replicate in countries that are characterized by high collectivism, high power distance, and strong uncertainty avoidance (e.g., Korea). In such cultural contexts, the measurement holds better with the addition of some emic (i.e., culturally relevant) items. Cross-cultural variations are also found with respect to the correlates of OC. Antecedents and consequences of OC in Western European countries (e.g., Germany) were very similar to those in North America. Affective commitment was the best predictor of turnover

intention in individualist cultures with high employment mobility, whereas normative commitment was the best predictor in collectivist cultures emphasizing long-term employment. Desirable organizational outcomes (e.g., low absenteeism, job satisfaction, employee well-being) are strongly correlated with affective commitment in North America but are linked to normative commitment in collectivist, high-power distance, and paternalistic cultures where loyalty is valued and rewarded.

OCB is described as the contribution of employees to the welfare and effectiveness of the organization by engaging in behaviors that are neither required nor expected. These behaviors are referred to as the "good soldier" or "extra-role" behaviors. Similar to the case of OC, cross-cultural research on OCB addresses the issues of whether or not its meaning and structure transcend culture. The original theory, as proposed in North America by Organ in 1988, describes five dimensions of OCB: altruism (e.g., voluntarily offering help to others), conscientiousness (e.g., being punctual, following company rules and regulations), courtesy (e.g., being mindful and respectful of others' rights), civic virtue (e.g., participating in the political life, attending meetings, speaking up on issues), and sportsmanship (e.g., not complaining, not nursing petty grievances, not gossiping, not falsely magnifying problems). Limited cross-cultural research in this area suggests that some of these behaviors, as described in the North American literature, are not appropriate in some cultural contexts (e.g., speaking up in high-power distance cultures), whereas others are not considered as in-role rather than extra-role behaviors (e.g., voluntarily offering help in collectivist cultures). Some evidence suggests that there are emic dimensions of OCB in collectivist cultures such as protecting company resources. Because of such variations, it is tentatively concluded that the dimensionality (e.g., the factor structure) of the Western OCB construct and its measures do not transcend culture. Similarly, there are variations in the predictors and outcomes of OCB. OCB is found to be predicted by the organizational justice perception in Western literature, whereas collectivist values are seen as better predictors of OCB in Eastern traditional cultures. OCB was correlated more strongly with turnover intention and productivity in individualist cultural contexts rather than in collectivist ones.

## 6.3. Organizational Justice

Organizational justice is concerned with the perception of fair treatment of employees in organizations. Justice perception has important implications for organizations

because it relates to OC, productivity, trust, and compliance with decisions. For more than a century, philosophers have been debating the concept of fairness with little agreement. Disputes over this issue are magnified when cross-cultural differences in values and norms regarding justice are involved in research. Despite such challenges, organizational justice has proven to be a fruitful area that has attracted a great deal of cross-cultural I–O research attention.

The literature draws a distinction between two types of organizational justice. The first is referred to as “distributive justice” and is concerned with the fairness of norms or criteria used in distributing resources. The second is “procedural justice” and refers to the fairness of the means (e.g., procedures, practices) used to achieve the organizational results.

Earlier cross-cultural research on distributive justice was limited to examination of the criteria used in performance appraisal. With respect to procedural justice, the focus was on the fairness perception of various leadership styles (e.g., participative leadership). The overarching assumption was that criteria influencing justice perception had similar meanings across cultures, but the cross-cultural differences consisted of different weights being given in different cultures. Social and organizational psychology literature during the 1980s offered an “interactionist” perspective purporting that the social context cued the application of fairness criteria. Such sophisticated models advanced the literature in such a way that some authors claimed that cross-cultural research had a significant influence on the mainstream (mainly North American) theories in the area of justice.

In the area of distributive justice, cross-cultural research has investigated the differential norms that are used in distributing resources. Resources are distributed on the basis of three norms: equity, equality, and need. The equity principle uses the input–output ratio as a method of allocating resources to employees, the equality principle advocates similarity in resource allocation regardless of the differences in inputs, and the need principle promotes the distribution of resources on the basis of need. Earlier cross-cultural research found that employees in individualist cultures perceive justice as being done when the equity norm is used, whereas those in collectivist cultures favor use of the equality and need norms. However, later studies showed that collectivist cultures prefer different norms depending on the context.

In 1997, Leung proposed a model describing situational variables interacting with individualism–collectivism in cueing differences in distribution criteria. According to the model, goal-directed behavior

is the most immediate predictor of selection of the distributive criterion. For instance, the equity principle is perceived to be more appropriate in collectivist cultures when applied to out-group members if the salient goal is to increase productivity, whereas the need principle is favored when the immediate goal is to protect the well-being of others under the condition of resource scarcity. In applying the equity principle, what constitutes inputs and outcomes is culturally determined, but there is surprisingly little research on this topic. Limited available research suggests that in collectivist cultures, loyalty, family size, and networks (e.g., *guanxi*) are considered to be appropriate inputs. The criteria to be used also depend on power distance. In hierarchical societies, in-group favoritism is perceived to be fair in distributing resources. Moreover, in interpersonal comparisons, it is accepted that the more powerful ones are entitled to receive more benefits than are those at the bottom of the hierarchy. Researchers urge caution regarding changes in the perception of power, and hence the perception of justice, over time.

Cross-cultural research in procedural and interactional justice is very limited, but available evidence suggests that the perception of fairness in procedures and interpersonal treatment depends largely on cultural norms. For example, in individualist cultures, having a feeling of control over decisions increases the perception of justice, and this finding is also replicated in some collectivist cultures. Similarly, voice (i.e., the opportunity to express one’s opinions) is perceived to be fair in individualist cultures, and this perception is mediated by one’s relationship with the people in authority. In collectivist cultures, procedures that follow tradition are accepted as fair, suggesting that norms have a stronger impact on the perception of justice in collectivist cultures than in individualist ones.

There are three factors that increase the perception of procedural fairness. The first is the benevolent intent of the authority. Cross-cultural variations exist in the perception of the intent underlying the behavior of superiors. Some research suggests that workers in high-power distance cultures are more likely to tolerate harsh treatment by their superiors, partly because this behavior is perceived to be for the good of the workers and the organization. Therefore, such actions are perceived to be less unfair and less likely to harm loyalty. The second factor is neutrality, that is, the belief that behavior is unbiased. In conflict resolution among workers, supervisors’ active involvement and interference are expected and appreciated in high-power distance and paternalistic cultures, whereas it is perceived to be breaching

supervisors' neutrality in low-power distance and individualist cultures. Finally, the third factor is status recognition, that is, the extent to which decision makers show respect toward employees. People feel valued when they are shown respect, but what constitutes respecting behavior is culturally determined. There is a dearth of cross-cultural research in this exciting area.

### See Also the Following Articles

Attitudes ■ Industrial/Organizational Psychology, Overview  
 ■ Organizational Justice ■ Personnel Selection  
 ■ Recruitment ■ Work Motivation ■ Work Teams

### Further Reading

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). New York: Academic Press.
- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance, and normative commitment to the organization. *Journal of Occupational Psychology*, 63, 1–8.
- Blum, M. L., & Naylor, J. C. (1968). *Industrial psychology: Its theoretical and social foundations*. New York: Harper & Row.
- Erez, M., & Earley, P. C. (1993). *Culture, self-identity, and work*. New York: Oxford University Press.
- Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison–Wesley.
- Industrial Democracy in Europe International Research Team. (1981). *Industrial democracy in Europe*. New York: Oxford University Press.
- Leung, K. (1997). Negotiation and reward allocations across cultures. In P. C. Earley, & M. Erez (Eds.), *New perspectives on international industrial/organizational psychology* (pp. 640–675). San Francisco: New Lexington Press.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
- Vroom, V. H. (1964). *Work and motivation*. New York: John Wiley.



# Industrial/Organizational Psychology, Overview

Sheldon Zedeck

*University of California, Berkeley, California, USA*

1. Definition
  2. History
  3. Contents of the Section
  4. Other Topics
  5. Future Trends
  6. Conclusion
- Further Reading

The section on industrial/organizational (I/O) psychology presents a broad overview of the multiple topics studied under its rubric. Some of the topics are quite broad, such as motivation taxonomies, while some are rather specific, such as system safety. The goals of the section are to provide the reader with an introduction to the many areas of concern to I/O psychologists with respect to the areas in which they conduct research and practice. The purpose of this overview article is to (1) define and describe industrial and organizational psychology as a field, (2) provide a brief history of the field, (3) summarize the contents of the section, (4) provide brief summaries of topics not covered in this section, and (5) present one view of the trends in research and practice.

## 1. DEFINITION

Broadly speaking, psychology is the study of behavior. With this general definition as the basis, industrial/

organizational (I/O) psychology can be defined as the study of behavior that takes place within the context of an organization (e.g., a work setting). Adopting such a broad definition provides the opportunity for I/O psychologists to address research and practice questions that most other psychologists study most often in isolation of a context. For example, cognitive psychologists study information processing. An I/O psychologist may also study information processing, but in the context of decision making. Or, a personality psychologist may study the basic dimensions of personality. An I/O psychologist may study personality psychology to determine the traits and characteristics of individuals that will lead to effectiveness as leaders or success in organizations. Yet another example can be drawn from social psychologists, who study small group behavior and attitudes. I/O psychologists could study similar topics, such as team behavior and job satisfaction, respectively. One additional example can be taken from biological psychology, which can focus on stress and circadian rhythms. Here, I/O psychologists might be interested in stress due to conflict between work and family, and in the attitudinal and behavioral outcomes that result from shift work in organizations. These examples demonstrate the broad range of topics that I/O psychologists are concerned with and the issues that they address. They also illustrate the approach taken by I/O psychologists, which is to conduct research that derives principles

of individual, group, and organizational behavior and to apply that research base to the solution of problems in organizational settings.

One cautionary note about the preceding illustrations: they are not intended to suggest that I/O psychologists predominantly or exclusively address practical, field, or applied issues. Though many I/O psychologists conduct their research in viable organizations, and may do so to address particular problems inherent to those environments, their research is both basic and applied. That is, much of the research that is conducted by I/O psychologists contributes to the understanding of behavior as well as serves to remedy or prevent problems in organizations.

For a more complete overview of the profession of I/O psychology, see the Web site of the professional association for I/O psychologists, the Society for Industrial and Organizational Psychology ([www.siop.org](http://www.siop.org)). This site provides complete information about the profession, which has over 6,000 members from 50 U.S. states and approximately 40 countries. I/O psychologists conduct research and consult with organizations on the following topics:

Employees:

- Testing
- Selection and placement
- Training and development
- Employee attitudes
- Employee motivation
- Leadership

Organizations:

- Change management
- Organizational climate
- Organizational culture
- Job design
- Job evaluation
- Organizational structure
- Team building
- Decision making
- Workplace planning

Before proceeding with the history of I/O psychology, another term, organization, must be defined. An organization exists when two or more people get together for the purpose of accomplishing a common goal. This is a very broad and general definition of an organization, adopted to highlight the flexibility available to I/O psychologists for their context of study. The organization can be one that is global, national, or local; it can be an organization that is intended to last

for a long period of time, or it can be one that is assembled for a short-term mission, such as a political campaign. The members of the organization may be paid for their services or may be volunteers; thus, the organization can be either for profit or non-profit. The members of the organization may be selected or elected to be part of the organization. The organization can be in business for the purposes of being in private industry, the public sector (government agency), or an educational institution.

One final topic for this section of the overview pertains to who is an I/O psychologist. Generally, I/O psychologists receive their training from doctoral level psychology departments that have special programs in I/O psychology. Thus, these individuals generally have a broad background in the research and methodology of psychology as well as in data analysis. Another route for education is to obtain a doctorate in organizational behavior (OB) from schools of business administration. Such programs often interact closely with psychology departments to provide the basic background in psychology needed to conduct research and practice in the domain of behavior. In addition, these programs provide a greater emphasis on business-related topics such as strategic planning, finance, and marketing. Often, however, the perspectives of the I/O psychologist and OB student may vary. For example, the former may be interested in individual differences and micro-level issues regarding how the individual behaves within the context, whereas the latter may prefer to focus on more macro-level issues such as the organization's impact on the individual. Note that this distinction is fuzzy and not rigid—the ideal prototype I/O psychologist is one who understands both micro- and macro-level issues.

## 2. HISTORY

This section identifies the primary individuals and events that have defined and distinguished the field of I/O psychology. This history is presented in a chronological order to show how the field has developed over the last 100 years and how the field has been influenced by historical and social events.

The formal history of the field begins in the early 1900s. One of the earliest psychologists recognized for his application of psychological principles to problems of the work setting is Hugo Munsterberg. Munsterberg studied in the laboratory of Wilhelm Wundt, an experimental psychologist who studied perception in Germany. When Munsterberg arrived in the United

States, he established a psychological laboratory at Harvard University, where he studied issues concerned with industrial accidents and the relationship between abilities and performance and industrial efficiency. He wrote one of the earliest books on I/O psychology in 1913, *Psychology and Industrial Efficiency*. Perhaps one of his earliest contributions was his applied work, his studies on the characteristics of successful motormen for the Boston streetcar system.

Another early pioneer, who also studied with Wundt, was James McKeen Cattell. Cattell studied individual differences in people; the critical notion for him was that differences between people should not be considered as “errors,” which was the way experimental psychologists viewed them, but rather as reliable characteristics that differentiated people and that explained why people differed from each other. Cattell was one of the first to use the term “mental test,” and, in many ways, is the founder of the field of selection and personnel.

Two other psychologists who were active in the early 1900s are Walter Dill Scott and Walter Van Dyke Bingham. Scott was active in two areas: he explored the field of advertising as well as the issue of selecting sales personnel. Regarding the latter, he was one of the earliest to demonstrate that the typical interview lacked reliability and validity. He was also interested in training and learning curves, performance rating systems, trade tests, occupational families, and job specifications. Another distinction for Scott is that he was a professor of applied psychology at Carnegie Institute of Technology in the mid-1910s.

Another landmark in the 1910s was the work of Frank and Lillian Gilbreth. They applied the “scientific management” principles of Frederick W. Taylor (1911) to educational institutions. Scientific management basically promoted the simplification of work processes. Work methods were designed to be efficient, and workers were selected who could be trained to perform the tasks. The Gilbreths applied time and motion studies to the ways in which workers performed tasks, both to understand how those tasks were performed and to design a more efficient way in which to perform those tasks in order to reduce fatigue and to increase productivity. These events were the forerunners for work on job analysis as well as human factors engineering—both topics of relevance today for I/O psychologists.

A significant historical event impacted the role of psychologists in the early 1900s—the entry of the United States into World War I in 1917. Scott, along with Bingham, adapted an intelligence test that was

designed for individual testing to a format that could be used with groups of test takers; such a test was needed for use with large numbers of candidates to determine who was fit to serve in the armed forces. The result was the Army Alpha test, which was used to select recruits into the armed forces. This was an instance of applying principles of individual differences to selection problems; it was also the forerunner of major efforts to develop selection tests and systems for organizations, a trend that continues today.

The 1930s introduced another trend in I/O psychology: the study of human relations and attitudes. Elton Mayo was interested in studying the relationship between work—particularly boring work—and attitudes and performance. Mayo, Roethlisberger, and Dickson are the main researchers identified with a classic set of studies—the Hawthorne studies—that took place in the Hawthorne plant of the Western Electric Corporation. The study began in 1926 and lasted for more than 10 years. The initial goal of the study was to explore the relationship between productivity and plant conditions, such as lighting, heating, work hours, rest breaks, and other manipulations that were under the control of the organization. The results were contrary to what was expected: even when the conditions were made worse, productivity increased. This finding led the researchers to interview the workers, who indicated that they were “pleased” that the organization was paying attention to them, as represented by the efforts to study the problems; as a result, they increased their performance. This finding—that performance can be increased when workers are attended to—is now known as the Hawthorne Effect. This research led to a trend toward greater concern for the individual workers and an emphasis on social relations, known as the human relations movement. Topics that resulted from this movement include motivation, job satisfaction, and morale.

World War II (early 1940s) was also instrumental in inspiring psychologists to begin new explorations for the purpose of helping the war effort. Of particular significance is the role psychologists played in the design of cockpits for the new fighter planes that were being used in the war. The different types of planes and their cockpit controls required pilots who could function in environments with all sorts of new airplane controls, gauges, and displays. Consequently, the psychologists helped in the design of the cockpit, to facilitate quicker recognition of the dials and controls and with less error in their use. This is another example of human factors engineering as well as work design.

In addition to human factors engineering, World War II was the catalyst for research on selection and assessment. New selection tests were developed to help match recruits to jobs that fit their abilities. In addition, attempts to identify personnel who could serve as spies for the United States resulted in the development of the “assessment center” procedure. The assessment center is a mixture of multiple types of assessment, e.g., individual tests, personality assessment, interviews, role-plays, and group exercises, many of which are simulations of actual job behavior. The set of exercises is designed to provide a picture of the individual and how he or she is likely to behave in a work setting. Today, assessment centers are a major industry used to select people for all levels within an organization.

After WWII, I/O psychology became more recognized, and as a result, degree programs were established at major universities and colleges in the United States. Many of the early pioneers in the field brought their training in experimental or social psychology to the programs, producing new Ph.D.s with interests in applied psychological problems. Some of the issues studied were leadership, participative decision making, and job satisfaction.

The role of the I/O psychologist was again challenged in the 1960s, with the passage of the 1964 Civil Rights Act. This act made it illegal to discriminate in selection, hiring, and promotion of employees based on race, ethnicity, national origin, and gender. As a result of this act, the hiring and promotion practices of many organizations were challenged, which led to the organizations' attempts to demonstrate the validity and usefulness of their selection procedures. For I/O psychologists, this meant increased efforts and research in understanding issues of selection and validation as well as better understanding of how tests can be used in the workplace environment. This law spurred research on validation models, the development of job analysis procedures, concerns about the measurement of performance, and the concept of validity generalization.

In the latter part of the 20th century (1970s to 2000), even though there was still a major focus on personnel psychology, a considerable amount of research and practice focused on the individual within the organization. Topics such as goal setting, satisfaction, organizational commitment, motivation, organizational climate, person–organization fit, organizational citizenship behavior, and justice received increasing research attention. From an economic perspective, some issues of interest were mergers, acquisitions, and downsizing.

The late 1990s and the early part of the 21st century have continued to see the interest in the social aspects of behavior. But this also has been influenced by the changing nature of work and the reliance on technology for that work. This trend will be revisited at the end of this overview. For now, and as a summary, the presentation of this brief history was intended to show that practical problems, within a context, often influenced by politics, economics, conflicts (war), and legislation, have been instrumental in defining what I/O psychologists studied.

### **3. CONTENTS OF THE SECTION**

One way of structuring how I/O psychologists study their topics is to view the study of organizations in three levels or phases along with their associated issues: (1) the need to develop the organization, (2) the need to obtain members for the organization, and (3) the need to maintain (and retain) the organization as a viable entity with members who contribute to its mission. Developing the organization refers to the issues that must be considered when starting an organization. These topics include organizational philosophy, organizational theory, organizational structure, leadership, communications, decision-making strategies, and other issues that relate to how the organization and its jobs will be designed, established, and conducted.

#### **3.1. Development Phase**

This section of the encyclopedia contains articles relevant to the development phase on the topics of organizational structure (authored by Tetric and Camburn), power, authority, and leadership (Munduate Jaca and Medina), organizational culture and climate (Peterson and Fischer), decision making (Jungermann), and organization development (Burke). These are more global topics that focus on organizational characteristics. But there are also very specific concerns, such as how the jobs will be designed and analyzed. A particular article within this section that is relevant to this concern focuses on job analysis, design, and evaluation (Smither). In addition to interest in the tasks needed to be performed and the knowledge, skills, and abilities (KSAs) needed to perform those tasks, organizations are also concerned with the safety of work, which is discussed in articles on system safety (Wilpert) and work safety (Zohar).

The topics described above are critical for the establishment of an organization, since they impact the type

of recruits attracted to the organization and the types of systems that will be developed to maintain the organization. Note, however, that as stated earlier, these are not isolated topics that are restricted to the development phase of setting up the organization. Here, as in the other two phases, the individual topics are appropriate for discussion and consideration for all phases. For example, although it is critical to identify the organizational structure at the outset, concerns about performance and satisfaction could lead to the need for organizational development activities that would take place several years after the establishment of the organization. After studying an organization that has been in operation for several years, it may be necessary to make changes in the structure to accommodate changes in the nature of the workforce, the demands of the public regarding the product produced by the organization, and new laws that govern how products should be produced and what materials they contain. The critical points to appreciate are that all of the topics are interrelated and all are important at any and all phases in the life of an organization.

### 3.2. Selection Phase

Once an organization is designed and there are plans for its operations, the next issue is the second phase: obtaining members to participate in that organization. When considering how to select employees at all levels within the organization, the I/O psychologist is concerned with the bridge between design of jobs and how recruits are selected to perform those jobs; that bridge is based on the previously mentioned article on job design and analysis. In addition, there is the related issue of how employees will be compensated for the work they perform in their roles, which is discussed in an article on compensation (Thierry).

Once jobs are designed, it is necessary to understand how employees will be evaluated in their positions or jobs—the topic of performance appraisal. Here, I/O psychologists need to understand what contributes to success and effectiveness in the job and organization, and how to measure employees on the relevant criteria. Two articles deal with the measurement of individual and organizational performance on the tasks: one on competence at work (DeNisi) and one on productivity (Pritchard). High individual performance and high organizational productivity are the criteria that the organization selects employees to achieve. Another criterion used to measure the organization's success is turnover; this is discussed in the article on labor

turnover (Hom). The selection system is designed to identify those who can meet these goals.

This section also contains three critical articles that set the stage for the other topics within the selection phase—occupational choice (Hansen), recruitment (Breugh), and personnel psychology (Chan). The occupational choice article considers how people choose the careers they want to pursue, focusing on the individual and what he or she wants to do as a career. The personnel psychology article focuses on the organization's choice of individuals, which is obviously related to what individuals decide to do in terms of careers. How the organization goes about identifying candidates for its positions is discussed in the article on recruitment. In other words, there is a reciprocal relationship between one's choice of a career and the organization's recruitment and choice of people.

Once the organization is ready to implement a selection system, it needs to consider the issues discussed in the articles on employment discrimination (Tenopyr) and employment interviewing (McDaniel and Whetzel). The former deals with laws that exist, particularly in the United States, that influence what selection devices can be used and under what conditions, as well as the litigation surrounding the use and validation of the tests. Related to the topic of discrimination and selection is the issue of affirmative action, which is discussed in an article by Kravitz. The topic of validation is covered in the previously mentioned article on personnel psychology (Chan) as well as in the article entitled "Personnel Selection" by Ones, which appears in another section of the encyclopedia.

A link between obtaining members for and maintaining the organization is the topic of training and development. Organizations can select employees who have the aptitude to be trained, or employees who are already trained and capable of producing and performing rather quickly after entry into the job. Training is a critical issue of concern for I/O psychologists. A special topic related to training is that of executive coaching and executive development, discussed in an article by Silzer, which is concerned with working with employees to enhance and develop their skills for management positions. This not only is an issue when deciding who to select and whether training is needed, but also can serve as a motivational tool for employees.

### 3.3. Maintenance Phase

Once the organization has determined how it will obtain its employees and implements the systems needed

to obtain those members, it turns its attention to the maintenance phase: how to retain the members in the organization and to further maintain and develop the organization to sustain itself in the marketplace. This phase is concerned with issues such as reward systems, motivation, job satisfaction, organizational climate and culture, work and family, conflict, and stress.

Articles relevant to the maintain phase are those on the topics of motivational taxonomies (Vancouver), motives and goals (Kleinbeck), and boredom (Balzer, Smith, and Burnfield). These articles address issues such as how to motivate employees to perform better and how to make the work more motivating and satisfying for the employees. Articles that deal with group and team issues include those on organizational participation (Heller), organizational socialization (Bauer), and cooperation at work (Salas). These issues relate, in part, to whether there is a person–environment fit, discussed in an article by Ostroff and Aumann that deals with selection, motivation, organizational characteristics, recruitment, and other factors that determine whether the organization’s choice and the individual’s choice are consistent. There may be problems, or conflict, discussed in an article by Castro and Ryan.

While attempting to maintain and further enhance the organization’s position in the marketplace, the organization can undertake programs designed to assess how well it is functioning. This issue is discussed in the article on organizational diagnosis (Bussing). Another article in this section pertains to downsizing and outplacement (Cascio), a topic important and timely in today’s economic environment. Organizations are merging and downsizing in order to achieve greater economies of workforce and productivity.

Note that the above trichotimization into development, selection, and maintenance phases is not exhaustive or mutually exclusive; rather, the topics are integrated. That is, it would be difficult to consider issues of selection without considering issues of training and motivation. This means that if an organization determines that it needs to place employees into position “X,” it could do so by selecting those who already have the “X” abilities and skills; it also means their compensation would be affected by the fact that the hired are skilled. In contrast, the organization could decide to hire those individuals who demonstrate the aptitude or ability to learn and gain the abilities and skills needed to perform the tasks. This would mean that there would be greater emphasis on training those hired, which in turn might mean that their compensation could initially be lower because they

are relatively less skilled. For both types of hires, however, motivation is critical. How do organizations motivate workers to be fully productive? Does the fact that the organization provides the training to relatively less skilled employees mean that those employees will be more committed to the organization, and consequently the organization will have a lower turnover rate? These issues and questions need to be considered jointly. Although the articles in this section treat the topics separately, the challenge for the reader is to recognize the necessary integration.

#### **4. OTHER TOPICS**

An encyclopedia often provides information on all topical phrases that are used within the discipline. Those who are somewhat familiar with I/O psychology might wonder why there are no specific articles in this section on topics such as job satisfaction, incentives, test fairness, and work performance. The reason for the omission of articles on such specific topics is because their essence is conveyed in other articles appearing in this or other sections of the encyclopedia. For the topics just mentioned, for example, the reader is referred to the articles on boredom, motivational taxonomies, personnel psychology, and competence at work, respectively. The point is that this section on I/O psychology is reasonably complete and provides an overview of the major topics and issues in the field.

#### **5. FUTURE TRENDS**

An obvious point of interest for readers of this section is what the future holds for the field of I/O psychology. Our crystal ball suggests that the major topics covered in this section will remain the same, but that the nature of work—how it is performed, what it is composed of, and where it is performed—will change. That is, I/O psychologists will always be interested in how organizations should be developed, how to obtain people for those organizations, and how to maintain and promote the survival and growth of the organization. But what people do in the form of work, and how and where they perform that work is likely to change.

To elaborate, let us return to the history of the field. Reviewing the “older” textbooks, such as one by Viteles, shows that their tables of contents cover many of the topics included in this section of the encyclopedia. Among the 27 chapters of the Viteles text are chapters

on individual differences, the social foundations of industrial psychology, vocational selection, job analysis, the interview, tests for different skills and industries, safety, training methods, industrial fatigue, motives, monotony, and supervision. The topics have remained the system—the way that I/O psychologists have approached the research problems and analyzed the data have become more sophisticated and elegant.

So, what has changed and what will continue to change? The answer is that (1) work, (2) the worker, and (3) the context will change. First, the nature of work will change. It is obvious that more and more work will be driven by computer technology and that employees will be producing less manually. More and more employee work will focus on providing information and services. Thus, I/O psychologists need to be concerned with systems that can describe the tasks to be performed (job analysis and job design), identify the KSAs needed to perform those tasks (personnel psychology), and figure out how to measure effectiveness on those tasks (competence at work). Work will become more fluid; jobs will not be well-defined. Much of this work is likely to be done in groups and teams, which means emphasis on how to select and motivate team members. Also, it is likely that much of the work will be done in virtual teams. An obvious implication of all of this change is that the relationship between the worker and the client will change. There will be less direct contact and more reliance on the worker fulfilling the client's expectations without being on site to push for services.

Second, workers may interact differently with their peers and supervisors. More and more meetings may take place via teleconferencing instead of face-to-face. Accordingly, I/O psychologists will need to pay more attention to cross-cultural psychology, to topics of personality and interpersonal relations, and to work and non-work balance.

The third issue that will impact I/O psychology, context, is concerned with the globalization of doing business and multiculturalism of the workforce. I/O psychologists will need to address the performance of work in the context of the globalization of the economy with a workforce that is more culturally diverse than it was in the past. This means that workers will need to be more adaptable to where they work and with whom they work. Not only will workers be asked to interact with others from different cultures and backgrounds, but they also may be asked to work in locations foreign to them (e.g., other countries).

The challenge for the I/O psychologist will be to develop systems (selection, training, compensation,

motivation, etc.) that recognize the changes in the work that is being performed, how it is being performed, who is performing it, and in what context. Workers will need to be more flexible and agile, and work at faster paces and speeds. Workers may need to be more self-driven and work in contexts in which they are empowered and in which there will be self-learning. Workers will need to adjust to uncertainty and the temporary nature of their work. They will need to adapt to flexible work patterns. And, they may need to assume that they will have multiple careers over their life.

## 6. CONCLUSION

This section focuses on the study of behavior for those who operate within an organizational context. It presents the issues and interests of I/O psychologists as they attempt to study how work is performed, who performs it, and where it should be performed. The topics contained within this section are integrated and should be studied in relation to each other—focusing on one article without considering others will lead to inappropriate assessments, conclusions, and solutions. Hopefully, the section conveys the message that the study of work and workers is complex and that many variables need to be considered when attempting to provide solutions to work problems. The literature shows that less than 50% of the variability in individual and group performance can be explained by the variables I/O psychologists study as they attempt to explain and predict that behavior. The goal for the future is to explain more of this variability.

### See Also the Following Articles

Decision Making ■ Employment Interviewing ■ Industrial/Organizational Psychology across Cultures ■ Organizational Culture and Climate ■ Organization Development ■ Organizational Socialization ■ Personnel Selection ■ Power, Authority, and Leadership ■ Recruitment ■ Work Environments ■ Work Motivation ■ Work Role, Values Sought in the ■ Work Teams

### Further Reading

Benjamin, L. (1997). A history of Division 14 (The Society for Industrial and Organizational Psychology). In D. A. Dewsbury (Ed.), *Unification through division: Histories of the divisions of the American Psychological Association*

- (Vol. 2, pp. 459–466). Washington, D.C: American Psychological Association.
- Goldstein, I. R., & Ford, J. K. (2002). *Training in organizations* (4th ed.). Belmont, CA: Wadsworth.
- Howard, A. (Ed.) (1995). *The changing nature of work*. San Francisco, CA: Jossey-Bass.
- Ilgen, D. R., & Pulakos, E. D. (Eds.) (1999). *The changing nature of performance*. San Francisco, CA: Jossey-Bass.
- Katzell, R. A., & Austin, J. T. (1992). From then to now: The development of industrial-organizational psychology in the United States. *Journal of Applied Psychology*, 77, 803–835.
- Koppes, L. L. (1997). American female pioneers of industrial and organizational psychology during the early years. *Journal of Applied Psychology*, 82, 500–515.
- Landy, F. J. (Ed.) (1993). Early influences on the development of industrial and organizational psychology. In, *Exploring applied psychology* (pp. 81–118). Washington, D.C: American Psychological Association.
- Landy, F. J. (1997). Early influences on the development of industrial and organizational psychology. *Journal of Applied Psychology*, 82, 467–477.
- Landy, F. J., & Conte, J. (2004). *Work in the 21st Century: An introduction to industrial and organizational psychology*. New York: McGraw-Hill.
- Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.
- Munsterberg, H. (1913). *Psychology and industrial efficiency*. New York: Houghton Mifflin.
- Taylor, F. W. (1911). *Principles of scientific management*. New York: Harper.
- Viteles, M. S. (1932). *Industrial psychology*. New York: W. W. Norton.



# Injury in Sport

Albert J. Petitpas and Britton W. Brewer

Springfield College, Springfield, Massachusetts, USA

1. Introduction
  2. Psychosocial Risk Factors for Sport Injury
  3. Psychological Effects of Sport Injury
  4. Sport Injury Prevention
  5. Psychological Interventions in Sport Injury Rehabilitation
- Further Reading

## GLOSSARY

**athletic identity** The degree to which one identifies with the athlete role.

**proprioceptive** Sensory or perceptual awareness of body movement and position, particularly with respect to the limbs.

**psychosocial risk factors** Physical and psychological conditions existing prior to injury that predict the occurrence of injury.

**rehabilitation adherence** The extent to which one follows a prescribed rehabilitation protocol.

**sport injury** Physical injury sustained during participation in recreational or competitive sport activity.

**tough-mindedness** A personality orientation characterized by assertiveness, self-assurance, independence, and a tendency to place stress on oneself.

Psychological factors are important elements in most aspects of the injury process in athletes. Psychological factors have been identified in both the occurrence and aftermath of sport injury, and psychological interventions have been used to reduce injury risk

and to facilitate physical and psychological recovery from sport injury.

## 1. INTRODUCTION

Physical injury is a common by-product of sport participation. Literally millions of athletes around the world sustain one or more injuries each year, incurring billions of dollars of medical costs in the process. Although most of the injuries are minor and require minimal medical treatment, many of the injuries require extensive and costly diagnostic testing, surgery, and rehabilitation. In the short term, sport injuries can impair performance both on and off the athletic field. Athletes can be prevented from training or competing as a result of injury, and their daily activities can be disrupted by injury-related pain or disability. In the long term, certain sport injuries may place athletes at increased risk for developing arthritis and other chronic medical conditions. Some of the more severe injuries may bring about the end of athletes' sport careers and result in permanent disabilities.

Historically, injury has been considered primarily as a physical phenomenon within the sport community, with an emphasis on identifying the physical factors that cause injuries to occur and on helping athletes to recover from injuries. Unfortunately, even with all of the medical and technical advances that have been made over the past several decades in equipment, playing surfaces, and other safety features, the number of sport-related

injuries has continued to increase. This unexplained rise in injury rates has been accompanied by a growing recognition of the possible role of psychological factors in the sport injury process. Psychological factors are now thought to contribute to not only the occurrence but also the rehabilitation of sport injuries. Indeed, psychological interventions have been implemented to assist in the prevention and treatment of sport injury.

## **2. PSYCHOSOCIAL RISK FACTORS FOR SPORT INJURY**

Athletes and coaches are well aware of physical factors that can precipitate injury, including overuse, muscular tension, and trauma. Psychosocial factors can also place athletes at increased risk for injury. Over the past three decades or so, elevated life stress has emerged as a highly consistent predictor of sport injury. In American football and a variety of other sports, athletes who have recent histories of stressful life events are more vulnerable to injury than are those without such stressful experiences. Life events perceived as negative (e.g., loss of a loved one, getting fired from a job) typically have the most adverse impact on the occurrence of sport injury, although even events perceived as positive (e.g., academic achievements, new romantic relationship) can be stressful and can result in increased injury risk.

In addition to life stress, personality and coping resources are psychosocial factors that can affect susceptibility to sport injury. Injury risk appears to be elevated for athletes low in social support or high in trait anxiety, Type A behavior, negative mood, and/or tough-mindedness. Athletes who are not satisfied with the amount of technical assistance, informational guidance, and emotional comfort they receive from people important to them are at increased risk for injury. Similarly, athletes who are generally anxious, hard driving, hurried, hostile, assertive, self-assured, depressed, and/or angry are more vulnerable to injury than are those without such characteristics. Personality and coping resources can also influence the impact of life stress on the occurrence of sport injury. For example, the adverse effect of stressful life events on sport injury is especially pronounced for athletes who tend to be anxious in competitive sport situations, lack social support, and/or have difficulty coping with stress.

It is not known exactly how life stress, personality, and coping resources affect vulnerability to sport injury, although there is reason to believe that both

cognitive and physiological factors are directly involved in injury occurrence. Life events, both positive and negative, often elicit stress reactions that can cause athletes' muscles to be tense, peripheral vision to narrow, and attentional focus to diminish, all of which may place athletes at increased risk for injury during sport participation.

## **3. PSYCHOLOGICAL EFFECTS OF SPORT INJURY**

In addition to the obvious detrimental effects of sport injury on physical functioning, sport injury can have a deleterious effect on various aspects of psychological functioning. In particular, injury can have an adverse impact on the thoughts, feelings, and actions of athletes. With respect to thinking, sport injury is associated with reduced self-esteem and physical self-confidence as well as with elevated confusion. From an emotional standpoint, it is not uncommon for athletes to experience negative feelings such as anger, confusion, depression, fear, and frustration following injuries. Although clinically meaningful levels of emotional distress have been found in approximately 5 to 24% of athletes with injuries, most negative emotions tend to dissipate over the first month following the injuries. Once athletes adjust to the initial changes caused by their injuries, they are typically able to focus primary attention on their rehabilitation programs. For example, injury appears to spur athletes' use of active coping strategies such as vigorously pursuing rehabilitation goals and seeking support or information from others. As athletes become more invested in their rehabilitation regimens, their moods return to typical preinjury levels.

Athletes' psychological responses to injuries are influenced by characteristics of the athletes themselves and the situations in which they sustain their injuries, including characteristics of the injuries themselves. Athlete characteristics associated with emotional reactions to injury include age, athletic identity, and hardiness, with athletes who are younger, more strongly identified with the athlete role, and less hardy reporting greater emotional disturbance following injuries. The extent to which athletes adhere to their injury rehabilitation programs—a behavioral response to injury—is related to athlete characteristics such as internal health locus of control, pain tolerance, self-motivation, and tough-mindedness. Athletes who consider themselves

responsible for maintenance of their health, tolerate pain well, and are self-motivated and tough-minded tend to demonstrate high levels of adherence. Injury is also a prominent factor leading to dropout in youth sport participation.

With respect to characteristics of the situations in which injuries occur, athletes report greater postinjury emotional disturbance when they perceive their injuries as severe, their life stress as high, their rehabilitation progress as poor, and their social support as low. Athletes tend to stick with their prescribed rehabilitation programs better when they perceive their injuries as severe, their level of social support as high, the treatment as effective, the clinical environment as comfortable, and the rehabilitation appointments as conveniently scheduled. Emotional and behavioral responses to injury are also related to each other; that is, athletes with high levels of mood disturbance tend to adhere well to their rehabilitation programs, whereas athletes with low levels of mood disturbance tend to adhere poorly to their rehabilitation programs.

Athletes who incur sport career-ending injuries are vulnerable to considerable psychological distress, particularly if they are competing at a high level and view themselves primarily or exclusively as athletes. Intercollegiate athletes who sustain career-ending injuries tend to be less satisfied with life 5 years after the conclusion of their playing careers, particularly if they believe that they have not achieved their sport goals.

Individuals incurring catastrophic sport injuries that result in permanent functional disabilities, typically from damage to the head or spinal cord, must make profound physical, emotional, social, and lifestyle adjustments. Although catastrophic injuries are rare occurrences, they happen with enough frequency that the National Center for Catastrophic Sports Injury Research was established in 1982 to track prevalence rates and make recommendations for safety and prevention that have been implemented in most contact sports as well as in gymnastics and cheerleading.

In addition to injury severity, other aspects of the injury itself—onset, course, history, and type—can affect psychological responses to sport injury. The onset of a sport injury can be either acute or gradual. Sudden and unexpected injuries are classified as acute and often place different demands on athletes' personal coping resources and elicit a different response from their social support systems compared with nagging or overuse injuries that eventually cause athletes to stop

competing due to the gradual accumulation of physical damage or discomfort.

Sport injuries can also be classified by the course to recovery. Progressive-type injuries show a continual stepwise process toward either full recovery or permanent disabilities. The speed of the progression may vary, but the course is predictable with few surprises or unknowns. Constant-type injuries require athletes to adjust to physical impairments that remain stable or invariant over time. In episodic-type injuries, athletes must learn to cope with injuries that recur or flare up periodically and unexpectedly. Injury course is important because individuals' perceptions of control and predictability in various situations have been linked to the amount of anxiety that they manifest in those situations.

History and type of injury can also affect athletes' psychological responses to injury. Athletes who have previous histories with the same injury tend to be more knowledgeable of what to expect and may have less anxiety than do athletes who experience the injury for the first time. Athletes with injuries of a type that requires a cast or involves some form of visible impairment may experience different reactions from their social support networks compared with athletes with nonvisible injuries who are forced to answer repeated questions about why they are not competing.

The ways in which athletes respond psychologically to injury may have important implications not only for the athletes' quality of life following injury but also for their physical recovery. As shown in [Table I](#), a variety of psychological factors are related to sport injury rehabilitation outcomes. In general, athletes who recover faster or more effectively tend to be those who demonstrate a favorable mental health profile, receive adequate social support, and adopt an approach to their rehabilitation in which they maintain a positive attitude, apply psychological coping strategies, take responsibility for achieving desired rehabilitation outcomes, and adhere to their rehabilitation programs.

Athletes' readiness to resume training postinjury requires both physical and psychological assessment. Warning signs that could help to identify athletes who are not ready psychologically to return to competition include dwelling on minor somatic complaints just prior to the recovery target date, withdrawing suddenly from teammates and/or coaches, exhibiting striking changes in affect or behavior, and engaging in exaggerated storytelling or bragging about accomplishments in and out of sports.

**TABLE I**  
**Examples of Psychological Factors Associated with Sport Injury Rehabilitation Outcomes**

<i>Variable type</i>	<i>Variable</i>
Athlete factors	Age Athletic identity Gender Optimism Personality
Situational factors	Social support
Cognitive factors	Attitude toward rehabilitation Attributions for recovery Goal-setting use Imagery use Injury appraisals Pain coping
Emotional factors	Mood disturbance Psychological distress Rehabilitation anxiety
Behavioral factors	Adherence to rehabilitation Physical activity Social support seeking

#### 4. SPORT INJURY PREVENTION

Just as psychological factors can contribute to sport injury, psychological treatments can help to prevent sport injury. Initial evidence of the potential preventive impact of psychological interventions on the occurrence of sport injury was obtained almost by accident as an unintended side effect of interventions designed to enhance sport performance. Training athletes in attentional control, biofeedback, imagery, and relaxation can reduce conditions (e.g., distraction, muscular tension, perceived stress) that not only hamper sport performance but also increase vulnerability to injury. Teaching athletes psychological strategies for managing stress (due to both positive and negative life events) can lower their risk of sustaining sport injuries. In situations where athletes have experienced successful competitive performance, such interventions can help athletes to reduce their injury risk by maintaining their focus and avoiding complacency.

Developing athlete social support groups that provide forums for individuals to share their life events, both positive and negative, may buffer the amount of

stress induced by these experiences and thus moderate injury susceptibility. Peer mentoring systems within teams or athletic departments can also strengthen the overall social support system and offset the potentially adverse consequences of stress.

#### 5. PSYCHOLOGICAL INTERVENTIONS IN SPORT INJURY REHABILITATION

As shown in Table II, a variety of psychological interventions have been successfully implemented to alter the psychological responses, rehabilitation processes, and rehabilitation outcomes of athletes with injuries, particularly knee injuries. As with interventions targeted at preventing sport injuries, the reasons why the interventions work are not fully understood. However, most of the interventions are thought to increase motivation for rehabilitation, and this may in part account for the therapeutic effects of the interventions. It has also been proposed that biofeedback enhances athletes' awareness of their body movements and positioning and that relaxation and imagery influence physiological processes central to healing such as immune/inflammatory responses and the regeneration and repair of body tissues.

**TABLE II**  
**Psychological Interventions in Sport Injury Rehabilitation**

<i>Intervention</i>	<i>Therapeutic outcome</i>
Biofeedback	Knee/Leg strength Electromyogram (EMG) output Range of motion
Goal setting	Rehabilitation adherence Rehabilitation self-efficacy Knee strength
Relaxation/Imagery	Pain Reinjury anxiety Knee strength
Positive self-talk	Leg strength
Multimodal treatment package	Physical recovery Pain Anxiety Positive mood Positive attitude

**See Also the Following Articles**

Goal Setting and Achievement Motivation in Sport  
■ Overtraining and Burnout in Sports ■ Performance  
Slumps in Sport: Prevention and Coping

**Further Reading**

- Brewer, B. W. (2001). Psychology of sport injury rehabilitation. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 787–809). New York: John Wiley.
- Pargman, D. (Ed.). (1999). *Psychological bases of sport injuries* (2nd ed.). Morgantown, WV: Fitness Information Technology.
- Petitpas, A., & Danish, S. J. (1995). Caring for injured athletes. In S. M. Murphy (Ed.), *Sport psychology interventions* (pp. 255–281). Champaign, IL: Human Kinetics.
- Wiese-Bjornstal, D. M., Smith, A. M., Shaffer, S. M., & Morrey, M. A. (1998). An integrated model of response to sport injury: Psychological and sociological dimensions. *Journal of Applied Sport Psychology*, *10*, 46–69.
- Williams, J. M. (2001). Psychology of injury risk and prevention. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 766–786). New York: John Wiley.



# Intelligence and Culture

Fons J. R. van de Vijver

Tilburg University, Tilburg, The Netherlands

1. Culture and Intelligence: An Emerging Field
  2. Three Views on the Relationship between Intelligence and Culture
  3. The Measurement of Intelligence in a Cross-Cultural Context
  4. Formal Versus Informal Studies of Intelligence
  5. Everyday Definitions of Intelligence
  6. Conclusion
- Further Reading

## GLOSSARY

**absolutism** The viewpoint that holds that cognitive processes are the same across the globe and that emphasizes the context-independent nature of cognitive processes and basic intellectual processes.

**bricoleur** A metaphor often underlying studies in the informal tradition; a bricoleur (i.e., “jack-of-all-trades”) is not interested in a solution that is correct as defined by an external standard (which is typically the case in the formal tradition) but rather is interested in a solution that solves a practical problem (e.g., a leaking roof) using available means.

**culture-fair test** A test that has been designed in such a way that culture does not differentially influence performance on the test (i.e., item formats are so easy that everyone is supposed to not experience any problem with the test); the concept has been abandoned as unattainable.

**culture-free test** A test that has been designed in such a way that culture does not influence performance on the test; the concept has been abandoned as unattainable.

**everyday intelligence** The study of the relationship between culture and intelligence by the observation of intellectual processes in everyday behavior (e.g., arithmetic behavior of shoppers who compare bargains).

**formal tradition** The study of the relationship between intelligence and culture by the administration of psychometrically adequate tests in different cultures; the scientific approach is seen as the normative model of good problem solving, with an emphasis on the solution of formalized problems that are unlikely to be met in everyday life and on the correctness of solutions.

**informal tradition** The study of the relationship between intelligence and culture by studying how people solve everyday intellectual problems; the “bricoleur” (i.e., jack-of-all-trades) is seen as the implicit model of problem solving.

**nature–nurture debate** The scientific controversy about the question as to what extent cross-cultural differences in scores on intelligence tests are due to genetic factors (nature) or environmental factors (nurture).

**relativism** The viewpoint that holds that intelligence comes into existence in a specific cultural context and that intelligence and culture are so closely linked that any attempt to compare intelligence across cultures is futile.

**universalism** The viewpoint that basic cognitive processes are universal but that manifestations of these processes may well vary across cultures.

This article discusses three views on the relationship between culture and intelligence. The first, called the relativist viewpoint, holds that intelligence comes into existence in a specific cultural context and that intelligence and culture are so closely linked that any

attempt to compare intelligence across cultures is futile. The second, called the absolutist viewpoint, maintains that cognitive processes are the same across the globe and emphasizes the context-independent nature of cognitive processes and basic intellectual processes. The third, called the universalist viewpoint, also holds that basic cognitive processes are universal but argues that manifestations of these processes may well vary across cultures. Assessment has always been an important part of the work on intelligence. This article provides a brief history of cross-cultural assessment of intelligence, ending with an overview of what are currently seen as major issues. Then, it presents a description and examples of two kinds of studies of the relationship between culture and intelligence: formal (usually based on a comparison of scores on intelligence tests) and informal (based on the observation of everyday behavior). Then, the article discusses everyday definitions of intelligence and argues that, particularly in non-Western countries, social aspects such as obedience and social skills are seen as part of intelligence alongside reasoning and memory. The article concludes that the universalist position, according to which the basic constituents of intelligence are panhuman, is best supported. However, within predefined limits, cultures define the way in which these constituents are merged into ability patterns.

## **1. CULTURE AND INTELLIGENCE: AN EMERGING FIELD**

It was more than a century ago that the French psychologist Alfred Binet was asked to devise a test that could be used to inform teachers about which pupils have learning difficulties. Tests to predict school performance nowadays are known as intelligence tests and are one of the most successful and widespread applications of psychological assessment. Intelligence test scores were, and still are, the best predictors of success in school and at work. The success has undoubtedly added to the received view that intelligence tests measure intelligence both within and across cultures. However, on closer scrutiny, the relationship between an intelligence quotient (IQ) score and the theoretical concept of intelligence turns out to be complex. At the beginning of the 20th century, Porteus set out to measure the cognitive skills of various cultural groups. He administered his Maze Test, a paper-and-pencil instrument in which the testee must find a way out of a schematically drawn

maze, to various cultural groups, including the Bushmen in the Kalahari Desert. Porteus concluded that the Bushmen's visuospatial skills (an important aspect of intelligence) were poorly developed, obviously not realizing that his conclusion clearly contradicted the superb tracking skills of Bushmen in the relatively cueless environment of the Kalahari.

Two further examples are presented to illustrate the close link between intelligence and culture. In the cognitive anthropological literature, there are various examples of expert cognitive performance. For example, the Trukese of the Pulawat atoll in the South Pacific have remarkable navigating skills; they sail over hundreds of miles of open ocean without compasses, sextants, or star tables. The second example comes from studies on Quranic schooling. Some Muslims learn (parts of) the Scripture by heart. In particular, when the learner does not speak Arabic (the language of the Scripture), massive training is needed. This is done by using an incremental learning method in which the first word is learned first, followed by the first and second words, followed by the first, second, and third words, and so forth. Studies have shown that this training leads to improvements of these people in learning in an incremental paradigm. However, their memory skills are not remarkable on tasks that do not require this type of learning. Cultural practice implies frequent exposure to a very specific task, and this leads to the advanced development of a skill.

## **2. THREE VIEWS ON THE RELATIONSHIP BETWEEN INTELLIGENCE AND CULTURE**

With some alleged simplification, one could say that in cross-cultural psychology, three kinds of views can be found on the relationship between intelligence and culture: relativist, absolutist, and universalist.

### **2.1. Relativist Viewpoint**

According to the relativist viewpoint, intellectual processes are closely linked to their cultural context and can be studied only with explicit reference to this context. The visuospatial skills of the Bushmen play an important role in tracking game needed for their everyday survival. The skills of Western individuals are developed to survive in a completely different environment. Therefore, the skills are embedded in different

contexts in both cultures. As a consequence, the skills are not comparable across these two cultures, and it is impossible to design an instrument that can be used to compare these skills across the cultures. Furthermore, there is no need to compare these skills given that all cultural groups are well adjusted to their natural habitats; one cannot say that Bushmen are better (or worse) adapted to the desert than Westerners are to the urban industrialized context. In this view, as socialization progresses, cognitive skill patterns of individuals from different cultures become increasingly dissimilar. Each culture produces its own type of intelligence that is well suited for everyday functioning in that specific cultural context. In this framework, it is important to identify these patterns: How does Bushman intelligence differ from Japanese intelligence?

### 2.2. Absolutist Viewpoint

The absolutist viewpoint is often associated with a small group of influential researchers such as Arthur Jensen, the late Hans Eysenck, Richard Lynn, and Philip Rushton. It is the opposite of the relativist viewpoint; absolutists argue that cognitive processes are the same across the globe. They emphasize the context-independent nature of cognitive processes. Basic cognitive operations, such as transitivity (e.g., "If John is taller than Mary and Mary is taller than Peter, then John is taller than Peter") and syllogisms (e.g., "If all humans are mortal and Socrates is a human, then Socrates is mortal"), are independent of their context of occurrence. Because of this independence, tests can be designed in such a way that they have a panhuman validity. A test of intelligence that is both reliable and valid in San Francisco will also be valid in the Australian outback and in Sub-Saharan Africa (assuming that the test does not measure information that is available only to a single cultural group). It is clear that with such a strong confidence in the value of intelligence tests, far-reaching conclusions about intellectual skills of various cultural groups can be drawn. Absolutists often assume that these differences are biologically rooted; not surprisingly, they often use the term "race" to describe these differences. Jensen has provided much evidence showing that European Americans show higher IQ scores than do African Americans (approximately 15 IQ points on average). Eysenck has argued that if the races are split up according to skin color, the "yellow race" has the highest IQ scores, followed by the "white race" and then the "black race."

### 2.3. Universalist Viewpoint

The universalist viewpoint holds that the same basic cognitive structure of humankind is shared by all cultural groups. Cognitive skills and structures, such as reasoning and memory, are present in all cultures, and these skills are called universal. Still, it is clear from cross-cultural studies that culture exerts a powerful influence on intellectual functioning. Culture defines the context of intellectual functioning by determining what, when, and how individuals will learn. A universalist perspective on the navigational skills of the Trukese would hold that these skills consist of universal building blocks that are combined and refined in a lengthy learning process, with the skill being learned by observing experts. In a universal perspective, there are no different types of intelligence. However, cultures differ in terms of which building blocks are used to a greater or lesser extent, and this could lead to cross-cultural differences in mean scores on the more extensively trained skills. Cultures also differ in the specific skills taught (e.g., navigational skills among the Trukese, computer games in Western societies). Such complex skills are based on the universal building blocks that are applied to cultural knowledge.

It is probably fair to say that the universalist viewpoint is the most popular one in cross-cultural psychology. A basic problem of the relativist position is that all cognitive processes are assumed to be context bound. If one is interested in a comparison of cognitive processes across various cultures, the position that these processes are not comparable by definition is not very fruitful. The main problem with this position is the implied overstatement of what could be a valid position in some domains of cognition. Take short-term memory span as an example. A number of items, usually digits, are presented at a rate of one item per second, and the respondent must repeat the items in the same order. The test starts with short series and then continues until the respondent can no longer recall the items correctly. There are strong indications that at least within school-going populations, short-term memory span is relatively invariant across cultures. A relativist viewpoint can quickly become counterproductive when studying these basic universalities in human cognitive functioning. The other extreme position, absolutism, can be equally counterproductive. It is difficult to understand how one could measure visuospatial skills of Bushmen and New Yorkers with one and the same test that gives a valid reflection of their skill levels and that could give a good prediction of real-life behavior in both

environments in which these skills are relevant. In general, it is naive to assume that Western cognitive tests “travel well” and that they easily allow for direct numerical score comparisons of different cultural groups. The popularity of the universalist viewpoint can be easily derived from the shortcomings of the relativist and absolutist viewpoints. The empirical record shows clearly that neither the culture-bound nature of the relativist viewpoint nor the essential denial of cultural influences on cognitive processes is strongly supported. The universalist viewpoint, which is more prudent, has the most face validity.

### **3. THE MEASUREMENT OF INTELLIGENCE IN A CROSS-CULTURAL CONTEXT**

The measurement of intelligence has always played an important role in the literature. If one works in a relativist tradition, one must make sure that his or her assessment of intelligence is adequate for the cultural context. A measure of vocabulary should provide a score that gives an adequate picture of the testee’s knowledge of the lexicon. Within the universalist tradition, one must deal with an additional question: Is a test (possibly translated or adapted) adequate to be used in different cultures? It cannot be taken for granted that a test (or its items) developed in one cultural context (typically a Western context) can be simply translated and employed in a different cultural context without any problems. For example, words in a vocabulary subtest might be inadequate or relatively easy or difficult for some cultural groups as compared with other cultural groups. In a Dutch study involving Turkish and Moroccan migrant children and Dutch autochthonous children, it was found that the word “bacon” was much more difficult for the migrant groups, presumably due to the food taboo on pork meat in Islam. The adequacy of a test to measure intelligence in a cross-cultural context cannot be assumed but must be demonstrated.

The question of whether intelligence tests can be applied in different cultural contexts is answered in different ways by relativists, universalists, and absolutists. Relativists argue that basic cognitive processes come into existence in a specific cultural context and that process and context constitute each other; in this view the search for a cross-culturally applicable intelligence test is needless and pointless. Within the absolutist tradition, and particularly within the universalist

tradition, the view is more common that there are good and bad intelligence tests, from a cross-cultural perspective, and that it is possible in principle to develop cross-culturally adequate intelligence tests.

Although the terms of relativism, absolutism, and universalism are recent, the theme of how to design intelligence tests that are adequate in various cultures is old. It was recognized a long time ago that familiarity with test stimuli (e.g., figures and words used in a test) could play a role in assessment; that is, people more familiar with the stimuli can be expected to obtain higher scores. Back in the 1940s, Cattell proposed using very simple stimuli in cognitive tests to deal with the problem that various groups were not equally familiar with test material. This led to the development of “culture-free tests.” However, the assumed independence of test performance and, among other things, socioeconomic status could not be demonstrated. Subsequently, “culture-fair tests” were developed. Not surprisingly, the existence of culture-free and culture-fair tests has come under critical scrutiny.

Tests that claim to be culture free or culture fair must deal with two related issues. The first is the problem of differential stimulus familiarity. If cognitive tests are administered to cultural groups with widely different educational backgrounds, differences in stimulus familiarity are nearly impossible to avoid and overcome. The second issue is the implicitness of cultural knowledge in tests. Tests and testing situations imply much cultural knowledge, even when this knowledge is not the topic of a test. For example, a situation in which an individual asks questions with answers he or she knows and expects the testee to answer the questions without the help from anyone else is uncommon in everyday life, with experience with this situation being more common among schooled populations. The implicit references to cultural knowledge and customs in intelligence tests are numerous. Indeed, it could be argued that these references are impossible to avoid. It has been argued repeatedly that cross-cultural differences in intelligence test scores that are often reported are strongly influenced by the differential knowledge of the culture of the test. Individuals who share a cultural background with the test developer are more likely to obtain higher scores than are people from other cultural groups.

During recent decades, various guidelines have been formulated to describe best practices in the design of cross-cultural tests and preparing translations. For example, the International Test Commission defined a set of 22 guidelines that are useful when designing cross-cultural instruments.

Various procedures can be employed to examine the adequacy of an instrument, for example, an analysis of the contents of a test by a person (or group) with a thorough knowledge of the target culture. This person should be able to answer the question of whether the cognitive constructs of the test (e.g., memory, visuospatial skills) could be measured in this group with this instrument. In addition, various statistical techniques can be applied. These techniques require that data be collected in various cultural groups. The most frequently employed techniques address the question of whether it is reasonable to assume that the data obtained in various cultural groups can be treated statistically as coming from a single population: Can the two cultural groups be seen psychologically as belonging to the same (imaginary) statistical population? A frequently employed technique is factor analysis that examines the intertest correlations. If subtests of different tests measure entirely different psychological processes in different groups, it is very likely that test scores are interrelated in different ways across these groups. Suppose that two tests have been administered in each of two countries. Both of the tests measure reasoning in one country, whereas one test measures reasoning and the other test measures memory in the other country. It can be expected that people in the first country who score high on the first test will also score high on the second test (leading to a positive correlation between the tests), whereas a different pattern is expected in the second country because the people with high reasoning skills are not necessarily also the people with high memory skills. The intertest correlations will be higher in the first country than in the second country. So, a study of correlations of (sub)tests can yield valuable information about the psychological composition of an intelligence test in different countries.

#### 4. FORMAL VERSUS INFORMAL STUDIES OF INTELLIGENCE

There are two research traditions in cross-cultural psychology to study the relationship between culture and intelligence: formal and informal (Table I). The formal and informal research traditions are comparable to (but not completely reducible to) the distinction between the universalist and relativist viewpoints, respectively.

In formal studies of intelligence, the scientific approach is seen as the normative model of good problem solving. There is an emphasis on the solution of

**TABLE I**  
**Differences of the Formal and Informal Traditions of the Study of Intelligence**

<i>Formal tradition</i>	<i>Informal tradition</i>
Closed problem spaces	Open problem spaces
Deterministic problems	Probabilistic problems
Formalized artificial problems	Problems with a high ecological validity
Focus on correctness of solution	Focus on practical value of the solution (uncertainty reduction)
Context independence	Context dependence
The scientist as the model of the problem solver	The bricoleur as the model of the problem solver
Product oriented (psychometric approach) or process oriented (Piagetian approach)	Process oriented
Cross-cultural comparisons of test performances	Intracultural studies
Algorithmic solutions	Heuristic solutions
Brief assessment of many problems	In-depth assessment of a single problem
Solution requires conceptual theoretical knowledge	Solution requires procedural practical knowledge

*Source.* Van de Vijver and Willemsen (1993).

formalized problems that are unlikely to be met in everyday life and on the correctness of solutions (e.g., "What is the capital of Poland?"). Nearly all intelligence tests are based on the formal tradition. Studies in this tradition have enlarged our insight on the relationship between culture and intelligence. First, administration of intelligence tests in various cultures has provided impressive evidence that the structure of intelligence is universal. The structure of cognitive abilities constitutes the universal "building blocks" that are combined by cultures in specific skill patterns useful in these cultures. (In a currently popular model of intelligence developed by Carroll in 1993, these cognitive abilities consist of fluid intelligence, crystallized intelligence, memory and learning, visual perception, auditory perception, fluency, and speed of cognitive and simple decision processes.) This finding provides a strong empirical argument against the view that all cognitive processes are culture bound. The importance of the panhuman structure of intelligence should not be underrated. It means that basic cognitive processes involved in memory are very likely to be the same across various cultures, even though

cultures have an impact on what kind of information is stored and retrieved. Views have been expressed that there are various kinds of intelligence across the globe and that, for example, Libyan intelligence is different from Japanese intelligence. Findings on the panhuman identity of intelligence show that the notion of multiple kinds of intelligence culturally cannot imply major differences in structure. The statement could also be interpreted to mean that Libyans and Japanese do not show the same scores on various subtests. The latter interpretation brings us to a second line of research in the formal tradition: determining and interpreting cross-cultural differences in mean scores on intelligence tests.

It is well established that cultures differ in mean scores on intelligence tests. The nature of these differences is debated, ranging from genetic differences to schooling differences. Furthermore, the quality of these tests to assess intelligence in groups for which such tests have not been developed and validated has been questioned. Although this controversy will not be solved in the near future, it is clear that the competition among these three competing explanations is unlikely to have a single winner. For instance, consider the evidence in favor of the genetic position. There are twin studies in behavior genetics in which a large proportion of individual differences in intelligence scores is found to be caused by genetic differences. Genes are nearly always found to be more important than the environment as determinants of score differences. However, the generalization of this finding to a cross-cultural context is not simple. The assessment problems discussed previously will also affect the size of the cross-cultural score differences observed. So, there is a problem of estimating the relative contributions of all relevant sources of variation. The current state of knowledge, and of the level of assessment procedures, is insufficient to deal with the question of the relative contributions in a decisive way.

A recent project by Georgas and colleagues compared data obtained with the third version of the Wechsler Intelligence Scale for Children (WISC-III) from various countries: Canada, Dutch-speaking countries (The Netherlands and Flanders), German-speaking countries (Germany, parts of Switzerland, and Austria), Greece, Japan, Lithuania, South Korea, Slovenia, Sweden, and the United States. The WISC-III test, one of the most widely employed intelligence tests for children, contains various subtests and yields an overall IQ score. The most salient finding of the study was the absence of major mean score differences on any subtest. No subtest showed a difference of more than 15 IQ points between the smallest and largest scores; in fact, the means of

most countries on most subtests differed by less than 5 IQ points. The authors did not find the major differences that had been reported previously. One of the reasons for finding only small differences may be related to the sampling procedure. The samples in all countries were carefully drawn so as to represent the national populations of that age as accurately as possible, whereas older studies usually did not work with random samples. Although the mean differences of the WISC-III scales were not large, their patterning was meaningful. Countries that are richer and that spend more money on education tended to show higher scores.

The informal tradition is known by various names such as “everyday cognition,” “indigenous cognition,” and “practical intelligence.” The bricoleur (i.e., “jack-of-all-trades”) is the implicit model of problem solving. A bricoleur is not interested in a solution that is correct as defined by an external standard (which is typically the case in the formal tradition) but rather is interested in a solution that solves a practical problem (e.g., a leaking roof) using available means. Various studies have examined how individuals solve everyday cognitive problems such as mental arithmetic by shoppers who compare prices of products when both prices and weight of brands differ, planning behavior by cooks, and estimation by carpenters of how much wood is needed for making a table.

Results of research in the informal tradition show some consistent findings. Contrary to what might be expected, there is little or no relation between performances in informal tasks and those in formal tests. Street vendors who are very skillful in addition and multiplication when dealing with numbers familiar to them do not always show high levels of performance on tests of mental arithmetic that cover areas and numbers less familiar to them. Second, there is little transfer of cognitive skills from a familiar problem-solving situation to an unfamiliar one. This problem is well known in schooling; when confronted with a new problem, students do not recognize it as an application of a solution procedure they know. Third, there are remarkable differences between experts and novices regarding accuracy and strategies they use in problem-solving tasks.

## 5. EVERYDAY DEFINITIONS OF INTELLIGENCE

The most frequently quoted definition of intelligence is that given by Boring in 1923: “Intelligence is what an

intelligence test measures.” This definition is nearly circular. A listing of the skills assessed by the subtests of an intelligence test, such as memory and reasoning, is not an informative definition of intelligence. The long tradition of theorizing and research about intelligence has not produced a satisfactory definition of the concept. Yet everyone seems to have an implicit idea of what the concept entails. During recent decades, an interesting series of studies of what laypeople mean by “intelligence” have been conducted, and many of these were done in infrequently studied groups (e.g., some rural groups in Kenya and Tanzania). In these studies, informants were asked to indicate the characteristics of both smart and dull children, to describe characteristics and behaviors of intelligent persons in their environment, and so forth. Basic constituents of Carroll’s model, such as reasoning, memory, and speed, seem to be viewed as core elements of intelligence in many cultures. However, social aspects of intelligence, usually not considered to be salient aspects of the concept in psychological theories of intelligence, are important in everyday definitions. Obedience, knowing one’s place in the family, knowing when to speak and when not to speak, and performing a task in the household without being asked were mentioned as examples of activities deemed to be relevant for intelligent children.

Recent developments in psychology can be seen as broadening the concept of intelligence and as being in line with this more context-informed conceptualization of intelligence. Sternberg introduced the concept of “successful intelligence,” and Salovey introduced the concept of “emotional intelligence.” Both of these concepts include elements that are hardly covered by classical theories, for example, creativity and empathy. The litmus test of these new tests is their power to predict important real-life criteria such as predicting which children will do well in school and which job applicant will become a good manager. The future will tell us whether predictions of real-life behavior can be improved by broadening the measures.

## 6. CONCLUSION

The relationship between intelligence and culture was not exactly love at first sight. Much intelligence research has been carried out from the implicit assumption that findings done in a single culture, typically a Western country, have panhuman validity. More recent studies have challenged this view. There is

growing support for the view that intelligence is always situated in a specific cultural context and that cross-cultural variation is an important feature of intelligence. Yet the cross-cultural variation is not endless. There is important evidence for the universality of the basic processes. Cultures define how these building blocks are combined to produce culture-relevant skills. When differences among cultures are large, the ability patterns may show quite a bit of variation. Many examples can be found in the literature of cognitive specializations that are part of regular socialization patterns (e.g., the learning of tracking skills among the Bushmen). An important task for the future of intelligence theory and research is to combine the patterns of cross-cultural differences and similarities into one or more overarching models specifying which aspects of intelligence are influenced by culture (thereby also implicitly indicating which areas are not influenced) and how these aspects are influenced. It is only by combining models and findings of the various research traditions that the universalities and cultural specificities of intelligence can be appreciated.

### See Also the Following Articles

Intelligence Assessment ■ Intelligence, Emotional ■ Intelligence in Humans

### Further Reading

- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.
- Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.
- Georgas, J., Weiss, L., Van de Vijver, F. J. R., & Saklofske, D. H. (Eds.). (2003). *Cultures and children’s intelligence: A cross-cultural analysis of the WISC-III*. New York: Academic Press.
- Greenfield, P. M. (1997). You can’t take it with you: Why ability assessments don’t cross cultures. *American Psychologist*, 52, 1115–1124.
- Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). New York: John Wiley.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

- Light, P., & Butterworth, G. (Eds.). (1992). *Context and cognition: Ways of learning and knowing*. London: Harvester Wheatsheaf.
- Mishra, R. C. (2001). Cognition across cultures. In D. Matsumoto (Ed.), *Handbook of culture and psychology* (pp. 119–135). Oxford, UK: Oxford University Press.
- Schliemann, A., Carraher, D., & Ceci, S. J. (1997). Everyday cognition. In J. W. Berry, P. R. Dasen, & T. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology* (2nd ed., Vol. 2, pp. 177–216). Boston: Allyn & Bacon.
- Serpell, R. (1993). *The significance of schooling: Life-journeys in an African society*. Cambridge, UK: Cambridge University Press.
- Serpell, R. (2000). Intelligence and culture. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 549–580). Cambridge, UK: Cambridge University Press.
- Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology*, 49, 479–502.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- Van de Vijver, F. J. R. (1997). Meta-analysis of cross-cultural comparisons of cognitive test performance. *Journal of Cross-Cultural Psychology*, 28, 678–709.
- Van de Vijver, F. J. R., & Willemssen, M. E. (1993). Abstract thinking. In J. Altarriba (Ed.), *Culture and cognition* (pp. 317–342). Amsterdam: North-Holland.



# Intelligence Assessment

Roberto Colom

Universidad Autónoma de Madrid, Madrid, Spain

1. What is Intelligence?
2. The Measurement of Intelligence
3. Intelligence Assessment Tools
4. Interpretation: Some Words
5. Recent Advances  
Further Reading

## GLOSSARY

**cognitive ability** An ability is a cognitive ability if (a) an individual's performance does not depend on any particular sensory or motor system and (b) individual differences in the ability are not correlated with sensorimotor measures.

**cognitive system design** An approach that includes a conceptual and procedural framework to interface item design principles with test validity; the conceptual framework broadens construct validation such as cognitive research results as supporting data, whereas the procedural framework includes several steps in which cognitive theory is integrated into test development.

**g** Psychometric construct resulting from the positive correlation among several diverse ability tests; *g* is what is shared or common to all of the different measures of intelligence that form a positive manifold.

**intelligence** General mental ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience.

**intelligence quotient (IQ)** A score on a test that has certain characteristics, the test's raw score is usually standardized in a large representative sample, and the standardization is done within narrow age intervals, so that the standardized

scores (IQ) will have the same mean and standard deviation at every age level.

In this article, intelligence is first defined and the structure of human cognitive abilities is described. Second, well-known data regarding the reliability and validity of intelligence measures are presented. Third, several remarkable intelligence assessment tools are briefly described, noting a movement from measures that are not derived from theory to measures that are theory driven. Finally, the cognitive system design approach is described as one noteworthy example of a recent advance in intelligence assessment.

## 1. WHAT IS INTELLIGENCE?

Intelligence is a very general mental capability involving the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience. Contrary to popular belief, psychologists are nearly unanimous in their view that reasoning and problem solving are crucial landmarks of intelligent behavior.

Human cognitive abilities can be organized hierarchically on three levels or strata. On Stratum I, there are specific cognitive abilities such as general sequential reasoning, language development, associative memory, spatial relations, speech sound discrimination, ideational fluency, numerical facility, and choice reaction

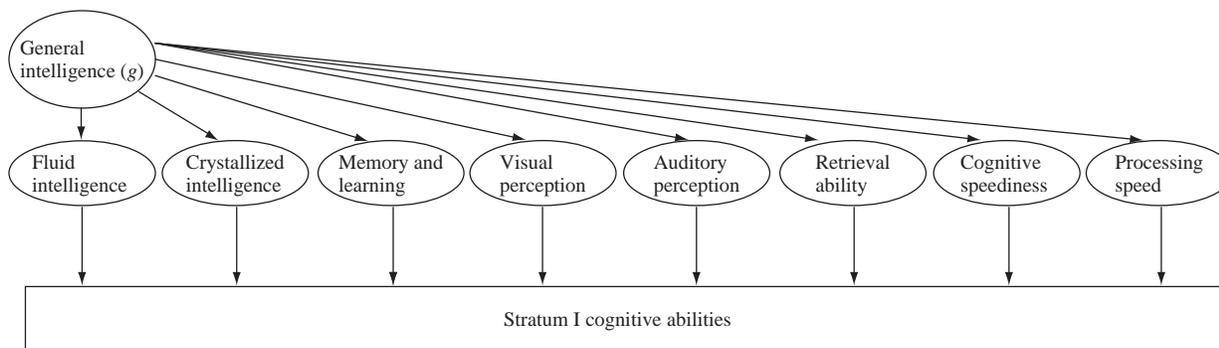


FIGURE 1 Hierarchical organization of cognitive abilities.

time. On Stratum II, there are several broad cognitive abilities such as fluid intelligence, crystallized intelligence, general memory and learning, broad visual perception, and broad auditory perception. At the top of the hierarchy, there is *g* or general intelligence. *g* accounts for half of the variability in intelligence in the general population. Figure 1 displays this hierarchical organization of cognitive abilities.

## 2. THE MEASUREMENT OF INTELLIGENCE

Intelligence tests are among the most reliable and valid of all psychological tests and assessments. Table I shows reliability coefficients of some of the most employed intelligence tests around the world. In addition to those reliability indexes, intelligence measured by standardized tests is the most stable psychological factor. Table II shows test–retest reliability coefficients taken from several classical longitudinal studies. These coefficients demonstrates that intelligence is a very

stable human trait. Furthermore, intelligence measured by standardized tests predicts more than 60 social phenomena. Table III shows some examples.

There are different types of intelligence tests, but they all measure the same intelligence (i.e., *g*). Some

TABLE II  
Correlations Computed in Several Longitudinal Studies between a Test and a Retest Measure

Test (Years)	Retest (Years)	Correlation coefficient
2	15	.78
11	77	.73
14	42	.68
17	48	.83
19	61	.78
25	65	.78
30	43	.71
50	70	.90

TABLE III  
Correlations between Intelligence Test Scores and Several Social Correlates

Correlate	Correlation coefficient
Crime	-.20
Income	.31
Health	.40
Academic success	.50
Job success	.51
Differentiation of dementia from controls	.52
Obtained level of education	.55
Job success	.63

TABLE I  
Reliability Coefficients of Several Intelligence Tests

Intelligence test	Reliability coefficient
Stanford–Binet IV	.98
WPPSI-R	.96
WISC-III	.94
WAIS-III	.98
Woodcock–Johnson III	.97
Cognitive Assessment System	.96
Kaufman Adolescent and Adult Intelligence Test	.97

**TABLE IV**  
**Criteria Associated with IQ Levels**

IQ level	Criteria
40	Adult can mow lawns, do simple laundry
50	Adult can do simple carpentry, domestic work
60	Adult can repair furniture, harvest vegetables, assist electrician
75	Adult can keep small store, perform in orchestra
90	Adult can perform jobs requiring some judgment
100	Average for total population
105	Fifty-fifty chance of passing in academic high school curriculum
110	Fifty-fifty chance of graduating from college
115	Mean of freshmen in typical 4-year college
120	Mean of college graduates
130	Mean of persons receiving PhD degrees

use words or numbers and require specific cultural knowledge. Others do not and instead use shapes or designs and require knowledge of only simple universal concepts. The measurement of intelligence through a standardized intelligence quotient (IQ) test arranges people. The spread of people along the IQ continuum, from low to high, can be represented by the normal curve. Table IV shows some examples of IQ levels for various external criteria.

Psychological science has several assessment devices with high-quality standards. The remainder of this article presents some of the most remarkable intelligence tests.

### 3. INTELLIGENCE ASSESSMENT TOOLS

Intelligence is currently assessed through tests that are not derived from theory and through tests that are theory driven. The Wechsler scales (Wechsler Preschool and Primary Scales of Intelligence–Revised [WPPSI-R], Wechsler Intelligence Scale for Children–Third Edition [WISC-III], and Wechsler Adult Intelligence Scale–Third Edition [WAIS-III]) are examples of the former type of tests, whereas the Das–Naglieri Cognitive Assessment System (CAS) and the Woodcock–Johnson III (WJ III) are examples of the latter type of tests. Those tests can be described in terms of theoretical background, the measurement

scales included, and psychometric properties such as reliability and validity.

### 3.1. The Wechsler Scales

David Wechsler supports the definition of intelligence as a global entity. The Wechsler scales measure intelligence from 3 to 89 years of age. The latest versions are the WPPSI-R for preschool and primary-grade children, the WISC-III for elementary and high school children, and the WAIS-III for older adolescents through old age.

#### 3.1.1. Wechsler Preschool and Primary Scale of Intelligence–Revised

The WPPSI-R battery provides a full-scale IQ but also has separate verbal and performance scales. The verbal subtests include information, comprehension, arithmetic, vocabulary, similarities, and sentences. The performance subtests include object assembly, block design, mazes, picture completion, geometric design, and animals/pegs. The battery allows the examiner to assist the child on early items to ensure that the child understands the test requirements.

The overall verbal IQ, performance IQ, and full-scale IQ are standard scores with the mean set at 100 and the standard deviation (SD) set at 15. The range of possible WPPSI-R full-scale IQs is 41 to 160. Standard scores are provided for the separate subtests, each with a mean of 10 and an SD of 3. The battery was standardized in the United States on 1700 children from 3 years to 7 years 3 months of age and matched 1986 census bureau estimates on the stratification variables of sex, race, geographic region, parental occupation, and parental education.

The internal consistency coefficients were .95 for the verbal IQ, .92 for the performance IQ, and .96 for the full-scale IQ. The coefficients for the individual performance subtests ranged from .63 for object assembly to .85 for block design, with a median coefficient of .79. For the individual verbal subtests, the values ranged from .80 for arithmetic to .86 for similarities, with a median coefficient of .84. The test–retest coefficient for the full-scale IQ was .91 for 175 children tested twice with a time interval of 4 weeks.

The battery's manual provides several validity studies, factor-analytic results, research overviews, and interpretive tables that provide a great amount of information.

### 3.1.2. Wechsler Intelligence Scale for Children—Third Edition

The WISC-III battery offers standard scores on four factors: verbal comprehension, perceptual organization, freedom from distractibility, and processing speed. The first two factors consist of verbal subtests, and the latter two factors consist of performance subtests. The verbal mandatory subtests are information, similarities, arithmetic, vocabulary, and comprehension. The performance mandatory subtests are picture completion, picture arrangement, block design, object assembly, and coding. There are verbal and performance complementary subtests: digit span for verbal and mazes and symbol search for performance. Processing speed results from symbol search and coding. Freedom from distractibility results from arithmetic and digit span. The range of possible WISC-III full-scale IQs is 40 to 160.

The battery was standardized in the United States on 2200 children 6 to 16 years of age and was stratified by age, sex, race, geographic region, and parents' education. The average reliabilities across the age groups were .95 for the verbal IQ, .91 for the performance IQ, .96 for the full-scale IQ, .94 for the verbal comprehension index, .90 for the perceptual organization index, .87 for the freedom from distractibility index, and .85 for the processing speed index.

### 3.1.3. Wechsler Adult Intelligence Scale—Third Edition

The WAIS-III battery was designed to parallel the WISC-III. The WAIS-III consists of verbal, performance, and full-scale IQs as well as indexes on four factors. There are seven verbal subtests and seven performance subtests. Three factors have the same names as do WISC-III factors: verbal comprehension, perceptual organization, and processing speed. The fourth factor is called working memory. In addition to arithmetic and digit span, the working memory factor includes the letter–number sequencing subtest.

The WAIS-III has six regular verbal subtests and five mandatory performance subtests. The verbal subtests are vocabulary, similarities, arithmetic, digit span, information, and comprehension. The performance subtests are picture completion, picture arrangement, block design, matrix reasoning, and digit–symbol coding.

The battery was standardized in the United States on 2450 adults (selected according to 1995 census data) and was stratified according to age, sex, race, geographic

region, and educational level. The participants were divided into 13 age groups, ranging from 16–17 to 85–89 years, with each group consisting of 100 to 200 people. The average split-half reliability coefficients across the age groups were .97 for the verbal IQ, .94 for the performance IQ, and .98 for the full-scale IQ. The average individual subtest reliabilities ranged from .93 for vocabulary to .70 for object assembly, with a median coefficient of .85. The stability coefficients were .95 for verbal IQ, .90 for performance IQ, and .97 for full-scale IQ (those coefficients were computed from 394 participants tested twice with a time interval of 5 weeks).

The range of possible WAIS-III full-scale IQs is 45 to 155. The WAIS-III computes scaled scores for each individual based exclusively on chronological age. The use of four indexes derives from the results of factor analysis. Several factor analyses have shown that the Wechsler performance scale measures visual–spatial ability (*Gv*), but some researchers argue that the performance scale measures a mixture of *Gv* and nonverbal reasoning (*Gf*).

## 3.2. Cognitive Assessment System

The CAS was developed according to the PASS (Planning, Attention, Simultaneous, Successive) theory of intelligence. The theory proposes that human cognitive functioning is based on four essential activities:

- Planning (which provides cognitive control, use of processes and knowledge, intentionality, and self-regulation to achieve a given goal)
- Attention (which provides focused, selective cognitive activity over time)
- Simultaneous processing (one form of operating on information)
- Successive processing (another form of operating on information)

The subtests are organized into four scales. Planning subtests require devising, selecting, and using efficient plans of action to solve the problems, regulate the effectiveness of the plans, and self-correct if necessary. Attention subtests require attending selectively to a particular stimulus and inhibiting attending to distracting stimuli. Simultaneous processing subtests require integrating stimuli into groups. Successive processing subtests require integrating stimuli in their serial order or appreciating the linearity of stimuli.

The CAS yields scores for the planning, attention, simultaneous, successive, and full scales, with a

normative mean of 100 and an SD of 15. All subtests are set at a normative mean of 10 and an SD of 3.

The battery was standardized in the United States on 2200 participants 5 to 17 years of age and was stratified by age, sex, race, geographic region, educational placement, and parents' education according to census reports and matches population characteristics in the variables used.

The internal consistency reliability estimates for the CAS were as follows: full scale = .96, planning and attention = .88, and simultaneous processing and successive processing = .93. The test–retest reliabilities for the CAS were as follows: full scale = .91, planning = .85, attention = .82, simultaneous processing = .81, and successive processing = .86. A total of 215 participants were tested twice, with a time interval of 21 days.

### 3.3. Kaufman Adolescent and Adult Intelligence Test

The Kaufman Adolescent and Adult Intelligence Test (KAIT) is based mainly on the Horn–Cattell theory of intelligence. Fluid intelligence (*Gf*) involves culture-fair novel tasks and taps problem-solving skills and the ability to learn. Crystallized intelligence (*Gc*) refers to acquired skill, knowledge, and judgments learned within a given society.

Thus, the battery provides fluid, crystallized, and composite IQs, with a mean of 100 and an SD of 15. It consists of a core battery of six subtests and an expanded battery. Each subtest yields age-based scores with a mean of 10 and an SD of 3. Crystallized intelligence measures the acquisition of facts and problem-solving ability using stimuli derived from schooling and cultural experiences. Fluid intelligence measures a person's adaptability when solving new problems. The KAIT fluid subtests emphasize reasoning rather than visual–spatial ability (contrary to the Wechsler performance scale).

The KAIT was standardized in the United States on 2000 participants 11 to 94 years of age and was stratified according to sex, race, geographic location, and socioeconomic status. The reliability coefficients were .95 for *Gc* and *Gf* and .97 for composite IQ. Test–retest reliabilities were .94 for *Gc* and composite IQ and .87 for *Gf*. A total of 153 participants were tested twice, with a time interval of 1 month.

Both exploratory and confirmatory factor analyses support the construct validity of the KAIT. Moreover, a correlation of .82 was observed between WISC-R full-scale IQ and the KAIT composite IQ, and a correlation

of .85 was observed between WAIS-R full-scale IQ and the KAIT composite IQ.

### 3.4. Woodcock–Johnson III

The WJ III consists of two distinct batteries: the WJ III Tests of Cognitive Abilities and the WJ III Tests of Achievement. Together, these batteries comprise a wide age range system for measuring general intellectual ability (*g*), specific cognitive abilities, oral language, and academic achievement. This subsection reviews only the WJ III Tests of Cognitive Abilities, where the standard battery and the extended battery can be found.

The standard battery includes verbal comprehension, visual–auditory learning, spatial relations, sound blending, concept formation, visual matching, numbers reversed, incomplete words, auditory working memory, and delayed visual–auditory learning. The extended battery includes general information, retrieval fluency, picture recognition, auditory attention, analysis–synthesis, decision speed, memory for words, rapid picture naming, planning, and pair cancellation.

The WJ III is based on current theory and research on the structure of human cognitive abilities. The theoretical basis is derived from the Cattell–Horn–Carroll theory of cognitive abilities (Fig. 1). The WJ III is a measurement model of this theory, and the design criteria place emphasis on providing the greatest practical breadth in Stratum II abilities: comprehension knowledge, long-term retrieval, visual–spatial thinking, auditory processing, fluid reasoning, processing speed, and short-term memory. The measured abilities are as follows:

- *Gf* (fluid intelligence) through concept formation (induction), analysis–synthesis (general sequential deductive reasoning), and planning (general sequential reasoning)
- *Gc* (crystallized intelligence) through verbal comprehension (lexical knowledge and language development) and general information (general verbal information)
- *Glr* (long-term retrieval) through visual–auditory learning (associative memory), delayed visual–auditory learning (associative memory), and retrieval fluency (ideational fluency)
- *Gv* (visual–spatial thinking) through spatial relations (visualization and spatial relations), picture recognition (visual memory), and planning (spatial scanning)
- *Ga* (auditory processing) through sound blending (phonetic coding synthesis), incomplete words

(phonetic coding analysis), and auditory attention (speech sound discrimination and resistance to auditory stimulus distortion)

- *Gs* (processing speed) through visual matching (perceptual speed), decision speed (semantic processing speed), rapid picture naming (naming facility), and pair cancellation (attention and concentration)
- *Gsm* (short-term memory) through numbers reversed (working memory), auditory working memory (working memory), and memory for words (memory span)

The data for WJ III norms were collected from a large, nationally representative sample of 8818 participants gathered in more than 100 geographically diverse U.S. communities. The preschool sample (2–5 years of age) was composed of 1143 participants. The kindergarten through 12th-grade sample was composed of 4783 participants. The college/university sample consisted of 1165 undergraduate and graduate students. The adult sample was composed of 1843 participants. The norming sample was selected to be representative of the U.S. population from 24 months to 90 years of age or over. The participants were randomly selected within a stratified sampling design that controlled for census region, community size, sex, race, type of school, type of college/university, education of adults, occupational status of adults, and occupation of adults in the labor force.

The WJ III general intellectual ability score is a general intelligence (*g*) score. It represents the first principal component obtained from principal component analyses. Using weights based on principal component analyses means that all subtest weights are optimal. In contrast, tests such as the Wechsler intelligence scales weight all subtests equally. Using principal component analyses of cognitive measures as a basis for prescribing different test weights gives the best statistical estimate of general intelligence.

Reliability coefficients ranged from .74 for planning to .97 for rapid picture naming. The construct validity of WJ III was tested by confirmatory factor analyses. The breadth of abilities measured by the WJ III is described by *g* plus several broad cognitive abilities plus several narrow cognitive abilities. Nearly all tests load exclusively on a single factor, suggesting that the cognitive tests have minimized the influence of construct-irrelevant variance.

The WJ III general intellectual ability scores had correlations ranging from .67 to .76, with composite

scores from the WPPSI-R, the WISC-R, or the Stanford–Binet IV. The WJ III matches its intended theory closely. It is a good translation of theory into practice.

#### 4. INTERPRETATION: SOME WORDS

The examiner must produce hypotheses about an individual's strengths and weaknesses based on theory, research knowledge, and clinical ability. The survival hypotheses must form practical recommendations.

Intelligence assessment devices must be selected prudently. These devices must meet sound psychometric standards, but the examiner should keep in mind that each examinee has unique characteristics. Alan Kaufman articulated five basic principles for intelligence assessment:

1. The items on an intelligence test measure what the individual has learned.
2. The subtests included on an intelligence test are nonexhaustive samples of behavior.
3. The items on an intelligence test assess mental functioning under fixed conditions.
4. The subtests on an intelligence test are useful when interpreted within an information processing model: how information enters the system, how it is interpreted and processed, how it is stored for later retrieval, and how it is expressed.
5. The hypotheses produced from an intelligence assessment should be supported from several sources: the person's behavior during test administration, the pattern of responses across several subtests, background information from parents, teachers, and/or peers, and so forth.

These basic principles are appealing, but the examiner must be cautious because there is always the temptation to neglect objective data in the face of other sources of much more subjective evidence. Therefore, the examiner must consider asking some colleagues when the distance between the objective result and the subjective appreciation increases.

#### 5. RECENT ADVANCES

Intelligence assessment is moving from an exclusively empirically rooted test construction to a more

theoretically based one. The WJ III and the KAIT are some examples. But there are other instances as well. The following focuses briefly on the recent applied research development known as the cognitive system design (CSD) approach.

The CSD approach emphasizes manipulating construct validity by the cognitive complexity of the items. These items can be designed to measure targeted aspects of processing by detailed specifications of the stimulus features that influence processing. The stages are as follows:

- Specify general goals of measurement.
- Identify design features in task domain (task-general features [e.g., model, format, conditions] and task-specific features).
- Develop a cognitive model (review theories, select or develop model for psychometric domain, revise model, and test model).
- Evaluate cognitive model for psychometric potential (evaluate cognitive model plausibility on current test, evaluate impact of complexity factors on psychometric properties, and anticipate properties of new test).
- Specify item distributions on cognitive complexity (distribution of item complexity parameters and distribution of item features).
- Generate items to fit specifications.
- Evaluate cognitive and psychometric properties for revised tests domain (estimate component latent trait model parameters, estimate plausibility of cognitive model, evaluate impact of complexity factors on psychometric properties, evaluate plausibility of the psychometric model, and calibrate final item parameters and ability distributions).
- Conduct a psychometric evaluation (measuring processing abilities and bank items by cognitive processing demands).
- Assemble test forms to represent specifications (fixed content test and adaptive test).
- Validate.

Cognitive item response theory (IRT) models are mathematical models of cognitive processes and IRT models of response patterns. These models contain parameters to represent the cognitive demands of items as well as the person's ability.

Susan Embretson has applied the CSD approach to develop a test of abstract reasoning called ART. ART contains matrix problems very similar to those of the Advanced Progressive Matrices (APM) test. A bank of 150 items was generated analytically from a list of

objects and attributes. Empirical results supported the CSD approach to test development. Working memory capacity was the primary cognitive aspect of item solving. Memory load predicted item difficulty. Items were banked by their cognitive demands by use of the memory load scores from the cognitive model. Moreover, ART scores were indistinguishable from APM scores. ART achieved the same internal consistency and factorial validity as did APM but from fewer items, suggesting that the CSD approach may eliminate features from items that are irrelevant to abstract reasoning.

Designing intelligence tests from cognitive theory is now feasible. Good results may be expected for a broad array of items if a high-quality cognitive theory about the item stimulus features that influence processing is developed. Future testing may change if the generating potential of cognitive theory is realized. The potential size of the generable items bank is quite large. Even item construction could be fully computerized for future tests. The test of the future will be a set of generating principles with previously known relationships to people's performance. This obviously requires a fully developed cognitive model.

### See Also the Following Articles

Clinical Assessment ■ Intelligence and Culture ■ Intelligence, Emotional ■ Intelligence in Humans ■ Psychological Assessment, Standards and Guidelines for ■ Psychometric Tests ■ Psychophysiological Assessment

### Further Reading

- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge, UK: Cambridge University Press.
- Colom, R. (2003). Multiple cognitive abilities assessment. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment*. Thousand Oaks, CA: Sage.
- Colom, R., & Andrés-Pueyo, A. (2000). The study of human intelligence: A review at the turn of the century. *Psychology in Spain*, 4(1), 167–182.
- Fernández-Ballesteros, R., & Colom, R. (2004). The psychology of human intelligence in Spain. In R. J. Sternberg (Ed.), *International handbook of the psychology of human intelligence* (pp. 79–103). Cambridge, UK: Cambridge University Press.
- Horn, J. (1989). Cognitive diversity: A framework for learning. In P. Ackerman, R. Sternberg, & R. Glaser (Eds.), *Learning and individual differences*. New York: Freeman.

- Kaufman, A. S., & Kaufman, N. L. (1993). *Manual for Kaufman Adolescent and Adult Intelligence Test (KAIT)*. Circle Pines, MN: American Guidance Service.
- McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock-Johnson III*. Chicago: Riverside.
- Naglieri, J. A., & Das, J. P. (1997). *Das-Naglieri Cognitive Assessment System*. Chicago: Riverside.
- Neisser, U., Boodoo, G., Bouchard, T., Boykin, A., Brody, N., Ceci, S., Halpern, D., Loehlin, J., Perloff, R., Sternberg, R., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77-101.
- Van der Linden, W., & Hambleton, R. (1997). *Handbook of modern item response theory*. New York: Springer.
- Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R)*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children-Third Edition (WISC-III)*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III)*. San Antonio, TX: Psychological Corporation.



# Intelligence, Emotional

Cary Cherniss

Rutgers University, Piscataway, New Jersey, USA

1. Why the Interest in Emotional Intelligence?
  2. Historical Roots of and Precursors to the Concept
  3. Various Models of Emotional Intelligence
  4. Measurement of Emotional Intelligence
  5. The Relationship between Emotional Intelligence and Personality
  6. The Contribution of Emotional Intelligence to Health, Work, and Educational Adjustment
  7. Can Emotional Intelligence Be Taught?
- Further Reading

Emotional intelligence refers to a set of abilities that involve the way in which people perceive, express, understand, and manage their own emotions as well as the emotions of others. Although there are several different definitions and models found in the literature, there is considerable overlap among them. At the core of every definition are the abilities to perceive emotion clearly in oneself and in others, to use emotions to facilitate thought and action, to understand how emotions affect one's own behavior and that of others, and to regulate one's own emotional reactions.

## 1. WHY THE INTEREST IN EMOTIONAL INTELLIGENCE?

For more than 100 years, psychologists have been trying to identify, measure, and modify those aspects of personality and behavior that most strongly affect a person's ability to adapt successfully to the demands of living. Many of these efforts have focused on cognitive ability, that is, the way in which one acquires, codes, stores, and retrieves information about the environment. However, no matter how one measures cognitive ability, the amount of variability in outcomes attributable to it always proves to be rather limited. Estimates vary depending on the nature of the task and context, but cognitive ability, as measured by intelligence quotient (IQ) tests or other tests of general mental ability, usually accounts for approximately 10% of the variability in performance outcomes—and sometimes much

## GLOSSARY

**cognitive abilities** Higher order mental faculties that involve the acquisition, analysis, and categorization of information as well as reasoning and problem solving.

**competency** A learned capability that contributes to superior performance.

**conative abilities** Having to do with motivation as opposed to emotion or cognition.

**360-degree assessment** An assessment in which a number of people who know an individual answer questions about that individual's attitudes, ability, or behavior; in the workplace, the raters are usually the individual's supervisors, peers, and subordinates.

**tipping point** A point in the linear relationship between two variables at which an additional small increase in the first variable results in a sharp increase in the second variable.

less. Up to now, the largest percentage attributable to cognitive ability has been approximately 25%. In addition, many tests of cognitive ability prove to be culturally biased; thus, the use of such tests perpetuates discrimination against many groups within society.

The significant but ultimately limited role of cognitive ability alone in helping people to adapt successfully has led to several attempts to identify other kinds of abilities that contribute to effective performance in school, interpersonal relationships, and the workplace. Emotional intelligence (EI) represents one such set of abilities.

## **2. HISTORICAL ROOTS OF AND PRECURSORS TO THE CONCEPT**

Even before psychologists began to measure intelligence, Darwin wrote about the adaptive importance of emotion in human and animal life. In 1872, he published a book in which he proposed that emotions serve two important survival and adaptive functions. First, they energize and motivate activity. Fear, to take an obvious example, helps one to be more vigilant and, when danger is spotted, to run faster from it. Second, emotions also serve a signal function, conveying important information from one animal in a group to others. If an elk, grazing in a herd, notices a mountain lion off in the distance, the elk's fear reactions alert the rest of the herd, and the entire herd quickly sets off in flight, helping the elk to survive and pass on their genes to the next generation.

As early as 1920, some of the pioneers in the study and measurement of intelligence recognized the need to move beyond purely cognitive-based abilities in their conceptions and to include other kinds of abilities. Thorndike, for instance, wrote about "social intelligence" in an article published in 1920. In 1943, Wechsler wrote that "nonintellective" factors or "affective and conative abilities" should be considered part of general intelligence, based on their capacity to facilitate intelligent behavior. During the early 1980s, Gardner developed his concept of "multiple intelligences," which included "intrapersonal" and "interpersonal" intelligences as well as the more traditional cognitive-based types. Also during the early 1980s, Bar-On began to study what he later referred to as "emotional and social intelligence." Around the same time, Sternberg began to write about "practical intelligence." In 1990, Salovey and Mayer published an article in which they

used the term "emotional intelligence" in the title, and they subsequently began developing their own model and measures of the construct.

There was another stream of activity in psychology that was a relevant part of the concept's history, even though it was not explicitly linked to intelligence. For decades, personality and social psychologists had been studying personal attributes that seemed to help people be more effective in a variety of contexts, and some of these attributes involved one's ability to perceive, express, and/or manage emotions in oneself and in others. For instance, there was research on empathy, defined as the ability to sense what others are feeling and to take their perspective, suggesting that both children and adults who scored higher in this quality tended to be more successful in both vocational and academic contexts. Similarly, there was research suggesting that the ability to express emotion in a clear and compelling way could have a positive impact on individual and group performance in a variety of contexts. Moreover, clinical work on anger and anger management suggested that awareness of one's emotional reactions and what triggers them can greatly influence one's ability to regulate emotion, which in turn affects well-being and success. Current research and applications relating to EI draw on all of these historical roots.

## **3. VARIOUS MODELS OF EMOTIONAL INTELLIGENCE**

Several different models of EI have emerged since 1990, and such diversity has contributed to a certain amount of controversy. Although there are important differences among the models, there also are many similarities. In fact, there probably is as much consensus about the EI construct as there is about the concept of general intelligence, a topic that has certainly seen its own share of conflict and controversy.

Salovey and Mayer called their model an "ability" model. Using a deductive approach, they identified four "branches" that are related in a hierarchical way: (a) the ability to perceive emotions accurately, (b) the ability to use emotions to facilitate thought, (c) the ability to understand emotions, and (d) the ability to manage emotions.

A second popular model is Goleman's "performance-based" model. Like Salovey and Mayer's model, the first component of Goleman's model involves perception of emotion. In this case, however, the focus is on awareness

of one's own emotions (Self-Awareness). A second component of the model involves awareness of emotions in others (Social Awareness). These first two components involve recognition of emotion. The other two components involve regulation of emotion: Self-Management and Relationship Management. Also like Salovey and Mayer, Goleman proposed a kind of hierarchical relationship among these components, with Self-Awareness constituting the foundation for Self-Management and Social Awareness, which in turn provides a foundation for Relationship Management. Goleman added another level to his model, consisting of approximately 20 "competencies" that are linked to the four basic dimensions of EI. For instance, under the Self-Management dimension, there are competencies such as flexibility and initiative. Goleman stressed that these competencies are not synonymous with EI, although EI provides the foundation for them. They are learned capabilities that contribute to superior performance.

Another model that has attracted considerable attention is Bar-On's model, which also is a competency-based model. Bar-On's model is composed of five dimensions: (a) Intrapersonal, which includes competencies such as emotional self-awareness and assertiveness; (b) Interpersonal, which includes empathy and interpersonal relationships, among others; (c) Stress Management, which incorporates stress tolerance and impulse control; (d) Adaptability, which consists of flexibility, reality testing, and problem solving; and (d) General Mood, covering optimism and happiness.

One reason why the models differ is that they were developed in different ways and for different purposes. Salovey and Mayer's model was developed deductively. Those researchers began with a general conception of EI and then defined their components based on that conception. Goleman, on the other hand, developed his model inductively. He identified the competencies that previous research suggested are most predictive of superior performance and then arranged them in clusters that seemed closely related to the basic components of EI. Bar-On's approach also was more inductive. Beginning with clinical work on life adjustment, he identified a set of skills that seemed to help people cope with demands and pressures. Then, using factor-analytic procedures, he ended up with his model of five major dimensions and 15 competencies.

Despite these differences in the models, a careful comparison suggests that there is considerable overlap. For instance, Salovey and Mayer's Perception of Emotion and Understanding Emotion dimensions seem to be similar to Goleman's Self-Awareness and Social Awareness

dimensions as well as to the emotional self-awareness and empathy subscales of Bar-On's model. Similarly, Salovey and Mayer's Emotional Facilitation of Thinking and Managing Emotions dimensions seem to be similar to Goleman's Self-Management and Relationship Management dimensions as well as to the Interpersonal, Stress Management, and Adaptability dimensions of Bar-On's model. Thus, although there are important differences among the three models, there also are many similarities.

#### 4. MEASUREMENT OF EMOTIONAL INTELLIGENCE

Measures for various aspects of EI have existed for some time. For instance, there are measures of empathy that can be used to assess one's ability to perceive clearly another person's mood or emotional response. Similarly, there are measures of emotional self-awareness and self-regulation. However, during the past decade or so, there also has been a concerted attempt to develop and refine instruments explicitly designed to measure all of the major facets of EI.

The first such instrument is Bar-On's Emotional Quotient Inventory (EQ-i). The EQ-i is a self-assessment instrument that asks an individual to answer a series of questions about his or her own emotional and social competence based on the way in which the respondent feels, thinks, and behaves most of the time. Bar-On originally developed the instrument during the 1980s and has done considerable research on it over the years. Plake and Impara's critical review in 2002 suggested that the EQ-i has good reliability and validity.

Another instrument is the Mayer, Salovey, and Caruso Emotional Intelligence Test (MSCEIT), which is an ability test in which an individual performs a number of tasks designed to test various aspects of EI. For instance, in one section of the test, the individual is shown a series of faces and must rate how much sadness, anger, jealousy, and the like are displayed in each one. This section is designed to assess the Perception of Emotion dimension. The MSCEIT is the latest in a series of tests that this research group has developed; thus, it has less psychometric research behind it at this point. But a study by Mayer, Salovey and Caruso provided some promising support for the instrument's reliability and validity.

A third test that sometimes is used to measure EI is the Emotional Competence Inventory (ECI), developed by Boyatzis and Goleman. The original version of this test

was a “360-degree assessment” in which a number of people who know an individual answer questions about that person’s social and emotional competence. Many of the items are based on earlier research that identified the competencies most important for success in the workplace. Like the MSCEIT, the ECI is relatively new, but the body of research reported by Sala in 2002 suggests that the instrument has good reliability and validity.

There have been several other attempts to develop measures of EI, but these have received less attention than have the EQ-i, MSCEIT, and ECI. Also, there has been less psychometric research reported on those other measures.

### **5. THE RELATIONSHIP BETWEEN EMOTIONAL INTELLIGENCE AND PERSONALITY**

One of the controversies surrounding the concept of EI concerns whether it is simply a new name for a set of personality traits that have been studied often in the past. Empirical research has addressed this issue by examining the strength of the relationship between common measures of EI and various personality traits. Most attention has focused on the relationship between EI and the Five-Factor Model of personality. A meta-analysis by Van Rooy and Viswesvaran, based on 69 studies involving more than 12,000 individuals, found that although there is some overlap, the relationships generally tend to be modest. The highest correlate of EI among the “Big Five” factors was Extraversion ( $r = .34$ ). This same study also found that EI demonstrated significant incremental validity over the Big Five factors but that the Big Five did not add to EI, suggesting that EI is a better predictor of performance than is personality.

### **6. THE CONTRIBUTION OF EMOTIONAL INTELLIGENCE TO HEALTH, WORK, AND EDUCATIONAL ADJUSTMENT**

Although there is some controversy about how much EI contributes to health, work, and educational outcomes, there is considerable research showing that it does have an impact. The research comes from two sources. First, there are recent studies that have used one of the explicit measures of EI that have appeared during the past decade or so. Van Rooy and Viswesvaran’s meta-analysis

found that, across all studies and measures, the correlation between EI and work performance was .24. For academic performance, the relationship was weaker; the overall correlation was .10. As is often the case with meta-analyses, the strength of the relationship varied considerably depending on the context of the study and the measures used. For instance, studies using an early measure of EI developed by Salovey and Mayer found a correlation of .32 between EI and performance, whereas researchers using the EQ-i published canonical correlations between EI and occupational performance as high as .74. Some researchers also have reported stronger associations between EI and academic performance. For example, one study found a correlation of .41 between EQ-i scores and grade point average in a group of American high school students.

Another source of research on the link between EI and important outcomes are the countless studies that have examined specific facets of EI during the past half-century or so. Such research has shown, for instance, that awareness of one’s own emotional reactions plays a vital role in stress regulation, anger management, and other processes that affect both physical and behavioral health. Emotional self-regulation plays an even greater role in health outcomes and also is a strong predictor of success in school and the workplace. Similarly, empathy (i.e., the ability to perceive emotions in others) has been shown to predict academic and vocational adjustment and to provide the foundation for positive social relationships that contribute to physical and behavioral health in various ways.

Most of the research on the link between EI and important outcomes probably provides an underestimation of EI’s significance because researchers typically look only at simple linear relationships. There is some research and theory suggesting that the relationship between EI and health or performance is more complex. For instance, Goleman suggested that EI is especially important in determining which individuals, among those occupying a particular role in an organization or occupation, will become superior performers or “stars.” Research on competence in work organizations summarized by McClelland in 1998 tends to support this view. Also, McClelland and his students showed that there tend to be “tipping points” in the relationship between emotional competencies and performance. This means that up to a certain level of competence, increases in competence do not result in significant increases in performance; however, once the level of competence passes a tipping point, there is a sharp increase in performance. Such complex relationships

could tend to suppress the strength of simple correlational and regression estimates.

## 7. CAN EMOTIONAL INTELLIGENCE BE TAUGHT?

At least since Freud's early work in psychoanalysis, psychotherapists of various kinds have been teaching their patients to become more aware of their emotional responses, to develop greater understanding of their emotions, to perceive the feelings of others more accurately, and to regulate their own emotions more effectively. Similarly, since the beginning of civilization, parents and teachers have been helping children to better understand and control their emotions and those of others. If these efforts had not been effective, civilization as we know it probably could not exist. Thus, it seems obvious that EI can be taught.

However, less obvious is the question of how EI can best be taught and how much change an intentional intervention can actually bring about in children or adults. The evidence suggests that EI can be taught but that the process is more time-consuming and difficult than many self-appointed gurus, teachers, and trainers often suggest. Perhaps the clearest indication that EI can be taught comes from efforts to help people better manage their emotions. Recent research on stress and anger management programs suggests that methods are now available to help many people who have problems with self-regulation of emotion. There also have been experimental procedures developed to help people become more aware of others' emotions and to identify those emotions more accurately. Moreover, a number of applied psychologists have developed educational curricula and teaching methods that seem to be effective in helping many children to develop social and emotional competence.

Unfortunately, all of these intervention strategies require a highly favorable context to be effective for many individuals. First, and most important, the individual usually must be highly motivated and willing to devote considerable time and effort to the change effort. Second, the teacher or trainer must have a high

degree of skill, particularly in the social and emotional domain, to be helpful. Third, the social environment needs to provide support and encouragement of various kinds. When all of these critical ingredients are present, well-designed and -implemented interventions can help children and adults to become more emotionally intelligent.

### See Also the Following Articles

Emotion ■ Intelligence and Culture ■ Intelligence Assessment ■ Intelligence in Humans ■ Personality Assessment ■ Traits ■ Work Adjustment

### Further Reading

- Bar-On, R., & Parker, J. D. A. (2000). *The handbook of emotional intelligence*. San Francisco: Jossey-Bass.
- Cherniss, C., & Goleman, D. (2001). *The emotionally intelligent workplace*. San Francisco: Jossey-Bass.
- Ciarrochi, J., Forgas, J. P., & Mayer, J. D. (2001). *Emotional intelligence in everyday life: A scientific inquiry*. Philadelphia: Taylor & Francis.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam.
- Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science and myth*. Cambridge, MA: MIT Press.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (in press). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*.
- McClelland, D. C. (1998). Identifying competencies with behavioral-event interviews. *Psychological Science*, 9, 331-339.
- Plake, B. S., & Impara, J. C. (Eds.). (2002). *The fourteenth Buros mental measurement yearbook*. Lincoln, NE: Buros Institute for Mental Measurement.
- Sala, F. (2002). *Emotional Competence Inventory technical manual*. Boston: Hay Group.
- Salovey, P., & Mayer, J. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9, 185-211.
- Van Rooy, D. L., & Viswesvaran, C. (in press). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*.
- Wagner, R. K. (1997). Intelligence, training, and employment. *American Psychologist*, 52, 1059-1069.



# Intelligence in Humans

Robert J. Sternberg

Yale University, New Haven, Connecticut, USA

1. Theories of Intelligence
  2. Development of Intelligence
  3. Measuring Intelligence
  4. The Heritability and Malleability of Intelligence
- Further Reading

## GLOSSARY

**crystallized intelligence** One of two major subfactors of general intelligence; it represents the accumulation of knowledge over the life span of the individual and may be measured by tests in areas such as vocabulary, general information, and achievement.

**deviation IQs** A means of determining intelligence test scores, based on deviations from an average score, calculated such that the normative equivalent for the median score is 100, approximately 68% of the scores are computed to fall between 85 and 115, and approximately 95% of the scores are computed to fall between 70 and 130.

**emotional intelligence** The ability to perceive and express emotion, assimilate emotion in thought, understand and reason with emotion, and regulate emotion in the self and others.

**fluid intelligence** One of two major subfactors of general intelligence; it represents the acquisition of new information or the grasping of new relationships and abstractions about known information.

**heritability coefficient** The degree to which heredity contributes to individual differences in intelligence, expressed in terms of a number on a scale from 0 to 1 such that a coefficient of 0 means that heredity has no influence on

variation among people, whereas a coefficient of 1 means that heredity is the only influence on such variation.

**intelligence** The ability to implement goal-directed adaptive behavior.

**mental age** A means of indicating a person's level of intelligence (generally in reference to a child), based on the individual's performance on tests of intelligence, by indicating the chronological age of persons who typically perform at the same level of intelligence as the test taker.

**mental retardation** Low level of intelligence, usually reflected by both poor performance on tests of intelligence and poor adaptive competence, that is, the degree to which a person functions effectively within a normal situational context.

**ratio IQ** A means of indicating performance on intelligence tests, based on a quotient of mental age divided by chronological age multiplied by 100.

**reaction range** The broad limits within which a particular attribute (e.g., intelligence) may be expressed in various possible ways given the inherited potential for expression of the attribute in the particular individual.

**theory of multiple intelligences** A theory suggesting that intelligence is composed of eight distinct constructs that function somewhat independently but may interact to produce intelligent behavior: bodily-kinesthetic intelligence, interpersonal intelligence, intrapersonal intelligence, linguistic intelligence, mathematical-logical intelligence, musical intelligence, naturalist intelligence, and spatial intelligence.

**triarchic theory of successful intelligence** A theory of intelligence asserting that intelligence is composed of three aspects that deal with the relationship of intelligence to the internal world, to experience, and to the external world.

Intelligence is the ability to implement goal-directed adaptive behavior. In this article, the concept of intelligence in humans is examined. The article opens with a review of theories of intelligence, then discusses development of intelligence, next measuring intelligence, and finally, the heritability and malleability of intelligence.

## 1. THEORIES OF INTELLIGENCE

### 1.1. Implicit Theories

How do psychologists know what intelligence is? One way in which they find out is to ask people. For example, they might ask experts. In a 1921 symposium on the definition of intelligence, an American psychologist, Lewis M. Terman, emphasized the ability to think abstractly. Another American psychologist, Edward L. Thorndike, emphasized learning and the ability to give good responses to questions. In a similar 1986 symposium, however, psychologists generally agreed on the importance of adaptation to the environment as the key to understanding both what intelligence is and what it does. They also emphasized learning skills and understanding one's own cognitive processes.

In a set of studies published in 1981, Sternberg and colleagues asked laypeople in the United States what they thought intelligence was. Three factors emerged from the responses: practical problem solving, verbal ability, and social competence. However, the conceptions that arise depend on who is asked. In a comparable set of studies carried out in Taiwan, five factors emerged: cognitive skills, getting along with other people, understanding oneself, knowing when to show one is smart, and knowing when not to show one is smart. In other studies conducted in Kenya, cognitive skills also were emphasized less than in the West, whereas obedience, respect, and understanding people were emphasized more.

Most theories of intelligence are explicit rather than implicit; that is, they are elicited not by asking people what they mean by intelligence but rather by having people perform tasks alleged to require intelligence and then analyzing the way in which people perform. There are several different kinds of explicit theories.

### 1.2. Psychometric Theories

Psychometric theories have generally sought to understand the structure of intelligence: What form does it take, and what are its parts (if any)? Such theories have

generally been based on and tested by the use of data obtained from tests of mental abilities. These tests include items such as vocabulary, completing series of numbers, seeing analogies between words, and visualizing what forms would look like if they were rotated in space.

The first of the major psychometric theories was that of the British psychologist Charles E. Spearman. In a 1904 article, Spearman argued that just two kinds of factors underlie all individual differences in test scores. Spearman called the first and more important kind of factor the "general factor" or *g*. It was said to pervade performance on all tasks requiring intelligence. The second kind of factor was specific to each test. But what exactly is *g*? In 1927, Spearman proposed that it might be something he labeled "mental energy."

An American psychologist, L. L. Thurstone, suggested instead that seven factors, or "primary mental abilities," pervade test performance: verbal comprehension (knowledge of vocabulary and in reading), verbal fluency (writing and producing words), number (solving simple arithmetical computation and reasoning problems), spatial visualization (mentally visualizing and manipulating objects), inductive reasoning (completing a number series or predicting the future based on past experience), memory (remembering people's names or faces), and perceptual speed (rapidly proofreading to discover typographical errors in a typed text).

Other psychologists, including Raymond B. Cattell and John B. Carroll, suggested that abilities are hierarchical. At the top of the hierarchy is *g* or general ability. But below *g* in the hierarchy are successive levels of gradually narrowing abilities, ending with Spearman's specific abilities.

Cattell suggested that general ability can be divided into two broad kinds of abilities: fluid and crystallized. Fluid abilities are the reasoning and problem-solving abilities measured by tests such as the analogies, classifications, and series completions. Crystallized abilities can be said to derive from fluid abilities and are viewed as their products, which would include vocabulary, general information, and knowledge about specific fields. John L. Horn suggested that crystallized ability more or less increases over the life span, whereas fluid ability increases during the earlier years and decreases during the later years.

Carroll proposed a "three-stratum" model of intelligence that is considered by some to be the most definitive psychometric model of intelligence because it is based on reanalysis of hundreds of data sets. According to this model, general ability is at the top of a hierarchy

of abilities. At the next lower stratum are various broad abilities such as learning and memory processes and the effortless production of many ideas. At the bottom of the hierarchy are many narrow specific abilities such as spelling ability and reasoning speed.

J. P. Guilford, an American psychologist, proposed a “structure of intellect” theory that, in its earlier versions, postulated 120 abilities. For example, in an influential 1967 work, Guilford argued that abilities can be divided into five kinds of operations, four kinds of contents, and six kinds of products. These various facets of intelligence combine multiplicatively for a total of  $5 \times 4 \times 6$  or 120 separate abilities. An example of such an ability would be cognition (operation) of semantic (content) relations (product), which would be involved in recognizing the relation between lawyer and client in the following analogy problem: lawyer: client: doctor:?. In 1984, Guilford increased the number of abilities proposed by his theory, raising the total to 150.

### 1.3. Cognitive Theories

In a 1957 address to the American Psychological Association, Lee J. Cronbach proposed that psychologists unite the “two disciplines of scientific psychology”—experimental and differential (i.e., the study of individual differences). His proposal led to cognitive theories of intelligence that are derived by experimental means.

Underlying most cognitive approaches to intelligence is the assumption that intelligence is composed of a set of mental representations (e.g., propositions, images) of information and a set of mental processes that can operate on the representations. A more intelligent person is assumed to mentally represent information better, and also to operate more quickly on these representations, compared with a less intelligent person.

In 1973, Earl B. Hunt, Nancy Frost, and Clifford E. Lunneborg showed that a critical ability underlying verbal intelligence is that of rapidly retrieving lexical information, such as letter names, from memory. A few years later, Sternberg identified key mental processes alleged to underlie many cognitive tasks, especially ones involving inductive reasoning. They included encoding stimuli, inferring relations between stimuli, and applying what one has learned.

Ian Deary and colleagues have sought to understand intelligence through the study of inspection time. They have found that more intelligent individuals can discriminate the lengths of the lines with less stimulus duration (inspection) times than can less intelligent individuals.

Two leaders in the field of cognitive psychology, Allen Newell and Herbert A. Simon, used computers to model intelligence. The underlying idea is that computers can, in some sense, show a kind of intelligence similar to that shown by humans. During the late 1950s and early 1960s, they worked with a computer expert, Clifford Shaw, to construct a computer model of human problem solving. Called the “General Problem Solver,” it could solve a wide range of fairly structured problems such as logical proofs and mathematical word problems. In 1972, Newell and Simon proposed a general theory of problem solving, much of which was implemented on the computer. It involved heuristics such as means–ends analysis, whereby at each step of problem solving, one seeks to reduce as much as possible the distance between the solution and where one is in the problem.

More recently, Marcel Just and Patricia Carpenter showed that complex intelligence test items, such as figural matrix problems involving reasoning with geometric shapes, could be solved by a computer program at a level of accuracy comparable to that of human test takers.

The models described heretofore are serial-processing models, whereby the computers take steps in sequence. David E. Rumelhart and Jay L. McClelland proposed what they called “parallel distributed processing” models of the mind. These models postulated that many types of information processing occur at once rather than just one at a time.

### 1.4. Cognitive–Contextual Theories

Cognitive–contextual theories deal with the way in which cognitive processes operate in various environmental contexts. Two of the major theories of this type were proposed by Howard Gardner and Sternberg.

In 1983, Gardner proposed a theory of what he called “multiple intelligences.” In the 1999 version of the theory, the multiple intelligences include (at a minimum) linguistic, logical–mathematical, spatial, musical, bodily–kinesthetic, naturalist, interpersonal, and intrapersonal intelligences. Gardner also speculated as to the existence of an existential intelligence.

An alternative theory that also takes into account both cognition and context is Sternberg’s “triarchic” theory of successful intelligence. According to Sternberg, intelligence is the ability to succeed according to one’s own definition of success within one’s sociocultural context. People succeed by recognizing and capitalizing on strengths and by recognizing and either compensating for or correcting weaknesses. Intelligence has three

aspects. These aspects relate intelligence to what goes on internally within a person, to what goes on in the external world, and to experience that mediates between the internal and external worlds.

The first aspect is the set of cognitive processes and representations that form the core of all thought. Sternberg distinguished among three kinds of processes: those involved in deciding what to do and later in deciding how well it was done, those involved in doing what one had decided to do, and those involved in learning how to do it in the first place. The second aspect is the application of these processes to the external world. According to Sternberg, mental processes serve three functions in the everyday world: adapting to existing environments, shaping existing environments into new ones, and selecting new environments when old ones prove to be unsatisfactory. According to the theory, more intelligent persons not only can execute many cognitive processes quickly or well but also know what their strengths and weaknesses are and capitalize on the strengths while remedying or compensating for the weaknesses. More intelligent persons, then, find a niche in which they can operate most efficiently.

The third aspect of Sternberg's triarchic theory is the integration of the internal and external worlds through experience. One measure of intelligence is the ability to cope with relatively novel situations. The abilities to cope with relative novelty and to automatize cognitive processing are seen as interrelated; the more a person is able to automatize the tasks of daily life, the more mental resources there are left to cope with novelty.

In 1990, Peter Salovey and John Mayer proposed the construct of emotional intelligence. This construct was then popularized by Daniel Goleman in a 1995 book. Emotional intelligence is the ability to perceive accurately, appraise, and express emotion; the ability to access and/or generate feelings when they facilitate thought; the ability to understand emotion and emotional knowledge; and the ability to regulate emotions to promote emotional and intellectual growth. Several tests are now available to measure emotional intelligence. They generally show modest to moderate correlations with conventional tests of intelligence.

### **1.5. Biological Theories**

Biological theories are based in the neuropsychological functioning that produces, at least in part, intelligent behavior. Several biological approaches have been used. One approach has been the investigation of types of intellectual performance as related to the regions of

the brain from which they originate. A researcher in this area, Jerre Levy, found that the left hemisphere is superior in analytical functioning, with the use of language being a prime example. The right hemisphere is superior in many forms of visual and spatial performance, and its functioning tends to be more synthetic and holistic, compared with the left hemisphere.

A second approach to research has involved the use of brain wave recordings to study the relation between these waves and either performance on ability tests or performance in various kinds of cognitive tasks. Researchers have found a relationship between certain aspects of electroencephalogram (EEG) and event-related potential (ERP) waves and scores on standard psychometric tests of intelligence.

A third approach involves the measurement of blood flow in the brain, which is a fairly direct indicator of functional activity in brain tissue. In such studies, the amount and location of blood flow in the brain is monitored while participants perform cognitive tasks. Horn found that older adults show decreased blood flow to the brain relative to younger adults. He also found that such decreases are greater in some areas of the brain than in others and that the decreases are particularly notable in those areas that are responsible for close concentration, spontaneous alertness, and the encoding of new information. Using positron emission tomography (PET), Richard Haier discovered that people who perform better on conventional tests of intelligence often show less activation in relevant portions of the brain than do those who do not perform as well. Presumably, this pattern of results reflects the fact that the better performers find the tasks to be easier and, thus, invoke less effort than do the poorer performers.

## **2. DEVELOPMENT OF INTELLIGENCE**

There have been diverse approaches to studying the development of intelligence. Psychometric theorists, for instance, have sought to understand how intelligence develops in terms of changes in the factors of intelligence over time and changes in the amounts of the various abilities that children have. For example, the concept of mental age was popular during the first half of the 20th century. A given mental age was believed to represent an average child's level of mental functioning for a given chronological age. Thus, an average 10-year-old would have a mental age of 10 years, but an above-average

8-year-old or a below-average 12-year-old might also have a mental age of 10 years. The concept of mental age has fallen into disfavor, however, and is now used only rarely. The concept does not work well over a chronological age of roughly 16 years, and its assumption of perfectly smooth continuous mental development is questionable.

### 2.1. Piaget's Theories on Intelligence

Jean Piaget, a Swiss psychologist, suggested that the child explores the world, observes regularities, and makes generalizations much as a scientist does. Two fundamental cognitive processes, assimilation and accommodation, are alleged to work in somewhat reciprocal fashion. Assimilation involves incorporating new information into an already existing cognitive structure. Accommodation involves revising or adapting a cognitive structure to incorporate new information. Cognitive development, according to Piaget, represents a dynamic equilibrium between these two processes of assimilation and accommodation.

Piaget also postulated that there are four major periods in intellectual development: sensorimotor, preoperational, concrete–operational, and formal–operational. The first, the sensorimotor period, extends from birth until roughly 2 years of age. During this period, the child learns how to modify reflexes to make them more adaptive, to coordinate actions, to retrieve hidden objects, and (eventually) to begin representing information mentally. The second period, the preoperational period, is from approximately 2 to 7 years of age. During this period, the child experiences the growth of language and mental imagery. The child also learns to focus on single perceptual dimensions such as colors and shapes. The third period, the concrete–operational period, is from approximately 7 to 12 years of age. During this period, the child develops an important set of skills that are referred to as conservation skills. The child will recognize that substances stay the same in amount, regardless of their form. Finally, children emerge into the fourth period, the formal–operational period, which begins at approximately 12 years of age and continues throughout life. The formal–operational child develops thinking skills in all logical combinations and learns to think with abstract concepts.

### 2.2. Post-Piagetian Theories

Later, so-called “post-Piagetian” theories of intellectual development, which have been heavily influenced by

Piaget (although they have substantially departed from his ideas), have taken several courses. One has been to expand on Piaget's work by suggesting a possible fifth (adult) period of development such as problem finding. A second course has been to suggest periods of development related to, but different from, those suggested by Piaget. A third course has been to accept that intellectual development occurs through the periods that Piaget proposed but to hold that the cognitive bases of development differ from those recognized by him. Some of these theories emphasize the importance of memory or knowledge rather than reasoning capacity.

### 2.3. Vygotsky's Theories on Intelligence

Lev S. Vygotsky, a Soviet psychologist, suggested that intellectual development may be largely influenced by a child's interactions with others. That is, a child sees others thinking and acting in certain ways and then internalizes and models what is seen. Vygotsky also proposed the notion of a zone of proximal development (ZPD), which is the range of ability between what a person can do on his or her own versus what the person can do with socially derived guidance. The zone is not directly observable but may be detected by providing a context in which the individual can perform both with and without guidance. The ZPD is sometimes termed the zone of potential development.

## 3. MEASURING INTELLIGENCE

### 3.1. Early Historical Background

In 1859, the publication of Charles Darwin's *The Origin of Species* had a profound effect on many lines of scientific work. Darwin suggested that the capabilities of humans are, in some sense, continuous with those of lower animals. Hence, they can be understood through scientific investigation. One person who was strongly influenced by Darwin's thinking was Sir Francis Galton, Darwin's cousin. For 7 years, between 1884 and 1890, Galton maintained a laboratory at the South Kensington Museum in London. For a small fee, visitors could have themselves measured on a variety of psychophysical tasks. These tasks included weight discrimination and sensitivity to musical pitch. Galton believed that these tests measured more than just psychophysical abilities. He believed that psychophysical abilities are the basis of intelligence and, hence,

that his tasks were measures of intelligence. Therefore, Galton's intelligence test required a person to perform simple tasks such as deciding which of two weights was heavier and showing how forcefully the person could squeeze his or her hand. The Galtonian tradition was taken to the United States by the psychologist James McKeen Cattell.

### 3.2. The IQ Test

A more influential tradition of mental testing was developed in France by Alfred Binet and his collaborator, Theodore Simon. In 1904, the minister of public instruction in Paris named a commission to study or create tests that would ensure that mentally retarded children received an adequate education. The minister was also concerned that certain children were being placed in classes for the retarded not because they were retarded but because they had behavioral problems and teachers did not want them in their classrooms. Binet and Simon proposed that tests of intelligence should measure skills such as judgment, comprehension, and reasoning—the same kinds of skills measured on most intelligence tests today. Binet's early test was taken to the United States by a Stanford University psychologist, Terman, whose version came to be called the Stanford–Binet test. This test has been revised frequently and continues in use.

The Stanford–Binet test and others like it have traditionally yielded, at the very least, an overall score referred to as an intelligence quotient (IQ). In its most recent form, this test yields an overall IQ and other scores as well, as does the Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Intelligence Scale for Children (WISC).

More recent tests of intelligence have expanded the range of abilities tested. For example, in 1997, J. P. Das and Jack Naglieri produced a test, the Cognitive Assessment System, based on a theory of intelligence first proposed by a Russian psychologist, Alexander Luria. The test measures planning abilities, attentional abilities, and simultaneous and successive processing abilities.

IQ was originally computed as the ratio of mental age to chronological (physical) age multiplied by 100. Thus, it is sometimes referred to as a ratio IQ. For example, if an 8-year-old child had a mental age of 10 years (i.e., performed on the test at the level of an average 10-year-old), the child would be assigned an IQ of  $(10/8) \times 100$  or 125. If the 8-year-old had a mental age of 6 years, the child's IQ would be  $(6/8) \times 100$  or 75. A score of 100, whereby the mental age equals the chronological age, would be average.

As discussed previously, the concept of mental age has fallen into disrepute, and few tests continue to involve the computation of mental ages. Many tests still yield an IQ, but it is most often computed on the basis of statistical distributions. The scores are assigned on the basis of what percentage of people in a given group would be expected to have a certain IQ. Scores computed in this way are called deviation IQs.

### 3.3. The Distribution of IQ Scores

Intelligence test scores follow an approximately normal distribution, meaning that most people score near the middle of the distribution of scores. Scores drop off fairly rapidly in frequency in either direction from the center of the distribution. For example, on the IQ scale, approximately two-thirds of all scores fall between IQs of 85 and 115 and approximately 95% of all scores fall between IQs of 70 and 130. Put another way, only 1 of 20 scores differs from the average IQ (100) by more than 30 points.

It has been common to associate certain levels of IQ with labels. For example, at the upper end, the label "gifted" is sometimes assigned to people with IQs over a certain point such as 130. At the lower end, mental retardation has been classified into various degrees depending on IQ so that, for example, IQs of 70 to 84 have been classified as borderline retarded, IQs of 55 to 69 as mildly retarded, IQs of 40 to 54 as moderately retarded, IQs of 25 to 39 as severely retarded, and IQs below 25 as profoundly retarded.

Many psychologists now believe that IQ represents only a part of intelligence and that intelligence is only one factor in both mental retardation and giftedness. Most current definitions of mental retardation stress adaptive skills as well as IQ, and they also emphasize attributes such as creativity, motivation, and achievement in conceptions of giftedness.

## 4. THE HERITABILITY AND MALLEABILITY OF INTELLIGENCE

Historically, intelligence has been viewed as a more or less fixed trait. This view conceives of intelligence as something with which people are born. Thus, the function of development is to allow this genetic endowment to express itself. A number of investigators have suggested that intelligence is highly heritable and that it is transmitted through the genes. Heritability is here

defined as the proportion of individual differences variance that is genetically transmitted. Other investigators believe that intelligence is only minimally heritable, if at all. Most authorities take an intermediate position. However, all agree that heritability operates within a reaction range, meaning that a given genotype for intelligence can result in a wide range of phenotypes, depending on how environmental factors interact with genetic ones.

Several methods are used to assess the heritability of intelligence. Perhaps the most well known is the study of identical twins reared apart. For a variety of reasons, identical twins are occasionally separated at or near birth. If the twins are raised apart, and if it is assumed that when the twins are separated they are randomly distributed across environments (often a dubious assumption), the twins would heritability have in common all of their genes but none of their environment except for chance environmental overlap. As a result, the correlation between their performances on tests of intelligence can provide an estimate of the proportion of variation in test scores due to heredity. Another method of computing the hereditary effect on intelligence involves comparing the relationship between intelligence test scores of identical twins and those of fraternal twins.

It appears that roughly half of the variation in intelligence test scores is caused by hereditary influences. However, Robert Plomin and others have shown that the heritability of intelligence increases with age, suggesting that genetic factors become more important and that environmental factors become less important to individual differences in intelligence with increasing age. During adulthood, heritability may reach as high as 70% or more. The estimates are computed, for the most part, on the basis of intelligence test scores, so that the estimates are only for that part of intelligence measured by the tests.

Whatever the heritability factor of IQ, an entirely separate issue is whether intelligence can be increased. Work by a New Zealand researcher, James Flynn, has shown that scores on intelligence tests rose rather steadily throughout the world during the mid- and late 20th century. The precise reason for the increase is unknown, although speculations include better education, better nutrition, and advances in technology.

Despite the general increase in scores, average IQs continue to vary both across countries and across socioeconomic groups. For example, many researchers have found a positive correlation between socioeconomic status and IQ, although they disagree over the reason for the relationship. Most psychologists agree that

differences in educational opportunities play an important role, and some investigators believe that there is a hereditary basis for the difference as well. But there is simply no broad consensus on the issue of why the differences exist. Again, the differences are based on IQ, not broadly defined intelligence.

No matter how heritable intelligence is, some aspects of it are still malleable. Heritability of a trait is a separate issue from malleability of the trait. For example, height is highly heritable but also highly modifiable by nutrition and other environmental factors. Thus, with intervention, even a highly heritable trait can be modified. There is a growing body of evidence that aspects of intelligence can be modified as well. In the view of many psychologists, intelligence is not merely a fixed trait. A program of training in intellectual skills can increase some aspects of a person's level of intelligence. No training program—no environmental condition of any sort—is likely to make a genius of a person with low measured intelligence. But some gains are possible, and a number of programs have been developed for increasing intellectual skills. A main trend for psychologists working in the intelligence field has been to combine testing and training functions so as to enable people to optimize their intelligence.

### See Also the Following Articles

Intelligence and Culture ■ Intelligence Assessment  
 ■ Intelligence, Emotional ■ Psychometric Tests

### Further Reading

- Binet, A., & Simon, T. (1916). *The development of intelligence in children*. Baltimore, MD: Williams & Wilkins (Originally published in 1905).
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Ceci, S. J. (1996). *On intelligence*. Cambridge, MA: Harvard University Press.
- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger/Greenwood.
- Neisser, U. (Ed.). (1998). *The rising curve*. Washington, DC: American Psychological Association.
- Piaget, J. (1972). *The psychology of intelligence*. Totowa, NJ: Littlefield-Adams.
- Spearman, C. E. (1927). *The abilities of man*. New York: Macmillan.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1997). *Successful intelligence*. New York: Plume.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. New York: Cambridge University Press.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.



# Intentional Behavior

Martin Fishbein

Annenberg School for Communication, University of Pennsylvania, Philadelphia, Pennsylvania, USA

1. Introduction
2. Defining Intentions
3. Intentions as Predictors of Behavior
4. Moderators of the Intention–Behavior Relation
5. An Integrated Theoretical Model
6. Determinants of Behavior
7. Determinants of Intention
8. Determinants of Attitudes, Norms, and Self-Efficacy
9. The Role of Distal Variables
10. Conclusion  
Further Reading

## GLOSSARY

**attitude** One's overall feeling of favorableness or unfavorableness with respect to a given object or action.

**behavior** A directly observable action performed with respect to some target, in some context, at some point in time.

**behavioral categories** Broad categories of behavior (e.g., prejudice, exercising, dieting) that are inferred from the observation of one or more behaviors.

**behavioral intention** A readiness to engage in a particular behavior.

**behavioral norm** One's perception of whether important others are (or are not) performing a given behavior.

**correspondence** The principle that measures of behavioral determinants and behavior involve exactly the same action, target, context, and time elements.

**goals** Desired conditions, occurrences, or states that may be outcomes of performing one or more behaviors (e.g., losing weight, staying healthy, getting an A on an exam).

**meta-analysis** A statistical procedure to quantify and summarize results from a number of different studies.

**self-efficacy** One's perception that he or she can perform a given behavior, even in the face of various obstacles.

**subjective norm** One's perception of what important others think he or she should (or should not) do.

Although it has been argued that many behaviors are performed automatically, without reflection on or awareness of intention, it would be hard to argue that most behaviors of interest to an applied psychologist are either automatic or habitual. In fact, most people can quickly and easily report the likelihood that they will or will not perform given behaviors, and when these measures of intention (or self-prediction) are assessed appropriately, they are highly related to actual behavioral performance.

## 1. DEFINING BEHAVIOR

Although behavior is often taken as a given, it is important to distinguish among behaviors, behavioral categories, and goals. All too often, investigators fail to distinguish between behaviors per se and occurrences that may be the outcome of those behaviors. For example, weight loss and success (or failure) on an exam have often been used as behavioral criteria. Unfortunately, neither of these is a behavior. Success on an exam is a possible outcome of specific actions

such as attending lectures, reading books, memorizing materials, and even copying answers from another person's test paper. Another problem is that investigators often treat inferences from behaviors as if they themselves were behaviors. For example, aggression, discrimination, dieting, and exercising have often been viewed as behavioral criteria. Unfortunately, none of these is a directly observable behavior. Instead, these are broad behavioral categories that are inferred from the observation of one or more behaviors. That is, although it is possible to observe specific actions that are assumed to be instances of a general class or category of behavior, the category per se cannot be directly observed. Thus, knowing that someone was "exercising," for example, might reveal very little about the specific "exercise" behaviors that the person did (or did not) perform.

In contrast to goals (or outcomes) and behavioral categories, behaviors are directly observable actions. Note, however, that the direct observation of any behavior always occurs in a given context at a given point in time. This implies that both context and time should be included in the definition of a behavior. Moreover, because most actions are directed at some object or target, a behavioral definition should involve four elements: the action (e.g., enlisting, buying, using), the target (e.g., the army, cigarettes, condoms), the context (e.g., after graduating high school, at the supermarket, for vaginal sex with one's spouse), and the time (e.g., yesterday, during the past month). Clearly, a change in any one of the elements changes the behavior under consideration. For example, enlisting in the army is a different behavior from enlisting in the navy (a change in target), using a condom for vaginal sex with one's spouse is a different behavior from using a condom for vaginal sex with a new partner (a change in context), and asking someone whether he or she purchased a car during the past 3 months is assessing a different behavior from asking whether someone purchased a car during the past 2 years.

Although all four elements should be considered in arriving at one's behavioral criterion, the level of specificity or generality used to define each element should be determined by the substantive questions being asked. Thus, a behavior may be assessed at a fairly general level (e.g., people could be asked whether they have ever [time] purchased [action] cigarettes [targets] [context unspecified]) or at a more specific level (e.g., people could be asked whether they had purchased [action] Marlboro Light 100s [target] from

their local supermarket [context] during the past 2 weeks [time]).

## 2. DEFINING INTENTIONS

Intentions can be conceptualized as a readiness to engage in a particular behavior. This readiness to act can be assessed by asking people to indicate the extent to which they agree (or disagree) with statements such as "I will engage in the behavior," "I intend to engage in the behavior," "I expect to engage in the behavior," "I am willing to engage in the behavior," and "I will try to engage in the behavior." A number of investigators, however, have proposed that some of these items assess distinct constructs. For example, it has been suggested that it is important to distinguish between intention and behavioral expectation. More specifically, it was hypothesized that behavioral expectations are better predictors of behavior than are behavioral intentions because the former are more likely to take into account possible impediments to performance of the behavior. Thus, items such as "I intend to . . .," "I will try to . . .," and "I plan to . . ." have been used to assess intentions and items such as "I expect to . . ." and "I will . . ." to assess behavioral expectations. Available evidence to date suggests that there is little to be gained by this proposed distinction. For example, a meta-analysis of a broad set of behaviors found no difference in the predictive validity of expectations and intentions. Thus, this article uses the term "behavioral intention" to refer to measures that assess one's readiness to engage in a specific behavior. In addition, it should be noted that people may also form intentions to engage in a behavioral category (e.g., to exercise) or to pursue some goal (e.g., to lose weight).

## 3. INTENTIONS AS PREDICTORS OF BEHAVIOR

Empirical research over the past three decades or so has led to the recognition that behaviors can be predicted with considerable accuracy by appropriately assessing intentions to engage in the behaviors under consideration. For a measure of intention to correlate closely with behavior, the intention measure must involve exactly the same elements as does the behavior itself.

### 3.1. The Principle of Correspondence

The requirement that measures of intention and behavior involve exactly the same action, target, context, and time elements is known as the principle of correspondence or compatibility. The principle of correspondence is sometimes misunderstood to mean that intentions and behaviors should always be defined and measured at a very specific level. In reality, the principle simply suggests that the two variables—intention and behavior—should be measured at equivalent levels of generality or specificity. If, for whatever reason, investigators are interested in a very specific behavior performed in a particular context and at a given point in time, the intention should be assessed at that same level of specificity. However, investigators might not be interested in a particular context or time, and they can generalize across these elements in their measures of behavior and in their measures of intention. To the extent that the assessments of intention and behavior comply with the principle of correspondence, they should correlate highly with each other.

Many studies have substantiated the predictive validity of behavioral intentions. When measured appropriately, behavioral intentions account for an appreciable proportion of variance in actual behavior. Meta-analyses covering diverse behavioral domains have reported mean intention–behavior correlations ranging from .45 to .62. Studies in specific behavioral domains, such as condom use and exercise, have produced similar results, with intention–behavior correlations ranging from .44 to .56. A meta-analysis of these and other meta-analyses reported an overall correlation of .53 between intention and behavior.

Thus, there is considerable evidence that the performance or nonperformance of a given behavior is determined primarily by the strength of a person's intention to perform (or to not perform) that behavior, where intention is defined as the subjective likelihood that one will perform (or try to perform) the behavior in question. In contrast to the strong relationship between behavioral intention and behavior, however, intentions to reach goals and intentions to engage in behavioral categories might not provide very good prediction of goal attainment or the performance of a specific behavior within the behavioral category. In the latter case, the problem is largely methodological and in many ways is nothing more than a case of failure to obtain correspondence. That is, people are asked their intentions to engage in a class of behaviors (e.g., to diet

or exercise), but the behavioral criterion is a measure of whether they have or have not performed one or more specific behaviors that are assumed to be a good indicant of the category. With respect to goal attainment, it should be clear that achieving a goal often requires more than a behavioral performance on the part of an individual. Achieving a goal often depends on the decisions or actions of other people or on the occurrence of certain events. To put this somewhat differently, goal (or outcome) attainment is often not completely under an individual's control, and the less control one has over attaining some goal, the lower the intention–goal attainment relation. In the same sense, one could argue that the less a given behavior is under one's control, the lower the intention–behavior relationship. Fortunately, as pointed out earlier, most human behaviors (and certainly most that are of interest to an applied psychologist) are under volitional control. Nevertheless, people do not always act in accordance with their intentions.

### 4. MODERATORS OF THE INTENTION–BEHAVIOR RELATION

Perhaps the two major factors that prevent one from acting on his or her intentions are (a) a lack of necessary skills and abilities and (b) the presence of environmental constraints (or the absence of environmental facilitators). Other factors may also account for observed low intention–behavior relations. As discussed previously, low correlations may be due to a lack of correspondence between the measures of intention and behavior. In addition, intentions may change over time; the longer the time interval between the assessment of intention and the assessment of behavior, the greater the likelihood that the intention will change. This implies that the longer the interval between the assessment of intention and the observation of behavior, the lower the intention–behavior correlation.

### 5. AN INTEGRATED THEORETICAL MODEL

Although there are many theories of behavioral prediction and behavior change in the literature (e.g., the theory of planned behavior, the theory of subjective culture and interpersonal relations, the transtheoretical

model of behavior change, the information/motivation/behavioral skills model, the health belief model, social cognitive theory, the theory of reasoned action), a careful consideration of these theories suggests that there are only a limited number of variables that need to be considered in predicting and understanding any given behavior.

### 5.1. Attitudes, Perceived Norms, and Self-Efficacy

In general, one can identify three factors that may directly influence an individual's intentions and behaviors: (a) the individual's attitude toward performing the behavior (i.e., the person's overall positive or negative feelings of favorableness or unfavorableness with respect to performing the behavior) that is based on the person's beliefs that performing the behavior will lead to various positive or negative consequences (or outcomes), (b) perceived normative pressure that includes the perception that the individual's important others think that one should (or should not) perform the behavior in question as well as the perception that these important others are (or are not) themselves performing the behavior, and (c) self-

efficacy or personal agency that involves the individual's perception that he or she can perform the behavior in question under a variety of difficult or challenging circumstances. These and other considerations suggest the following integrative model of behavioral prediction (Fig. 1).

### 6. DETERMINANTS OF BEHAVIOR

In Fig. 1, it can be seen that any given behavior is most likely to occur if one has a strong intention to perform the behavior, if one has the necessary skills and abilities required to perform the behavior, and if there are no environmental constraints preventing behavioral performance. Indeed, if one has made a strong commitment (or has formed a strong intention) to perform a given behavior, if one has the necessary skills and abilities to perform the behavior, and if there are no environmental constraints to prevent the performance of that behavior, the probability is very high that the behavior will be performed.

In some populations or cultures, the behavior might not be performed because people have not yet formed intentions to perform the behavior. In other populations,

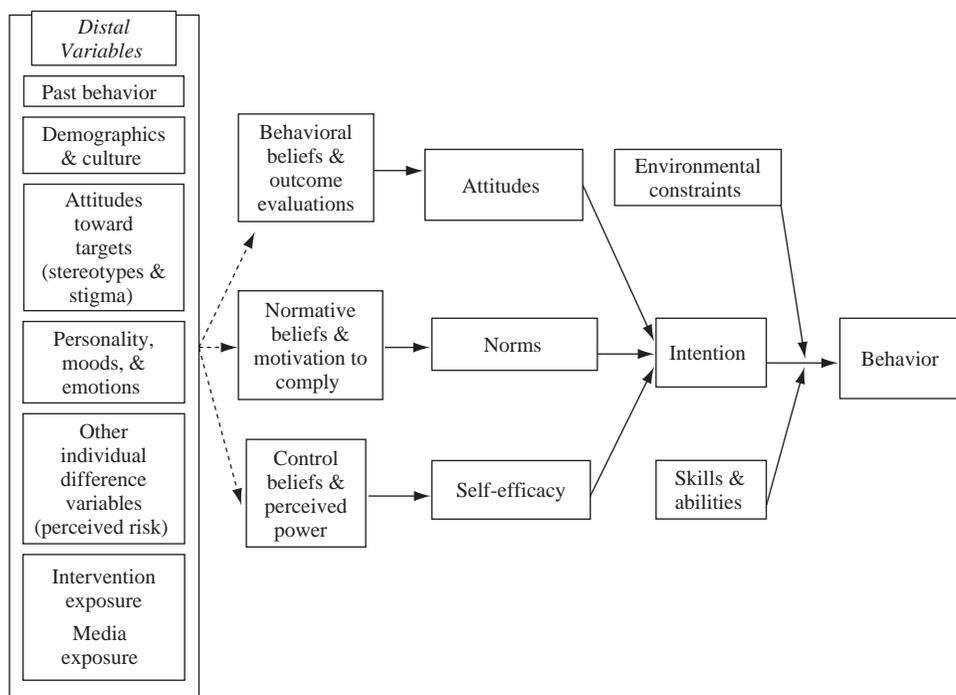


FIGURE 1 An integrative model of behavioral prediction.

the problem might be a lack of skills and/or the presence of environmental constraints. In still other populations, more than one of these factors may be relevant. Clearly, if people have formed intentions but are not acting on them, behavioral performance will depend on enhancing their skills and abilities or on removing (or learning to overcome) environmental constraints.

## 7. DETERMINANTS OF INTENTION

On the other hand, if strong intentions to perform the behavior in question have not been formed, the model suggests that there are three primary determinants of intention: the attitude toward performing the behavior, perceived norms concerning performance of the behavior, and one's self-efficacy with respect to performing the behavior. It is important to recognize that the relative importance of these three psychosocial variables as determinants of intention will also depend on both the behavior and the population being considered. For example, one behavior may be determined primarily by attitudinal considerations, whereas another behavior may be influenced primarily by feelings of self-efficacy. Similarly, a behavior that is attitudinally driven in one population or culture may be normatively driven in another population or culture. Thus, to understand intentional behavior, it is important to first determine the degree to which that intention is under attitudinal, normative, or self-efficacy control in the population in question.

## 8. DETERMINANTS OF ATTITUDES, NORMS, AND SELF-EFFICACY

The model in Fig. 1 also recognizes that attitudes, perceived norms, and self-efficacy are themselves functions of underlying beliefs—about the outcomes of performing the behavior in question, about the normative proscriptions and/or behaviors of specific referents, and about specific barriers to behavioral performance. Thus, the more one believes that performing the behavior in question will lead to “good” outcomes and prevent “bad” outcomes, the more favorable one's attitude toward performing the behavior should be. Similarly, the more one believes that specific others are performing the behavior and/or the more these others think that one should (or should not) perform the behavior in question, the more social pressure one will feel with respect to performing (or not

performing) the behavior. Finally, the more one perceives that he or she can (i.e., has the necessary skills and abilities to) perform the behavior even in the face of specific barriers or obstacles, the stronger one's self-efficacy will be with respect to performing the behavior.

It is at this level of underlying beliefs that the substantive uniqueness of each behavior comes into play. Clearly, a change in any one element in the behavioral definition will usually lead to very different beliefs about the consequences of performing that behavior, about the expectations of relevant others, and about the barriers that may impede behavioral performance and, thus, will usually lead to very different attitudes, subjective norms, and perceptions of self-efficacy and intentions. For example, the barriers to, and the outcomes (or consequences) of, always using a condom for vaginal sex with an occasional partner may be very different from those associated with always using a condom for vaginal sex with one's spouse. Yet it is these specific beliefs that must be considered if one wishes to fully understand intentions and behavior. Although an investigator can sit in her or his office and develop measures of attitudes, perceived norms, and self-efficacy, the investigator cannot know what a given population (or a given person) believes about performing a given behavior. Thus, one must go to members of that population to identify salient outcome, normative, and efficacy beliefs. To put this somewhat differently, one must understand the behavior from the perspective of the population that one is considering.

## 9. THE ROLE OF DISTAL VARIABLES

Finally, Fig. 1 also shows the role played by more traditional demographic, economic, personality, attitudinal, and other individual difference variables (e.g., perceived risk, sensation seeking). Although it is clear that these types of variables may often be important determinants of a given behavior, the model suggests that these types of variables play primarily an indirect role in influencing behavior. That is, these “background” factors may or may not influence the behavioral, normative, or self-efficacy beliefs underlying attitudes, norms, or self-efficacy. For example, although men and women may hold different beliefs about performing some behaviors, they may hold very similar beliefs with respect to performing other behaviors. Similarly, rich and poor, old and young, those from developing countries and those from developed countries, those who plan to go to college and those

who do not, those with favorable attitudes toward law enforcement and those with unfavorable attitudes toward law enforcement, those who are happy and those who are sad or angry, those who have used drugs and those who have not—all may hold different attitudinal, normative, or self-efficacy beliefs with respect to one behavior but may hold similar beliefs with respect to another behavior. In other words, there is no necessary relation between these “distal” or “background” variables and any given behavior. Nevertheless, distal variables, such as cultural and personality differences, moods and emotions, and differences in a wide range of values, should be reflected in the underlying belief structure.

## 10. CONCLUSION

In general, most human behavior is intentional; that is, people typically perform behaviors that they intend (or plan) to perform, and they do not perform behaviors they do not intend to perform. In addition, they usually intend to perform (or not perform) a behavior for one or more of three reasons: they think that performing the behavior is a “good” thing to do, they feel strong social pressure to perform the behavior, or they believe that they have the necessary skills and abilities to perform the behavior. These attitudes, perceived norms (or social pressure), and feelings of efficacy are themselves determined by underlying beliefs—about the expected outcomes of performing the behavior, about the expectations and behaviors of others, and about their ability to overcome specific barriers that could impede behavioral performance. The full understanding of the performance or nonperformance of a given behavior ultimately rests on understanding which of these underlying beliefs is (are) responsible for the performance or nonperformance of that behavior.

## See Also the Following Articles

Self-Control

## Further Reading

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior & Human Decision Processes*, 50, 179–211.
- Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 888–918.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Becker, M. H. (1974). The health belief model and personal health behavior. *Health Education Monographs*, 2, 324–508.
- Fishbein, M. (2000). The role of theory in HIV prevention. *AIDS Care*, 12, 273–278.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison–Wesley.
- Fishbein, M., Triandis, H. C., Kanfer, F. H., Becker, M., Middlestadt, S. E., & Eichler, A. (2001). Factors influencing behavior and behavior change. In A. Baum, & T. A. Revenson (Eds.), *Handbook of health psychology*. Mahwah, NJ: Lawrence Erlbaum.
- Fisher, J. D., & Fisher, W. A. (1992). Changing AIDS-risk behavior. *Psychological Bulletin*, 111, 455–474.
- Petratis, J., Flay, B. R., & Miller, T. Q. (1995). Reviewing theories of adolescent substance use: Organizing pieces in the puzzle. *Psychological Bulletin*, 117, 67–86.
- Prochaska, J. O., & DiClemente, C. C. (1992). Stages of change in the modification of problem behaviors. *Progress in Behavior Modification*, 28, 184–218.
- Sheeran, P. (2002). Intention–behavior relations: A conceptual and empirical review. In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 1–36). Chichester, UK: Wiley.
- Triandis, H. C. (1972). *The analysis of subjective culture*. New York: John Wiley.



# Intergroup Relations and Culture

Karmela Liebkind

University of Helsinki, Helsinki, Finland

1. In the Crossroads of Various Research Traditions
2. Culture: The Missing Link in Intergroup Relations
3. Intercultural Contact as Interpersonal Adjustment
4. The Social Psychology of Intergroup Bias
5. Culture as a Moderator of Intergroup Bias
6. The Role of History and Social Context in Intergroup Bias
7. The Reduction of Intergroup Bias

Further Reading

## GLOSSARY

**contact hypothesis** A theoretical framework that assumes that intergroup relations can, under certain conditions, be improved by increasing intergroup contact.

**cultural distance** Cultural differences between two groups; the more different the languages, family structures, religion, standard of living, and values of the two groups, the greater the cultural distance between them.

**discrimination** Unequal treatment of members of a social group on unacceptable grounds (e.g., ethnic/cultural origin, language, religion, skin color).

**immigrants** Persons of foreign nationality moving to a country for the purpose of staying there; in time, they may form new ethnic groups in the society of settlement.

**minority groups** Social groups in a numerical minority in the society; numerical inferiority is often combined with low status, but there are also high-status minorities.

**prejudice** Negative stereotypical beliefs about a social group combined with negative feelings toward that group.

**social identity** Those aspects of an individual's self-image that derive from the social categories to which the individual perceives himself or herself as belonging; these

aspects contribute positively or negatively to the individual's self-esteem.

**stereotyping** A generalizing view of a social group that ignores individual differences among the group members.

Intergroup relations have traditionally been studied by social psychologists, whereas intercultural relations have traditionally been studied by cross-cultural psychologists. All humans can fall prey to stereotyping, discrimination, and racism, but culture determines acceptable levels of expressed bias. Collective representations about the past influence current intergroup relations. Minority members have differing attitudes about retaining their culture of origin and becoming part of the mainstream society. These attitudes interact with those of the majority population and with official minority policies. The most harmonious intergroup relations are predicted when both majority and minority members share integration, assimilation, or individualist orientations. Explicit outgroup attitudes and stereotyping operate in a conscious mode, but there are also attitudes and stereotypes that function in an unconscious fashion. The translation of prejudiced beliefs (stereotypes) and evaluations (attitudes) into action (discrimination) depends on both personal and social factors. Thus, both individual and intergroup approaches can make important contributions to the reduction of intergroup bias. Interventions used for reducing bias usually focus on the need to increase the quantity and enhance the quality of intergroup contact. Some interventions can be effective

under particular contact conditions, whereas others may fail, particularly with respect to generalization from the individual out-group member to the out-group as a whole and to finding an intervention that works for both majority and minority groups.

## **1. IN THE CROSSROADS OF VARIOUS RESEARCH TRADITIONS**

Issues on cultural pluralism or multiculturalism, ethnicity, and migration are hotly debated in many countries. These issues pose many political, legal, social, and economic questions. In addition, they raise many social psychological questions regarding acculturation, integration, stereotyping, “new” or “subtle” racism, ethnic identity, and intergroup relations. It is obvious that a clear understanding of intergroup relations can facilitate the processes of immigration, acculturation, and integration.

Intergroup relations have been studied mostly by social psychologists, whereas intercultural relations have been studied mostly by cross-cultural psychologists, but traditionally these two research traditions have largely tended to ignore each other. In addition, generations of researchers in the field of intergroup relations have alternated individual versus contextual levels of analysis, sometimes pushing them as far as they can go before collapsing under the accumulated complexity of evidence (pro and con) and forgetting the advantages of previous approaches. Individual difference approaches to stereotyping, prejudice, and racism all apparently emerge from U.S. researchers. Coupled with an explicit constitutional ideology of equality, the U.S. cultural focus on individualism places responsibility for bias on individuals and privileges individual autonomy over ethnic group identity. In Europe, the histories of intergroup encounters are centuries older. Coupled with the importance of linguistic, cultural, religious, and geographical boundaries, a lesser focus on individualism makes Europe the logical birthplace of more contextual approaches to intergroup relations.

Intergroup behavior can be defined as actions by members of one group toward members of another group. The most studied social psychological phenomenon related to intergroup relations is intergroup bias. Forms of intergroup bias range from prejudice and stereotyping (via discrimination, injustice, perpetuation of inequality, and oppression) to ethnic cleansing and genocide. Various forms of prejudice (i.e., the holding of derogatory attitudes

about members of a social group) are more common than outright manifestations of intergroup conflict. In practice, therefore, the vast majority of social psychological studies have investigated weaker forms of bias, as expressed by persons with relatively mild prejudice.

Within both social and cross-cultural psychology, it has been customary to look at social behavior from a perspective that reflects the predominantly individualist emphasis of many contemporary North American researchers. Research on intergroup relations, however, more often applies a perspective that still focuses on individual behavior but that takes group membership as the determinant of that behavior. The individualist approach sees group behavior as an aggregated result of the social cognitions and resultant behaviors of its members. The more contextual approach reverses this presumed causal chain of events in that the individual's identity and consequent behaviors are seen as defined by his or her group memberships. The principal theorists proposing this perspective have been Europeans, but the influence of their ideas is now widely diffused. Most researchers would probably agree that social psychology should be able to offer explanations of both individual and small group behavior. European researchers tend to include explanations at the level of intergroup relations and of society as a whole. It is through an emphasis on these more macroscopic levels of influence that recent European work has become distinctive.

However, social psychology has also been criticized for being traditionally more concerned with general processes and structures than with particular contents related to historical, cultural, and contextual circumstances. Consequently, the growing interest within social psychology in content, culture, and context is very welcome. Perceptions of history and common ancestry, as well as assumed cultural characteristics, play a central role in interethnic/intercultural relations. Many questions and problems in studying cultural features are discussed extensively in cross-cultural and cultural psychology. It is the contextual, rather than the individualist, approach to intergroup and intercultural relations that dominates the bulk of this article.

## **2. CULTURE: THE MISSING LINK IN INTERGROUP RELATIONS**

### **2.1. What Is Culture?**

Various aspects of the environment are more or less important for members of different cultures. Culture

has been defined in numerous ways. It can be seen as consisting of a set of human-made objective and subjective elements that in the past have increased the probability of survival and resulted in satisfaction for participants in an ecological niche and, thus, have become shared among those who could communicate with each other because they had a common language and they lived in the same time and place. Cultural differences can also be seen as being tied to historical experiences, some of which are based on socioeconomic stratification of a relatively recent nature and others of which are rooted in a distant path of which nobody is conscious. This historically older culture is sometimes called “deep culture” because it is embedded in language, ethnicity, religion, and/or nationality as such.

For the purpose of this article, culture is conceptualized as shared patterns of belief and feeling toward issues such as child-rearing practices, family systems, and ethical values or attitudes. These are elements of what Triandis has called “subjective culture” that is created in interactions among people. In this perspective, there are as many cultures as there are languages and ways of life within each language (e.g., physicians have a culture that is somewhat different from that of lawyers). Thus, culture can be viewed as a relatively stable and characteristic system of shared meanings that define or determine perceptions and behavior. This approach is typically taken when different ethnic groups are compared, especially within cross-cultural psychology. Changes are not denied, but a more long-term perspective is taken. More gradual changes are examined in research on acculturation that focuses on psychological and behavioral changes that an individual or group experiences as a result of sustained contact with members of another cultural group(s).

## 2.2. Culture in the Research on Intergroup Bias

The groups involved in intercultural encounters are mostly large-scale ethnic or cultural groups. Yet much of even the more contextual branch of social psychological research on intergroup relations concentrates mainly on status differences between groups. Although an exclusive social position perspective offers important insights into intergroup relations, there are limits to using such a perspective. History and culture are the two main ingredients of ethnicity. Ethnicity, in turn, is most often thought of as culture that is transmitted across generations.

All humans can fall prey to stereotyping, discrimination, and racism. At a more global level, however, little is known about individual differences in these phenomena. In some Asian cultures, which have a history of being more ethnically homogenous within nations and more collective in general, social psychologists rarely study individual variation in stereotyping and prejudice. Likewise, Latin American and African social psychologists still have much to say on this topic. However, the importance of cultural and local norms in predicting discriminatory behavior should not be underestimated. Culture channels stereotyping and prejudice by defining who constitutes “us” and who constitutes “them.” Culture also determines acceptable levels of expressed bias, from subtle to overt. In addition, cultures differ in norms for describing perceived differences between social categories as either inherent and traditional or unacceptable and controllable.

It must be acknowledged that the vast majority of research on intergroup issues, such as stereotyping, prejudice, and discrimination, is carried out in societies that are low on cultural measures of hierarchy, power distance, and conservatism. Such systems tolerate considerable public discussion about inequalities and their legitimacy. Given the egalitarianism of such societies—or at least their ideal of egalitarianism—one must be cautious in generalizing globally some of the psychological processes that have been identified by this research.

## 3. INTERCULTURAL CONTACT AS INTERPERSONAL ADJUSTMENT

### 3.1. Intercultural Relations and Sociocultural Adaptation

When groups of individuals having different cultures come into continuous firsthand contact with each other, the consequence is likely to be subsequent changes in the original cultural pattern of either group or both groups. This is how the process of so-called acculturation is traditionally defined. This process refers to groups in contact, thereby explicitly placing acculturation within the realm of intergroup relations. However, within cross-cultural psychology, theoretical frameworks in relation to the acculturation process have not been borrowed from social psychology; instead, they have been borrowed from different areas of mainstream psychology, notably the stress and coping literature, on the one hand, and research on social learning and skills acquisition, on the other.

In both perspectives, acculturation is nearly exclusively seen as an individual-level process of adjustment and adaptation to a new culture. As a consequence, the acculturation literature has dealt more with interpersonal relations than with intergroup relations.

Ward and colleagues have argued that adaptive outcomes of acculturation are meaningfully divided into psychological (emotional/affective) and sociocultural (behavioral) domains. The former has traditionally been the focus of study for the stress and coping framework, and the latter has traditionally been the focus of study for the social learning framework. Individuals who are exposed to acculturative demands are motivated to maintain psychological well-being as well as to acquire culturally appropriate knowledge and skills. Empirically, these dimensions are moderately and positively related. This interrelation is understandable given that successful social adaptation to a new culture arises in part from psychological well-being and leads to a feeling of greater self-efficacy. This, in turn, enhances psychological well-being and vice versa.

It is obvious that good social relations with members of the other cultural group are a vital part of sociocultural adaptation. The more different the languages, family structures, religion, standard of living, and values of the two groups, the greater the cultural distance between the groups. The greater the cultural distance, the more the norms are likely to be different and the more difficult it is for the two groups to develop good intergroup relations. Cultural distance results in perceived dissimilarity, which can lead to conflict if there are different goals and interests. Individuals confronting new environments with greater cultural dissimilarities in values will experience more difficulties with adaptation. A robust relationship has been consistently observed between cultural distance and sociocultural adjustment problems—the greater the cultural similarity, the less difficulties with adaptation. This outcome also holds for perceived degrees of cultural differences.

Scholars approaching the problem of cross-cultural adaptation from a skills perspective emphasize the role of culture-specific knowledge for sociocultural adaptation. Depending on which particular cultural groups are involved, people will need specific knowledge of the rules of interaction with their hosts. Knowledge about normative patterns of behavior in the new culture will then have to be put into practice. For example, in training programs, Arab nonverbal skills can be taught to normally “stand-offish” English persons, thereby enhancing the impressions they give to Arabs. Various types of cross-cultural training can be very effective,

although some persons will already have culturally fitting skills through personalities or previous cross-cultural experiences. Powerful effects of cross-cultural training have been demonstrated for several factors such as better interpersonal skills with members of the other cultural group and better understanding of the social system and values of the other cultural group.

One important skill worth identifying separately is language competence. Greater fluency in the language of the other cultural group greatly increases the likelihood of successful intercultural communication. Of course, because the vast majority of cultural sojourners or immigrants will have no access to any cross-cultural training programs, they will have to adapt as best as they can. It is here that broad personality dispositions and the intercultural attitudes of the individuals, on the one hand, and the specific social, cultural, and historical context of the encounter, on the other, will influence the ultimate outcome of the intercultural contact.

## **4. THE SOCIAL PSYCHOLOGY OF INTERGROUP BIAS**

### **4.1. Individual- and Group-Level Explanations**

A popular view of the origins of prejudice is that that it is primarily a personality problem. The most influential example of this type of theory was that proposed by Adorno and colleagues in 1950. Working from a Freudian perspective, they believed that a child's natural aggression toward his or her parents (an inevitable consequence of being subjected to constraints) was displaced onto alternative targets due to the feared consequences of displaying it directly. The likely choice of targets would be those seen as weaker than or inferior to the child such as members of deviant groups or ethnic minorities. The end result was the so-called authoritarian personality. The association between authoritarianism and various forms of prejudice has been confirmed in several intergroup contexts. Recent work points to a positive relationship between prejudice and right-wing authoritarianism (i.e., submitting to established authorities and adhering to social conventions).

A similar individual-level theory of prejudice is the so-called frustration-aggression theory, which implies that aggression is often not directed at the true source of the frustration people feel (e.g., the capitalist system that causes an economic recession); instead, it is often diverted onto vulnerable and easily accessible targets

such as members of minority groups. Attempts to confirm this so-called scapegoat theory of prejudice have, however, met with mixed success. It seems that absolute levels of hardship and frustration are less potent instigators of aggression than is a sense of relative deprivation, that is, a perceived discrepancy between what one has and what one feels entitled to have.

However, it is important to note the limitations of this kind of individual difference perspective. It underestimates the importance of current social situations shaping people's attitudes and tends to neglect sociocultural determinants of prejudice. For example, Pettigrew found during the 1950s that although White South Africans showed very high levels of anti-Black prejudice, they did not appear to have particularly high levels of authoritarianism as compared with populations in other countries. In addition, the personality approach cannot explain the widespread uniformity of prejudice in certain societies or subgroups within societies. In prewar Nazi Germany, as well as in many other places since then, consistently racist attitudes and behavior were shown by hundreds of thousands of people who must have differed on most other psychological characteristics. Yet another problem concerns the historical specificity of prejudice. For example, increases in racist attacks on various ethnic minority populations in Western Europe during recent decades have sometimes taken place over the space of just a few years—much too short a time for whole generations to have adopted new forms of child-rearing practices giving rise to authoritarian and prejudiced children.

#### 4.2. Social Identity Theory

The individual difference approaches make the assumption that people's behavior in group settings is essentially similar to their behavior in all other situations. However, in group settings, people's behavior is often qualitatively different. This led Tajfel to suggest during the late 1970s that it is important to distinguish between interpersonal behavior and intergroup behavior, where the former means acting as an individual with some idiosyncratic characteristics and a unique set of personal relationships with others and the latter means acting as a group member. Any social behavior falls somewhere along a continuum defined by these two extremes. At the group level of contextual analysis of intergroup relations, the social identity theory, originally formulated by Tajfel and colleagues, has blossomed within the European context into a primary international approach to intergroup relations.

Social identity theory proposes that the social part of one's identity derives from the groups to which one belongs. By favorably comparing attributes of one's own groups with those of out-groups, this theory suggests that a person acquires both a positive sense of who he or she is and a clear understanding of how he or she should act toward in-group and out-group members. From a social psychological perspective, therefore, intergroup discrimination can be perceived as an attempt to establish a positively valued distinctiveness for one's in-group in order to achieve or maintain a positive social identity, that is, a positive collective self.

In this perspective, prejudice is perceived as stemming partly from the normal cognitive process of categorization; that is, people accentuate differences between categories and minimize differences within categories. They view categorized groups as homogenous and privilege category-confirming information in memory as well as category-confirming covariation in judgment. This means that prejudice-consistent information is attended to and rehearsed more frequently, and is remembered more accurately, than is inconsistent information.

Studies from a broad range of nations have shown that members of various ethnic groups often maintain biased systems of belief about the positive qualities of their own groups and about the negative qualities of other groups. For example, group membership determines group-serving attributions in that such membership can completely reverse the pattern of attributions made about an individual's behavior. Positive behaviors performed by members of one's own group are believed to arise from internal dispositions, whereas their negative behaviors are seen as the result of external forces. Similarly, desirable behaviors of out-group members are seen as externally caused, whereas undesirable behaviors are seen as internally caused.

However, there is a danger in overemphasizing the categorization approach to intergroup bias. It may suggest that prejudice might be inevitable, or even desirable, for cognitive adaptation. Although perceptual schemas undoubtedly simplify information, it is not clear that linguistic categories must do so. Language can be used to both categorize and particularize. The difference between prejudiced views and unprejudiced views lies in the content of these views and not the structure. It is true that people tend to interpret information in accordance with their category-based expectations. For example, if they are confronted with information about a group that disconfirms their stereotype of that group, they tend to perceive that information as nonrepresentative of the group; that is, they use the stereotype to "explain away"

deviant information. However, what is crucial here is the “category-based expectations”; whereas these expectations regarding out-groups are likely to be negative for a prejudiced individual, they may be as positive for the in-group as for out-groups for a tolerant individual. In the latter case, the person would need to explain away (i.e., particularize) the additional negative information as opposed to the additional positive information.

### **4.3. The Role of Perceived Threat and Unequal Position in Intergroup Bias**

An in-group is a set of people who perceive each other as having something in common. There are cultural differences in the way in which in-groups are defined; that is, the bases for judging other people as in-group members can be different. There are also degrees of being in-group members; that is, people identify with in-groups in varying degrees. Typically, the more people perceive themselves as belonging together because they experience a common fate, have a common ancestry, or are similar in their gender, age, ethnicity, geographic origin, language, religion, social class, occupation, way of life, attitudes, beliefs, norms, role definitions, and/or values, the more they see each other as in-group members. Also, in-groups are dynamic constructs; that is, people change their perceptions of in-groups and out-groups according to the situation. An out-group consists of a set of people who are not members of the in-group. Members of the in-group are likely to refer to out-group members as “them” and to other in-group members as “us”.

Research has identified a number of key moderators of intergroup bias. Potential moderators range from culture to education. Higher education is generally related to less intergroup bias. Culture as a moderator is discussed in more detail later. Other moderators studied include perceived threat, power or status, group size, identification, and self-esteem. Threat is a central explanatory concept in several theories of intergroup bias, although its interpretation varies greatly. Threat can be perceived in terms of the in-group’s social identity, its goals and values, its position in the hierarchy, and even its existence. Threat can be realistic (e.g., intergroup competition over scarce resources) or symbolic (e.g., blocking the in-group’s values, customs, and traditions). Perceived threat increases intergroup bias.

Intergroup relations among real groups tend to involve groups of unequal power. In general, members of high- and equal-power groups show more bias than do

members of low-status groups, and discrimination by members of numerical minorities with high power is especially strong. Although size, status, and power tend to become confounded outside of the laboratory, these factors have also been shown to have independent effects; that is, groups in numerical minorities express more bias than do those in numerical majorities, regardless of whether the groups are real or artificial. However, when identification is experimentally induced, both majority and minority groups show bias. Members of low-status groups show more bias when status differentials are perceived as unstable and/or illegitimate and when group boundaries are perceived as impermeable. However, low-status groups may simultaneously show out-group favoritism, especially on status-relevant evaluation and when they define their inferiority as legitimate and stable.

### **4.4. Identification and Strategies for Escaping an Unsatisfactory Social Identity**

A central idea in social identity theory is that biased intergroup comparisons are directly linked to social identification. Presumably, the more important a group is to its members, the more bias group members should show in its favor. However, in many studies, there have been rather unstable correlations between group identification and in-group bias. Hinkle and Brown suggested that the psychological processes proposed by social identity theory might not be operative in all groups. They hypothesized that this would depend on the prevailing level of individualism or collectivism in the group or group members, on the one hand, and their inclination to engage in intergroup comparisons, on the other. Group comparisons may be more important for relational groups (e.g., sports teams) and less important for autonomous groups (e.g., families). A strong link between group identification and in-group favoritism would be expected only in groups simultaneously characterized by collectivist and relational orientations. Some support has been found for this idea.

For low-status groups, the outcome of available group comparisons is often negative for their self-esteem. One reaction to this is simply to leave the group, and there are many examples of members of “inferior” groups distancing themselves physically or psychologically from their groups. Such an individualist strategy of social mobility might not always be possible, especially if the group boundaries are relatively fixed and impermeable, as is

the case with many ethnic and religious groups. In cases such as this, social identity theory suggests that a number of other avenues may be pursued. One is to limit the comparisons made to other similar or subordinate groups. Another is to sidestep the main dimensions of comparison and to either invent new dimensions or change the value of existing dimensions. These ways are expressions of social creativity. Yet another route is to directly confront the dominant group's superiority by agitating for social and economic change.

Which of these tactics will be chosen might well depend on the prevailing social climate. If it is such that no real alternatives to the status quo may be conceived, subordinate groups are unlikely to openly challenge the existing order and attempt social change. A key factor in generating social unrest among subordinate groups is a sense of relative deprivation, either in relation to their own groups in the past or (more often) in relation to the dominant group. The implications of broader social and cultural influences for intergroup bias are discussed further in the next two sections.

## **5. CULTURE AS A MODERATOR OF INTERGROUP BIAS**

### **5.1. Individualist and Collectivist Cultures and Individuals**

Research on culture as a moderator of intergroup bias is often conducted within an approach to cross-cultural psychology that combines individual and cultural levels of analysis and focuses on the self-concept and values. This approach has offered a perspective, the key concepts of which are individualism and collectivism (I–C). Originally, these terms were employed in reference to dimensions that differentiate cultures, describing individualist societies as those emphasizing autonomy and emotional independence and describing collectivist cultures as those showing concern for solidarity, sharing, and interdependence.

Individualist cultures, such as those found in Western societies, are defined by certain tendencies. First, people in individualist cultures see the self as independent of groups, whereas people in collectivist cultures (e.g., those found in Asia, Latin America, and most traditional societies) see the self as an aspect of one or more groups. This implies that in collectivist cultures, the goals of the in-group tend to dominate over personal goals, in-group norms are slightly more important than personal attitudes, and communal

relationships tend to be more important than relationships based on personal gain. Such differences result in different definitions of the in-group and out-group and in different pressures to conform to in-group norms, with major implications for intergroup relations.

However, Triandis and Trafimov in 2001 emphasized that individuals in different cultures include in their cognitive systems both individualist and collectivist elements in different mixtures, depending on the situation. Thus, the same constructs (I–C), when studied at the individual level of analysis, are renamed idiocentrism–allocentrism (I–A), although a similar distinction was also proposed by Markus and Kitayama during the early 1990s in their description of independent and interdependent selves. Triandis has argued that the social identity of allocentrics and collectivists is more core and salient than that of idiocentrics and individualists. Consequently, competition with out-groups and ethnocentrism also are more readily apparent in the former. These contentions have yet to be tested in the realm of ethnic perceptions and relations, but researchers consider I–C and I–A to be useful constructs in the analysis of ethnocentrism in a cross-cultural context.

### **5.2. Empirical Evidence and Limits of Generalizability**

Most commentators would agree that the hypotheses of social identity theory are more likely to be supported in collectivist cultures than in individualist ones. However, Lee and Ward during the late 1990s found only partial support for the hypothesis of stronger ethnocentrism among allocentrics. Although allocentrics had more positive in-group attitudes than did idiocentrics, no difference was found between allocentrics and idiocentrics on out-group attitudes. Thus, in-group favoritism in both Malays and Chinese was due to the ethnocentrism bias of allocentric respondents. These findings are in broad agreement with Triandis's theorizing on the heightened significance of social identity, social categorization, and social comparison for allocentrics. The evidence for allocentric ethnocentrism, however, was limited to an in-group bias. Contrary to Triandis's suggestions, there was no distinction in the out-group attitudes held by idiocentrics and allocentrics in Lee and Ward's study.

The taxonomic model proposed by Hinkle and Brown suggests that the relationship between group identification and intergroup bias would be strongest in groups that are both relational and collectivist,

whereas this link would be weak in individualist and autonomous groups. Although there is some empirical support for this model, there is little information as to whether a relational orientation is important within collectivist cultures. Triandis suggested that members of collectivist cultures are not interested in comparing themselves with out-groups. The social comparison in which allocentrics engage might not be quite the same as that envisaged by social identity theory. Allocentric group members might not be so preoccupied with their identity as group members. Instead, they may focus on their group as a whole and on threats to its well-being.

Although Triandis has taken great pains to distinguish I-C as a cultural-level label and I-A as an individual difference measure, the terms have been used interchangeably in the cross-cultural literature. In comparative studies, the constructs frequently have been operationalized through the examination of individuals from two or more cultures either a posteriori labeled or a priori selected on the basis of some overarching conceptualization of I-C. In the more sophisticated investigations, the selection, definition, and classification are based on previous empirical research, but even this approach can pose inadvertent problems, and researchers often fall prey to the "ecological fallacy," that is, the assumption that because two cultures differ on a specific dimension, individual members of those countries will reliably differ in the same predictable manner.

Lee and Ward's study among Chinese and Malays in Singapore provides a case in point. Singapore has been previously labeled and commonly classified as a collectivist culture. Triandis has argued vigorously that members of collectivist cultures are expected to display pronounced ethnocentric attitudes. However, the ethnocentric biases that Lee and Ward found in their study were limited to allocentric subjects. This highlights the necessity of the distinction between I-C and I-A as well as the awareness and acknowledgment of theory generalizability.

## **6. THE ROLE OF HISTORY AND SOCIAL CONTEXT IN INTERGROUP BIAS**

### **6.1. History and Social Context: Neglected Factors in Research**

Norms, beliefs, behavioral intentions, feelings, and stereotypes determine intergroup behavior. These may be both the result of the history of previous intergroup

contact and the cause of future intergroup behavior. However, they may also be affected by factors other than previous intergroup contact. Sometimes it is in the interest of certain members of a group to induce other members of that group to have negative thoughts about a different group. More generally, a variety of social factors (e.g., relative power of the in-group and out-group; zero-sum conflict resulting in competition, domination, or scapegoating) affect norms (e.g., antagonism, paternalism, persecution, submission, rebellion, repression, conciliation), stereotypes (e.g., threatening, inferior, superior, oppressive, powerful, weak), feelings (e.g., accepting/rejecting), and behavioral intentions (e.g., to hurt, to derogate, to punish, to submit, ambivalence) regarding out-groups.

Collective representations about the past influence current intergroup relations, perceptions, and behavior. One should be careful not to neglect the importance of more enduring identities with long-term commitments and connections to, for example, former and future generations. A sense of enduring group membership may be related to cultural features such as a strong collectivist value orientation in which "in-group-centrism" predominates and identification is more of a nominal variable than a continuous one. The history of intergroup conflict has generally been neglected in the social psychological literature. Especially in collectivist cultures, history is often a major factor in intergroup relations (e.g., the case of the former Yugoslavia).

Groups and group members derive their identities from particular contexts. However, the notion of context is ill defined across many psychological paradigms. For example, context can be taken to refer to the particular task or activity in which people are engaged, to historical and cultural circumstances, to general immigrant conditions, and/or to actual social situations (e.g., schools, neighborhoods, workplaces). Many existing psychological models are not specific enough for the study of ethnic minority groups. An understanding of such groups would require explicit attention to negative social circumstances, such as racism and discrimination, in relation to concrete environmental influences.

### **6.2. The Special Circumstances of Minority Groups**

Many real-life intergroup contexts consist of groups that hold either a minority or a majority position vis-à-vis each other. In real life, numerical asymmetries often, but not necessarily, covary with power or status asymmetries.

Simon and colleagues described the so-called cognitive-affective crossfire of minority members as follows. On the one hand, numerical distinctiveness makes group membership highly salient and difficult to ignore, both for minority members themselves and for others who may continually remind them of it by word and deed. In this way, group membership becomes a much more central self-aspect for minority members than for majority members. On the other hand, minority membership entails specific risks and stressful experiences that may be largely unknown to majority members. These risks and the negative affective consequences are further exacerbated when a numerical inferiority is also associated with status and/or power inferiority. In other words, there are stronger cognitive forces pushing minority members toward their group (or keeping them in it), while at the same time there may also be stronger affective forces pulling them away from it (or keeping them out of it).

Depending on the perceived opportunity structure of the social context (e.g., the permeability of group boundaries), minority members may opt for individualist strategies involving psychological dis-identification or actual exit from their group, for collectivist strategies involving assertive intergroup behavior, or for a combination of both. Majority group members can be more “mindless” than minority group members regarding group membership, at least so long as the former do not feel threatened by an assertive minority. Both minority and majority group members are likely to adopt an intergroup perspective when minority group members show assertive intergroup behavior to escape from or cope with the cognitive-affective crossfire (or worse) because such assertiveness may incite defensive intergroup behavior on the part of majority group members. Note that, on the surface, the assertive behavior of a minority and the defensive behavior of a majority may take the same form such as intergroup discrimination. From a theoretical point, however, the underlying meanings are different because such phenotypically similar behavior serves different purposes for minority and majority members.

### **6.3. The Special Circumstances of Immigrants**

Historical circumstances, group norms, and ideologies also play a crucial role in the intergroup relations between immigrants and host populations. Immigrant groups, as well as individual immigrants, arrive in a new country with differing attitudes about retaining their culture of origin and becoming part of the new

society. In the new society, these attitudes interact with the actual and perceived levels of acceptance of immigrants and with official policies toward immigration. For example, assimilationist societies expect immigrants to abandon their own cultural and linguistic distinctiveness for the sake of adopting the culture and values of the dominant group, whereas ethnist societies often define who can be citizens of the state in exclusive terms.

The particular norms and ideologies reflected in the public rhetoric on the integration of immigrants provide the specific environmental context within which hosts and immigrants develop their cognitive representations of each other. They can perceive each other as one single in-group, as two subgroups within a common in-group, as two separate groups, or as separate individuals.

The most harmonious intergroup relations are predicted when both host community members and minority group members share integration, assimilation, or individualist orientations. In contrast, negative intergroup stereotypes and discriminatory behaviors can result when immigrant group members prefer integration but host community members insist that immigrants assimilate to the host society. However, host community members who want to segregate or exclude immigrants are likely to foster the most conflictual intergroup relations with targeted immigrant groups. These host community members are likely to have negative stereotypes regarding immigrants and to discriminate against them in many domains. Nonviolent segregationists may simply think that cultures should not mix, whereas radical exclusionists are more likely to launch violent racist attacks against immigrants. Of the targeted immigrant groups, it is those with separatist attitudes that are most likely to resist or even retaliate against host community persecutions. This promising so-called interactive acculturation model, developed by Bourhis and colleagues during the late 1990s, has received some empirical support.

### **6.4. Is Intergroup Bias Ubiquitous?**

Stereotypes are likely to be particularly clear and consensually held if one's group has a long history of dealings with an out-group. If groups share a history of conflicts, out-group stereotypes are often extreme, simple, negative, and symmetrical, with each group rating its own group members positively while denigrating the out-group members. In general, group members rate their own group members higher on beneficence (e.g., honest, kind, loyal, trustworthy)

than they do out-group members. Out-group members do the same, resulting in a “mirror image” pattern.

The critical question for intergroup harmony is the size of the gap between the in-group and out-group ratings, especially on perceived values. Stereotypes can serve the societal function of acting as ideological representations that are used to justify and legitimize existing social and power relations within society. People can hold intense stereotypes about persons from other cultural and ethnic groups, even though they have never met them. These polarized attitudes probably arise from the home, the media, and the educational curricula. However, there are studies that have not found in-group bias in stereotyping. It is obvious that one must consider the effects of prevailing social and cultural norms on the expression of intergroup bias.

### **6.5. The Influence of Social and Cultural Norms on the Expression of Prejudice**

Theoretically, prejudice and stereotypes are related but still distinct concepts. Attitudes are evaluative responses, whereas stereotypes are specific beliefs about a group. Responses to explicit measures of intergroup bias are given consciously. Typically, explicit attitudes are assessed by self-report measures, including attribution of group traits (stereotypes), group evaluations (prejudice), and differential behavior toward in-group and out-group targets (discrimination). However, measures of these cognitive, affective, and behavioral components, respectively, are often only modestly or weakly related in most Western societies.

Explicit attitudes and stereotyping operate in a conscious mode, whereas implicit attitudes and stereotypes are evaluations and beliefs that are automatically activated by the mere presence (actual or symbolic) of the attitude objects. Implicit attitudes commonly function in an unconscious fashion and are typically assessed using response latency procedures, memory tasks, physiological measures (e.g., galvanic skin response), and indirect self-report measures. Thus, the promise of implicit measures is to assess the true extent of people’s bias given pressures to conform to socially desirable or politically correct norms. When such norms prevail, individuals may be motivated to express their prejudice in indirect ways that do not threaten their nonprejudiced self-images.

The translation of prejudiced beliefs (stereotypes) and evaluations (attitudes) into action (discrimination)

depends on both personal and social factors. Highly prejudiced persons feel less guilt and shame, less ambivalence, and less internal conflict about discrepancies between internal and/or social standards of egalitarianism and actual discriminatory behavior. Less prejudiced individuals are self-critical about their departures from more egalitarian standards prevailing in their cultural and social environments. There may be a greater correspondence between implicit and explicit attitudes for issues that are not socially sensitive than for those that are socially sensitive or that are associated with more prejudiced historical norms or traditional socialization.

However, there is more to discrimination than personal proclivities. Social groups may feel that they are under political and economic threat by other groups or may be fueled by ideologies of antagonism passed down from the past and sustained by racist talk common to in-group members. These social inducements to discrimination may be further unrestrained by a wider political ideology of intergroup harmony (e.g., Canada’s policies of multiculturalism and human rights observance). Such a volatile mix can be exacerbated by societal factors and cultural values supporting a sense of cultural superiority.

Discriminatory behavior need not result in open intergroup hostility if the social ideology surrounding differential access to material and social resources is accepted by members of all competing groups. Alternatively, unequal group access to resources may prevail and be resented by members of groups facing discrimination. However, the consequences of challenging this order may be too destructive to contemplate, or the social basis for group mobilization may be lacking. Conflict behavior is less in societies with crosscutting ties than in societies with ties that reinforce polarization between groups and in societies where scapegoating ideologies concerning specific out-groups are firmly entrenched.

## **7. THE REDUCTION OF INTERGROUP BIAS**

### **7.1. The Contact Hypothesis**

Intergroup bias has psychological, social, and cultural components; hence, both individual and intergroup approaches can make important contributions to the reduction of intergroup bias. Some interventions attempt explicitly to decrease bias on an individual level by invoking guilt for holding socially undesirable attitudes. Although this approach may sometimes seem successful, it capitalizes on the good intentions of

low-prejudiced people (whose bias is surely not the main problem) and demands appropriate levels of awareness, effort, and practice over time. Other individual-level approaches use more indirect means such as inducing empathy for out-groups.

However, most of the interventions used for reducing bias employ more intergroup methods and focus on the need to increase the quantity and enhance the quality of intergroup contact. This approach derives from the so-called contact hypothesis, which was originally developed by Allport during the 1950s and is one of the most influential ideas for the reduction of intergroup prejudice. Broadly speaking, it suggests that contact between members of different groups, under the appropriate conditions, lessens intergroup prejudice and hostility.

In 1998, Pettigrew's review of recent research pointed to the increased knowledge of the mediating processes by which contact can reduce intergroup bias and anxiety. Allport held that the positive effects of intergroup contact occur only in situations marked by four key conditions: equal group status, common goals, intergroup cooperation, and the support of authorities, laws, or customs. Many other factors have since been advanced, for example, the acquaintance potential and intimacy of the contact. Some researchers emphasize equal group status coming into the situation as part of the equality requirement, whereas others find this to be less important than group status within the situation. Social and institutional support means that those in authority (e.g., the politicians implementing new legislation and the judges and police monitoring its administration) should be unambiguous in their endorsement of measures designed to promote greater contact and integration. Creating a social climate in which more tolerant norms can emerge, and sanctioning the violation of such norms, constitutes the essence of social and institutional support.

Allport's formulation continues to receive support across a variety of situations, groups, and societies. Most studies report positive contact effects, even in situations lacking key conditions. Overall, face-to-face interaction between members of distinguishable groups is importantly related to reduced prejudice, and intergroup contact relates to a wide range of prejudice measures (e.g., belief, social distance, stereotype) and affective measures. Testing the possibility of an opposite causal sequence (i.e., that prejudiced people avoid contact more than contact lowers prejudice), it has been shown that the path from contact to reduced prejudice is stronger than the other way around. Although it concentrates on school and housing situations, the contact literature ranges from Chinese students in the United

States and interracial workers in South Africa to German and Turkish school children and Americans getting to know Southeast Asian immigrants.

## **7.2. Mechanisms by Which Contact Reduces Intergroup Bias: The Role of Category Salience**

The contact hypothesis does not specify the processes that are responsible for the positive effects or how the positive effects of contact generalize beyond the immediate situation to other situations, from the specific out-group member encountered to the out-group as a whole, or to other out-groups not involved in the contact. Recent work suggests that four interrelated processes operate through contact and mediate attitude change: learning about the out-group, changing behavior, generating affective ties, and in-group reappraisal.

Although cognitive research has uncovered a host of mechanisms that limit learning material that counters our attitudes and stereotypes, the research literature suggests that positive effects are more common than the contact hypothesis or cognitive analyses predict. The basic reason is that learning about the out-group is only one of several processes involved. For example, optimal and continuous intergroup contact has the potential to produce attitude change by revising attitudes in accordance with behavior and by reducing the anxiety often experienced in intergroup encounters. Positive emotions aroused by intergroup friendship may be pivotal in reducing intergroup prejudice. It has been shown that having an out-group friend may even generalize to other out-groups besides that of the particular friend. In addition, optimal intergroup contact reduces prejudice by providing insight about the in-group as well as the out-group, leading to a less provincial view of out-groups in general.

As for the generalization of contact across situations, research has generally shown that positive effects on intergroup attitudes might need repeated optimal situations. Research supporting the salient categorization model proposed by Hewstone and Brown during the 1980s shows that stereotype change generalizes best to the intergroup level when the individuals involved are typical out-group members. This approach argues that, to protect against loss of distinctiveness for groups involved in contact, two factors are important. First, the salience of group boundaries should be maintained during contact (i.e., group memberships are made salient or the partner is typical [rather than atypical] of the out-group as a whole) to promote generalization across members of

the target group. Second, groups should recognize and value mutual superiorities and inferiorities.

Thus, there seems to be some advantage in maintaining the original group boundaries in the interaction while otherwise optimizing contact conditions. Positive contact is more likely to be associated with favorable out-group attitudes when contact takes place with a typical out-group member. However, this poses a problem. Typical members of real groups are different in many ways, but people tend to seek out people with similar interests and status. Maintaining category distinctions during contact involves the risks of reinforcing perceptions of group differences and increasing intergroup anxiety, thereby increasing bias.

One way of maintaining category salience while still diminishing anxiety associated with intergroup contact is offered by a recent development of the contact hypothesis called the extended contact effect. This model proposes that the awareness that one's fellow in-group members have close friendships with out-group members can help to reduce prejudice toward the out-group. A likely reason for this is that the in-group "exemplars" provide normative information regarding how one should behave and also may contribute to the redefinition of the intergroup relationship as less negative. Such an "extended contact effect" does not require personal intergroup friendship for the perceiver. Correlational and experimental evidence of this idea is emerging. Underlying this model is the idea that both in-group and out-group role models should be seen as typical or representative of their groups and not as exceptions to the rule.

### **7.3. The Decategorization and Recategorization Models for Reducing Intergroup Bias**

The so-called decategorization model of intergroup relations holds that intergroup contact is more effective when group salience is low and the intergroup situation is personalized. However, individual-to-group generalizations are unlikely in very personalized contact situations. In contrast, the so-called common in-group identity model holds that intergroup attitudes are best improved when the category boundaries are drawn so that the out-group becomes subsumed into a new and larger superordinate category. There is now an impressive literature supporting this claim. However, most studies of this model have not examined generalization. The fundamental limitation of both the decategorization and recategorization models is that they threaten to deprive individuals

of valued social identities in smaller, less inclusive groups. A common in-group identity may be only short-lived or unrealistic in the face of powerful ethnic/racial categorizations (e.g., the former Yugoslavia). In addition, for groups with histories of antagonism and for minorities whose members are likely to resist assimilation into a superordinate category that is dominated by a majority out-group, the prospect of a superordinate identity may constitute a threat that actually increases bias.

### **7.4. Integrating Different Models of Contact**

To overcome these problems, various solutions have been proposed. Hewstone and colleagues in 2002 proposed that the categorization salience model of contact needs to be integrated with the more personalized models and that, in contrast to original social identity theory, interpersonal and intergroup contact should be viewed as orthogonal dimensions that together can create highly effective conditions of out-group contact (i.e., contact should simultaneously be both highly intergroup and highly interpersonal).

The dual-identity model developed by Gaertner and colleagues takes this idea as a starting point as it aims to maximize the benefits of both the "common in-group identity" and the mutual intergroup differentiation model. According to this model, subgroup identity is maintained while the groups also share membership in a common superordinate group. A dual identity leads to more positive out-group attitudes than does a superordinate identity alone. Gaertner and colleagues conducted a correlational study in a multiethnic high school and found less bias when students identified themselves at both the subgroup (i.e., ethnic group) and superordinate (i.e., American) levels rather than at the subgroup level only.

The dual-identity model resembles the bidimensional interactive acculturation model suggested by Bourhis and colleagues. According to this model, as well as to the original acculturation theory developed by Berry during the 1980s, members of cultural minorities can adopt different acculturation strategies. These strategies are formed by the crossing of two orthogonal orientations: in the interactive acculturation model, these are defined as identification with one's own culture (high/low) and identification with the dominant culture (high/low). Research suggests that the "intergration" strategy (high on both orientations) is often the preferred strategy among minority groups and seems to be associated with less acculturative stress and better educational outcomes.

A major problem for the dual-identity approach is that members of minority and majority groups may have different preferences for what model of intergroup relations to adopt. Dominant majority ethnic groups tend to favor assimilation, whereas racial/ethnic minorities favor pluralistic integration. Thus, dual identity may reduce bias for the minority group but not for the majority group. This asymmetry between minority and majority strategies is also reflected in the interactive acculturation model itself in that the strategies of majority members concern only their preferences for the identification strategies they would like the minority members to adopt. As a consequence, minority and majority members can, at best, agree only on the most preferred strategy for the minority group members. Even then, majority members might not be prepared to change their own identification patterns accordingly.

A successful superordinate category and identity must be inclusive rather than reflect the superior characteristics of a dominant majority group too strongly. Even if this is the case, however, subgroup categorizations and identities are likely to be stronger and more stable than superordinate ones. One simple way in which to reduce intergroup bias is to ensure that the contact partners can be simultaneously classified as in-group and out-group members on multiple dimensions. Shared or overlapping category memberships reduce bias simply because they make social categorization more complex and decrease the importance of any one in-group/out-group distinction. However, the effectiveness of crossed categorization is limited when one category distinction is functionally dominant and/or categories are correlated, and this is normally the case in real-life intergroup encounters.

All of the models for reduction of intergroup bias presented here should be seen as complementary and reciprocal rather than as competing and exclusive. This integrative approach responds to the fact that each model can be effective under particular contact conditions but that each model also has its weaknesses and limitations, particularly with respect to generalization and to finding an intervention that works for both majority and minority groups.

One way in which to combine the various models was proposed by Pettigrew in 1998. He suggested that time sequence is crucial. Diminished category salience can be important when intergroup contact is initiated, but category salience may be required later for the positive effects to generalize to the intergroup level. At an even later stage, the recategorization strategy proposed by the common in-group identity model may become possible. After continuous contact, people may begin to think of

in-group and out-group members as members in a more inclusive common in-group, although this final state is seldom reached by interacting groups.

### **7.5. Realistic Goals for Real-Life Intergroup Relations**

Prior attitudes and experiences influence whether people seek or avoid intergroup contact and what the effects of the contact will be. High intergroup anxiety and threat can impede both contact and its positive effect. In addition, institutional and societal norms structure the form and effects of contact situations. Optimal contact situations are difficult to achieve if a struggle for power fuels the larger intergroup conflict. In contrast, normative support makes attainment of optimal conditions much easier. Although the beneficial effects of contact with typical out-group members have been confirmed in a number of studies, if the intergroup encounter goes wrong, structuring the encounter as intergroup rather than interpersonal could make matters worse. Therefore, there is a fine line to be drawn in designing situations of intergroup contact. Enough category salience must be retained to permit the positive change to generalize, but not so much as to allow the regression into familiar patterns of prejudice that are associated with category divisions that are too firmly drawn.

Given the corrosive nature of many real-world conflicts, out-group liking is often unlikely; and achieving other outcomes may be more realistic and just as important, including increases in perceived group variability, out-group knowledge, and perspective taking as well as decreases in intergroup anxiety. To reduce full-blown intergroup conflict, effective interventions also need to build trust. It is true that relevant research and theoretical developments on intergroup bias have been directed at its relatively mild forms. However, social categorizations play a role in even extreme forms of prejudice. Severely prejudiced individuals and ideologies delegitimize victims, assigning them to an extreme social category that enjoys no projection, and dehumanize victims by morally excluding them from the in-group of humanity where justice, fairness, and morality prevail.

However, real-world intergroup relations owe at least as much to intergroup history, economics, politics, and ideology as they do to social psychological variables such as categorization, self-esteem, in-group identifications, group size, and group threat. Social conflict is more complex than intergroup bias. Consequently, intergroup bias should not be analyzed from just one disciplinary

perspective, regardless of how important the psychological antecedents and consequences of intergroup bias may be. The current trend to study culture in social psychology has yet to address intergroup bias in full force, but there is little doubt that it will. Culture and societal influences are neglected at our peril.

### See Also the Following Articles

Motives and Goals ■ Prejudice and Discrimination  
 ■ Stereotypes

### Further Reading

- Brown, R. (2001). Intergroup relations. In M. Hewstone, & W. Stroebe (Eds.), *Introduction to social psychology* (pp. 480–515). Oxford, UK: Blackwell.
- Fiske, S. (2000). Stereotyping, prejudice, and discrimination as seen between the centuries: Evolution, culture, mind, and brain. *European Journal of Social Psychology*, 30, 299–322.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575–604.
- Liebkind, K. (2001). Acculturation. In R. Brown, & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 386–406). Oxford, UK: Blackwell.
- Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49, 65–85.
- Smith, P. B., & Bond, M. H. (1999). *Social psychology across cultures* (2nd ed.). Boston: Allyn & Bacon.
- Triandis, H. C., & Trafimov, D. (2001). Culture and its implication for intergroup behavior. In R. Brown, & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 367–384). Oxford, UK: Blackwell.
- Verkuyten, M. (2000). The benefits to social psychology of studying ethnic minorities. *European Bulletin of Social Psychology*, 12(3), 5–21.



# International Classification of Diseases (WHO)

Jesús Rodríguez-Marín

Miguel Hernandez University, Alicante, Spain

1. Introduction
  2. International Statistical Classification of Diseases and Related Health Problems, 10th Revision
  3. ICD-10's Chapter 5: Mental and Behavioral Disorders
  4. Phenomenological Rigor, Consistency, and Diagnostic Stability of ICD-10 Psychiatric Diagnoses
  5. Clinical Utility of ICD-10
  6. Comparison between ICD-10 and DSM-IV
  7. Collaborating Centers for Classification of Diseases and Procedures for ICD-10 Updating
- Further Reading

The *International Statistical Classification of Diseases* (ICD) was worked out and proposed by the World Health Organization (WHO) to facilitate communication and study on organic and mental diseases among health workers and clinicians worldwide. In 1893, Jacques Bertillon, head of statistical works for the city of Paris, presented a new classification of *Diseases and Causes of Death* to the International Congress of Statistics in Chicago. The Bertillon classification was the first of a series that gave rise to the ICD, published by WHO from 1948 onward, to improve diagnosis and classification of diseases and to promote international comparability

in the collection, processing, classification, and presentation of mortality and morbidity statistical data. Since then, the ICD has been revised periodically to incorporate changes in the medical field. This article includes a historical approach to the ICD, and its general principles and structure. The article specifically discusses Chapter 5 of the 10th revision of the ICD (ICD-10), on mental and behavioral disorders, describing the phenomenological rigor, consistency, diagnostic stability, and clinical utility of the psychiatric diagnoses. Finally, it incorporates a brief comparison of ICD-10 and the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and mentions WHO collaborating centers for classification of diseases and procedures for ICD-10 updating.

## 1. INTRODUCTION

International collaboration began a century ago with the aim of drawing up a classification of diseases. The systematic classification of diseases began in the 18th century with the publication of the *Nosologia Methodica* by Lacroix and of the *Genera Morborum* by Linnaeus. In 1785, Cullen published the *Synopsis Nosologiae Methodicae*, which was used in various European countries until the middle of the 19th

century. A classification of diseases and death causes was approved in 1855 during the Second International Congress of Statistics in Paris. This classification, which combined the English tradition (based on anatomical region) and the French tradition (based on etiology), had wide repercussions but did not garner complete acceptance.

In 1893, Jacques Bertillon, head of statistical works for the city of Paris, presented a new classification of *Diseases and Causes of Death*, based on the one employed in Paris, to the International Congress of Statistics in Chicago. The Congress approved it and recommended its use and revision every 10 years. The Bertillon classification of diseases was the first of a series that gave rise to the *International Statistical Classification of Diseases* (ICD), edited and published by the World Health Organization (WHO) from 1948 onward, to improve diagnosis and classification of diseases and to promote international comparability in the collection, processing, classification, and presentation of mortality and morbidity statistical data. Since then, the ICD has been revised periodically to incorporate changes in the medical field. Bertillon himself did the work in 1900, 1910, and 1920. WHO assumed the task from the 4th revision (in 1929) onward. The 4th revision and the next one (in 1938) maintained the basic structure of the Bertillon classification. In both of these, the mental diseases were included without any special section; instead, they were distributed throughout 9 of the 14 sections.

In 1948, WHO's 6th revision included a chapter specifically dedicated to mental diseases (Chapter 5), and this remained unchanged until the 9th revision (ICD-9). This new chapter contained three sections dedicated to (a) psychosis, (b) psychoneurotic disorders, and (c) disorders of character, behavior, and intelligence. The 1955 edition dedicated more space to mental disorders but maintained these three sections.

Since the beginning of the 1960s, WHO has stimulated a series of works on mental disorders, supported projects to improve diagnosis and classification of mental diseases, and encouraged research on classification instruments. Emphasis was placed on the reliability, consistency, and diagnostic stability of classification instruments. These efforts culminated in a program with four main subprograms dedicated to (a) the improvement of the diagnosis and classification of psychiatric disorders, (b) cross-cultural research of the most important psychiatric disorders, (c) an epidemiological study, and (d) training in social and psychiatric epidemiology. A network of individuals and institutions

was created, and they worked on proposals to be included in the draft for the 8th ICD revision (in 1968). In this 8th revision, the diagnostic categories for mental diseases were reorganized into three sections: (a) psychosis; (b) neurosis, personality disorders, and other nonpsychotic mental disorders; and (c) mental retardation. A glossary defined each diagnostic category and standardized its use.

The 9th revision of the ICD (in 1978) contained four sections dedicated to mental disorders: (a) organic psychosis; (b) other psychoses; (c) neurotic disorders, personality disorders, and other nonpsychotic mental disorders; and (d) mental retardation. The difficulties of diagnosis and classification of some nosologic entities revived interest and research into psychiatric diagnosis, and WHO again took on the leadership in the preparation of the 10th ICD edition.

The rules and study areas for classification of mental disturbance were defined during an important conference in Copenhagen in 1982. Also identified were those areas in which improvement in classification required further research work and more evidence. A trial classification was then staged in 55 countries with the participation of more than 700 clinicians from at least 100 centers, coordinated by a group of 13 centers. More than 15,000 patients were assessed in the trial classification.

Some special notes should be made here with regard to the development of disease classification in the 10th revision (ICD-10) in 1992. First, it was developed simultaneously in several languages, and it enabled compatible definitions among many countries. Second, the different versions of ICD-10 allowed different uses in accordance with different necessities. Third, the classification was based on the evidence found in research. Fourth, the classification was accompanied by standard instruments for the evaluation of symptoms and other diagnosis elements.

The results of this project were very positive.

## 2. INTERNATIONAL STATISTICAL CLASSIFICATION OF DISEASES AND RELATED HEALTH PROBLEMS, 10th REVISION

In the 10th revision, the title was changed to clarify the content and goals and to reflect the progressive extension of the scope of the classification.

ICD-10 is much longer than the 9th revision. Printed in three volumes (the 9th revision was printed in two volumes), ICD-10 has adopted an alphanumeric coding scheme (rather than using numeric categories) and has significantly enlarged the number of categories available for classification. The first character in the coding system is now a letter for each chapter, as well as a number, and each chapter is assigned 10 two-character categories and 100 three-character major categories. ICD-10 has nearly twice as many categories as did the 9th revision. Finally, some slight changes have been made in the coding rules for mortality. For instance, in the case of Chapter 5 (on mental and behavioral disorders), the three-character categories go from F00 to F99. F20 is “schizophrenia,” and every category has different subcategories (e.g., F20.0 is “paranoid schizophrenia,” F20.5 is “residual schizophrenia”).

The first volume of ICD-10 contains 17 chapters dedicated to different diseases (from “certain infectious and parasitic diseases” to “congenital malformations, deformations, and chromosomal abnormalities”), special tabulation lists for mortality and morbidity, the definitions, and the nomenclature regulations. The second volume includes new background information, instructions, and guidelines for users of the tabular list and historical information about the development of the ICD classification. The third volume incorporates a detailed alphabetical index. The index includes diagnostic terms commonly used as synonyms for the terms officially accepted in the classification, and it facilitates more efficient coding.

The ICD-10 manual includes not only the disorders but also other attributes of patients’ interactions with their social contexts. Consequently, it includes a multi-axial system with three axes, with Axis I covering clinical diagnoses of both mental and physical disorders, Axis II covering disabilities due to impairments produced by the disorders, and Axis III covering the contextual or environmental and personal lifestyle factors influencing the appearance and course of the disorders. The ICD-10 manual is accompanied by the ICD-10 Multi-axial Diagnostic Formulation Form, the WHO Short Disability Assessment Schedule (WHO DAS-S), and Axis I and Axis III glossaries.

ICD-10 is intended for use in clinical, educational, and research activities. It has been prepared by an international team of experts and has been field-tested in 20 countries. It has been shown to be easy to use and applicable in a wide range of cultures and settings.

### 3. ICD-10’S CHAPTER 5: MENTAL AND BEHAVIORAL DISORDERS

ICD-10 dedicates Chapter 5 to mental and behavioral disorders. This chapter, under the letter F of the coding system, substantially extends the space given previously to mental disorders.

Chapter 5 in ICD-10 offers important differences compared with the previous revision. That chapter in the 9th revision had only 30 three-character categories (290 to 319), whereas it includes 100 major three-character categories in ICD-10. Two other important changes have been made. First, differentiation between psychoses and neuroses has been abandoned. Second, disorders showing common elements or properties are grouped together.

Chapter 5 in ICD-10 contains the following sections: F00 to F09, organic (including symptomatic) disorders; F10 to F19, mental and behavioral disorders due to psychoactive substances; F20 to F29, schizophrenia, schizotypal, and delusional disorders; F30 to F39, mood (affective) disorders; F40 to F49, neurotic, stress-related, and somatoform disorders; F50 to F59, behavioral syndromes associated with physiological disturbances and physical factors; F60 to F69, disorders of adult personality and behavior; F70 to F79, mental retardation; F80 to F89, disorders of psychological development; F90 to F98, behavioral and emotional disorders with onset usually occurring during childhood or adolescence; and F99, unspecified mental disorders.

The most recent Chapter 5 also contains clinical guidelines for the classification of mental disorders in several languages; compatible versions dedicated to classification, clinical, and investigation work; and a multi-axial presentation with three axes: clinical diagnosis, the development of factors related to disorders, and the disability level related to them. In addition, ICD-10 includes a version for primary care doctors.

The publication of the ICD-10 classification of mental and behavioral disorders has been an extremely important event for both psychiatry and clinical psychology. For the first time, psychiatry has produced a classification based on the worldwide consensus of experts and schools of psychiatry. ICD-10 constitutes an essential basis that enables the development of a unique language for national and international communication in the field of mental health, facilitating collaboration across cultures and countries.

Chapter 5 of ICD-10 has been translated into more than 35 languages and presented in different versions for

different purposes. Everywhere, the code numbers and titles of the disorders are exactly the same, but details and styles of presentation are adapted to specific purposes. Furthermore, two major documents have been developed from ICD-10: a version for general clinical practice and educational use (*ICD-10 Clinical Descriptions and Diagnostic Guidelines*) and another version for research purposes (*ICD-10 Diagnostic Criteria for Research*).

#### 4. PHENOMENOLOGICAL RIGOR, CONSISTENCY, AND DIAGNOSTIC STABILITY OF ICD-10 PSYCHIATRIC DIAGNOSES

*ICD-10 Clinical Descriptions and Diagnostic Guidelines* and *ICD-10 Diagnostic Criteria for Research* have been analyzed for their internal consistency and phenomenological soundness. The operational definitions of all categories have been considered both for their agreement with the core text of ICD-10 and for their internal consistency within each category. The soundness of all operational criteria and concepts used in the clinical descriptions has also been assessed. The results indicate a high degree of consistency and phenomenological rigor as well as a high level of interrater reliability. Nonetheless, some inconsistencies have been detected in a few categories related to mania, hypomania, cyclothymia, schizotypal disorder, and behavior disorder. Especially in the schizotypal disorder category, a low interrater reliability and poor face validity were found, and this could lead to the label of mental illness being applied to otherwise healthy people.

Temporal stability of diagnosis, a measure of the degree to which diagnoses remain unchanged at later hospital admissions, seems to be high for schizophrenic and affective disorders, whereas it is not so high for neurotic, obsessive-compulsive, stress-related, and adjustment disorders. Further research is required in this area.

#### 5. CLINICAL UTILITY OF ICD-10

ICD-10 has led to a major innovation in clinical psychiatry. The criteria-based classifications are no longer simple coding conventions; instead, they have become part of the conceptual framework of the discipline itself. Consequently, comprehensive clinical evaluation should work at two separate levels: (a) the syndrome

diagnosis and (b) the evaluation of factors possibly influencing the course and outcome of psychiatric disorders. Only when both are taken into consideration is it possible to get a complete understanding of the disorders and optimal treatment.

### 6. COMPARISON BETWEEN ICD-10 AND DSM-IV

During the 1940s, the American Psychiatric Association drew up a psychiatric classification system of mental diseases titled *Diagnostic and Statistical Manual of Mental Disorders* (DSM), and since then it has been the main reference for contrasting the ICD. The DSM is produced at intervals of at least 10 years, reflects U.S. national trends and opinions, and has probably been the most important of the national classifications. Therefore, although ICD-10 is the official coding system in many countries, DSM-IV appears to be more popular among mental health professionals.

Both systems, ICD-10 in the *Diagnostic Criteria for Research* and the 4th edition of the DSM (DSM-IV), have introduced descriptive, nonetiologic classifications with operationally defined categories based on diagnostic criteria and have proposed explicit diagnostic criteria. There are differences between the two systems, with the most significant ones being the avoidance of social dysfunction criteria in ICD-10 and the hierarchical order preventing comorbidity diagnoses in ICD-10. Results from separate studies show high levels of concordance in diagnoses of depression, dysthymia, substance dependence, and generalized anxiety disorder. They also show a moderate concordance in social phobia, obsessive-compulsive disorder, and the three panic/agoraphobia disorders. Results show only a low concordance for posttraumatic stress disorder and drug use/abuse. Concordances and differences basically reflect similarities and dissimilarities in concepts and diagnostic criteria between classification systems.

Both systems have multiaxial presentations, but it is doubtful that the ICD-10 multiaxial presentation could be used to the same extent as can the DSM-IV multiaxial assessment. ICD-10 has sought to place psychiatric disorders in the context of the different religions, nationalities, and cultures of the world community, whereas DSM-IV lacks this degree of cultural sensitivity, perhaps because it is a system that originates from only one country. The concern for cultural universality increased the length of the document but curtailed and

simplified some operational definitions by omitting culturally biased criteria. However, ICD-10 does contain some questions that are not culturally sensitive because it is based on premises that are inherent in its construction. Consequently, and in spite of efforts to coordinate ICD-10 and DSM-IV as much as possible, major differences in important areas remain, although the two systems use the same coding system in some cases.

### 7. COLLABORATING CENTERS FOR CLASSIFICATION OF DISEASES AND PROCEDURES FOR ICD-10 UPDATING

To assist users with problems in the development and use of disease classifications and in the use of the ICD, 10 WHO Collaborating Centers for the Classification of Diseases have been created: 3 centers for English-language users (in London, United Kingdom; Canberra, Australia; and Hyattsville, Maryland, United States), 1 center for French- users (Le Vésinet, France), 1 center for Arabic users (Safat, Kuwait), 1 center for Chinese users (Beijing, China), 1 center for Nordic users (Uppsala, Sweden), 1 center for Portuguese users (Sao Paulo, Brazil), 1 center for Russian users (Moscow, Russia), and 1 center for Spanish users (Caracas, Venezuela).

Proposals for changes or additions to ICD-10 should first be presented to the relevant collaborating center. All proposals for changes must be sponsored by, and submitted to WHO through, a collaborating center. Different meetings of WHO Collaborating Centers for Classification of Diseases have approved specific procedures for updating ICD-10.

#### See Also the Following Articles

Cognitive and Behavioral Interventions for Persons with Dementia ■ Dementia in Older Adults ■ Diagnostic and Statistical Manual of Mental Disorders

### Further Reading

- Andrews, G., & Slade, T. (1998). Depression, dysthymia, and substance use disorders: Sources of dissonance between ICD-10 and DSM-IV. *International Journal of Methods in Psychiatric Research*, 7, 116–120.
- Andrews, G., Slade, T., & Peters, L. (1999). Classification in psychiatry: ICD-10 versus DSM-IV. *British Journal of Psychiatry*, 174, 3–5.
- Bertelsen, A. (1999). Reflections on the clinical utility of the ICD-10 and DSM-IV classifications and their diagnostic criteria. *Australian and New Zealand Journal of Psychiatry*, 33, 166–173.
- Cooper, J. E. (1988). The structure and presentation of contemporary psychiatric classifications with special reference to ICD-9 and -10. *British Journal of Psychiatry*, 152(Suppl.), 21–28.
- Haghighat, R. (1997). Phenomenological rigour and consistency of ICD-10. *European Journal of Psychiatry*, 11, 90–93.
- Haghighat, R. (1999). Cultural sensitivity: ICD-10 versus DSM-III-R. *International Journal of Social Psychiatry*, 40, 189–193.
- Rice, J. P., Rochberg, N., Endicott, J., Lavori, P. W., & Miller, C. (1992). Stability of psychiatric diagnoses: An application to the affective disorders. *Archives of General Psychiatry*, 49, 824–830.
- Sartorius, N. (1991). The classification of mental disorders in the tenth revision of the *International Classification of Diseases*. *European Psychiatry*, 6, 315–322.
- World Health Organization. (1992a). *ICD-10 clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: Author.
- World Health Organization. (1992b). *International statistical classification of diseases and related health problems* (10th rev.). Geneva, Switzerland: Author.
- World Health Organization. (1993). *ICD-10 diagnostic criteria for research*. Geneva, Switzerland: World Health Organization.
- World Health Organization. (1996). *WHO ICD-10 diagnostic and management guidelines for mental disorders in primary care (ICD-10, Chapter 5, primary care version)*. Seattle, WA: Hogrefe & Huber.
- World Health Organization. (1997). *The multi-axial presentation of the ICD-10 for use in adult psychiatry*. Cambridge, UK: Cambridge University Press.



# International Conflict

Herbert C. Kelman

Harvard University, Cambridge, Massachusetts, USA

1. International Conflict as a Contemporary Problem
  2. Social–Psychological Propositions about the Nature of International Conflict
  3. Psychological Processes Promoting Conflict
  4. Implications for Conflict Resolution
- Further Reading

## GLOSSARY

**attribution mechanisms** Ways in which to assess the motivation for the behavior of self and others; in conflict relationships, these often lead to confirmation of the negative attitudes toward the adversary.

**coalition across conflict lines** A working relationship between members of opposing parties seeking common ground in the pursuit of their respective interests.

**conflict** A relationship between two parties characterized by the fact and/or perception that each party, in the pursuit of its own interests, threatens or undermines the interests of the other party.

**conflict norms** A set of social expectations within groups or societies in conflict that require members to manifest hostility and distrust toward the adversary in their feelings, language, and actions.

**enemy image** The tendency among parties in conflict to demonize and delegitimize the adversary and to perceive the other party as inherently aggressive, dangerous, and inauthentic.

**groupthink** A tendency within groups tasked with decision making in crisis situations to seek concurrence among their members, discourage dissenting views, and thereby maintain group cohesiveness.

**human needs approach** A view of conflict as generated and maintained by the nonfulfillment and threat to the parties' basic human needs and of conflict resolution as a process that addresses the basic needs of both parties.

**integrative solutions** “Win–win” solutions to a conflict that meet the needs and interests of both parties; these are in contrast to “win–lose” solutions, in which one party prevails at the expense of the other.

**interactive problem solving** An unofficial approach to conflict resolution in which the parties are encouraged to treat the conflict as a shared problem that requires joint efforts to shape mutually satisfactory solutions.

**mirror image formation** The tendency among parties in conflict to develop parallel images of self and the enemy, except with the values reversed, for example, for each party to see the other as inherently aggressive and the self as entirely defensive in response.

**mutual reassurance** A form of influence in conflict relationships designed to demonstrate each party's awareness of the other's concerns and requirements and thereby to reduce the other party's fears of the risks entailed by negotiation.

**negotiation of identity** An informal process whereby parties in conflict explore and invent ways in which to accommodate their group identities and national narratives to one another, particularly by removing negation of the other from their own identities and narratives.

**prospect theory** A theory of decision making under conditions of risk that proposes that people are more prone to take risks to avoid losses than to achieve gains, for example, to go to war in defense of threatened interests than to pursue a potential opportunity for peace.

**realistic empathy** Taking the perspective of the adversary to gain a realistic understanding of the other party's thoughts and feelings.

*responsiveness* An influence strategy based on attempts to elicit cooperation from the other party by addressing the other's needs, fears, and constraints.

*self-fulfilling prophecy* A relationship tendency whereby one party's expectations of the other cause the other party to behave in ways that confirm these expectations.

*zero-sum thinking* A view of the conflict as a relationship where every gain by the other party represents a loss for one's own side and where every loss by the other party represents a gain for one's own side.

The end of the cold war has changed the character of international conflict, but it has not reduced the prevalence of violent conflict, the human suffering that it causes, or the threat that it poses to the well-being and survival of the human species. Thus, during recent years, there has been a proliferation of deadly, deep-rooted conflicts between ethnic and other identity groups within and across nation-states—conflicts often marked by violence against civilians, ethnic cleansing, and genocidal actions. There has also been a rise in terrorism and counterterrorism around the world, undermining the peace and development of established and emerging states.

## 1. INTERNATIONAL CONFLICT AS A CONTEMPORARY PROBLEM

Psychological theory and research can make useful contributions to the understanding and amelioration of the violent manifestations of international conflict. Indeed, psychological concepts and findings have been used increasingly in the study of international conflict and international relations more generally as well as in the development of new approaches to conflict resolution. This article offers a perspective on international conflict that is anchored in social–psychological theory and research and that, in turn, informs the practice of conflict resolution. Social–psychological analysis is designed to complement (and not to replace) approaches based on structural or strategic analysis by providing a special lens for viewing international conflict that brings some of its less explored dimensions into focus.

In this spirit, the article begins with a discussion of several propositions about the nature of international conflict that flow from a social–psychological perspective and that have clear implications for conflict resolution. It then describes social–psychological processes characteristic of conflict interaction that contribute to the escalation and perpetuation of conflict and that must

be reversed if such conflict is to be resolved. [These ideas are presented in greater detail in Kelman's chapter in *Peacemaking in International Conflict* (edited by Zartman and Rasmussen), and linked to a discussion of approaches to conflict resolution in Kelman and Fisher's chapter in the *Oxford Handbook of Political Psychology* (edited by Sears, Huddy, and Jervis).]

## 2. SOCIAL–PSYCHOLOGICAL PROPOSITIONS ABOUT THE NATURE OF INTERNATIONAL CONFLICT

A social–psychological perspective can expand on and enrich the analysis of international conflict provided by the realist or neo-realist schools of international relations or other, more traditional approaches. While acknowledging the importance of objectively anchored national interests, it explores the subjective factors that set constraints on rationality. Without denying the primacy of the state in the international system, a social–psychological perspective opens the “black box” of the state as unitary actor and analyzes processes within and between the societies that underlie state action. Fully recognizing the role of power in international relations, it postulates a broader range of influence processes (and, indeed, of definitions of power) that play a role in international politics. And while affirming the importance of structural factors in determining the course of an international conflict, it conceives of conflict as a dynamic process, shaped by changing realities, interests, and relationships between the conflicting parties.

These observations suggest four general propositions about international conflict that call for social–psychological concepts and data. The four propositions are particularly relevant to existential conflicts between identity groups, that is, conflicts in which the collective identities of the parties are engaged and in which the continued existence of the groups is seen to be at stake. Thus, the propositions apply most directly to ethnic or ideological conflicts, but they also apply to more mundane interstate conflicts insofar as issues of national identity and existence come into play—as they often do.

### 2.1. Conflict as a Process Driven by Collective Needs and Fears

According to the first proposition, international conflict is a process driven by collective needs and fears

rather than entirely a product of rational calculation of objective national interests on the part of political decision makers. Human needs are often articulated and fulfilled through important collectivities such as the ethnic group, the national group, and the state. Conflict arises when a group is faced with nonfulfillment or threat to the fulfillment of basic needs, not only obvious material needs such as food, shelter, physical safety, and physical well-being but also psychological needs such as identity, security, recognition, autonomy, self-esteem, and a sense of justice—needs featured in the 1990 volume *Conflict* and other writings by Burton, a pioneer in the field of conflict resolution. Moreover, needs for identity and security and similarly powerful collective needs, as well as the fears and concerns about survival associated with them, contribute heavily to the escalation and perpetuation of conflict once it has started. Even when the conflicting parties have come to the conclusion that it is in their best interest to put an end to the conflict, they resist going to the negotiating table or making the accommodations necessary for the negotiations to move forward out of fear that they will be propelled into concessions that, in the end, will leave their very existence compromised. The fears that drive existential conflicts lie at the heart of the relationship between the conflicting parties, going beyond the cycle of fears resulting from the dynamics of the security dilemma discussed by Jervis in 1976.

Collective fears and needs, although more pronounced in ethnic conflicts, play a part in all international conflicts. They combine with objective factors—a state's resources, the ethnic composition of its population, its access to the sea (or lack thereof), and the like—in determining how different segments of a society perceive state interests and what ultimately becomes the national interest as defined by the dominant elites. Similarly, all conflicts—interstate as well as ethnic—represent a combination of rational and irrational factors, and in each type of conflict the mix may vary from case to case. Some ethnic conflicts may be preponderantly rational, just as some interstate conflicts may be preponderantly irrational. Furthermore, in all international conflicts, the needs and fears of populations are mobilized and often manipulated by the leadership, with varying degrees of demagoguery and cynicism. Even when manipulated, collective needs and fears represent authentic reactions within the population and become the focus of societal action. They may be linked to individual needs and fears. For example, in highly violent ethnic conflicts, the fear of

annihilation of one's group is often tied to a fear of personal annihilation (and for good reason).

The conception of conflict as a process driven by collective needs and fears implies, first and foremost, that conflict resolution—if it is to lead to a stable peace that both sides consider just and to a new relationship that enhances the welfare and development of the two societies—must address the fundamental needs and deepest fears of the populations. From a normative point of view, such a solution can be viewed as the operationalization of justice within a problem-solving approach to conflict resolution. Another implication of a human needs orientation noted by Burton is that the psychological needs on which it focuses—security, identity, recognition, and the like—are not inherently zero-sum, although they are usually seen as such in deep-rooted conflicts. Thus, it may well be possible to shape an integrative solution that satisfies both sets of needs, which in turn might make it easier to settle issues such as territory and resources through distributive bargaining. Finally, the view of conflict as a process driven by collective needs and fears suggests that conflict resolution must, at some stage, provide for certain processes that take place at the level of individuals and interactions between individuals, for example, taking the other society's perspective (what White called "realistic empathy" in 1984 and elsewhere) as well as creative problem solving, learning, and insight.

## 2.2. Conflict as an Intersocietal Process

Focusing on the needs and fears of the populations in conflict reminds one that international conflict is an intersocietal process, not merely an intergovernmental or interstate phenomenon. The conflict, particularly in the case of protracted ethnic struggles, becomes an inescapable part of daily life for each society and its component elements. Thus, analysis of conflict requires attention, not only to its strategic, military, and diplomatic dimensions but also to its economic, psychological, cultural, and social-structural dimensions. Interactions along these dimensions, both within and between the conflicting societies, shape the political environment in which governments function and define the political constraints under which they operate.

An intersocietal view of conflict points to the role of internal divisions within each society, which often play a major part in exacerbating or even creating conflicts between the societies. They impose constraints on

political leaders pursuing a policy of accommodation in the form of accusations by opposition elements that the leaders are jeopardizing national existence as well as in the form of anxieties and doubts within the general population that are both fostered and exploited by the opposition elements. However, the internal divisions may also provide potential levers for change in the direction of conflict resolution by challenging the monolithic image of the enemy that parties in conflict tend to hold and by enabling them to deal with each other in a more differentiated way. They point to the presence of potential partners for negotiation on the other side and, thus, provide the opportunity for forming what Kelman called pro-negotiation “coalitions across the conflict lines” in 1993. To contribute to conflict resolution, any such coalition must necessarily remain an “uneasy coalition,” lest its members lose their credibility and political effectiveness within their respective communities.

Another implication of an intersocietal view of conflict is that negotiations and third-party efforts should ideally be directed not to mere settlement of the conflict, in the form of a brokered political agreement, but to its resolution. A political agreement may be adequate for terminating relatively specific, containable interstate disputes, but conflicts that engage the collective identities and existential concerns of the adversaries require a process conducive to structural and attitude change, to reconciliation, and to the transformation of the relationship between the two societies. Finally, an intersocietal analysis of conflict suggests a view of diplomacy as a complex mix of official and unofficial efforts with complementary contributions. The peaceful termination or management of conflict requires binding agreements that can be achieved only at the official level, but many different sectors of the two societies must be involved in creating a favorable environment for negotiating and implementing such agreements.

### **2.3. Conflict as a Multifaceted Process of Mutual Influence**

International conflict is best understood as a multifaceted process of mutual influence and not just a contest in the exercise of coercive power. Much of international politics entails a process of mutual influence in which each party seeks to protect and promote its own interests by shaping the behavior of the other party. Conflict occurs when these interests clash, that is, when attainment of one party's interests (and

fulfillment of the needs that underlie them) threatens, or is perceived to threaten, the interests (and needs) of the other party. Therefore, in pursuing the conflict, the parties engage in mutual influence, designed to advance their own positions and to block the adversary. Similarly, in conflict resolution, by negotiation or other means, the parties exercise influence to induce the adversary to come to the table, to make concessions, to accept an agreement that meets their interests and needs, and to live up to that agreement. Third parties also exercise influence in conflict situations by backing one party or the other, by mediating between them, or by maneuvering to protect their own interests.

Influence in international conflict typically relies on a mixture of threats and inducements, with the balance often on the side of force and the threat of force. Thus, the U.S.–Soviet relationship during the cold war was predominantly framed in terms of an elaborate theory of deterrence, that is, a form of influence designed to keep the other side from doing what one does not want it to do, as described in important books by Schelling in 1963, by George and Smoke in 1974, and by Jervis and colleagues in 1985. In other conflict relationships, the emphasis may be on compellence, that is, a form of influence designed to make the other side do what one wants it to do. Such coercive strategies entail serious costs and risks, and their effects may be severely limited. For example, they are likely to be reciprocated by the other side and lead to escalation of the conflict, and they are unlikely to change behavior to which the other side is committed. Thus, the effective exercise of influence in international conflict requires a broadening of the repertoire of influence strategies, at least to the extent of combining “carrots and sticks,” that is, of supplementing the negative incentives that typically dominate international conflict relationships with positive incentives (e.g., economic benefits, international approval, a general reduction in the level of tension). An example of an approach based on the systematic use of positive incentives is Osgood's GRIT (Graduated and Reciprocated Initiatives in Tension reduction) strategy. Egyptian President Anwar Sadat, in his 1977 trip to Jerusalem, undertook a unilateral initiative with the expectation (partly prenegotiated) of Israeli reciprocation, but—unlike GRIT—he started with a large fundamental concession in the anticipation that negotiations would fill in the intervening steps.

Effective use of positive incentives requires more than offering the other party whatever rewards, promises, or confidence-building measures seem to be most readily available. It requires actions that address

the fundamental needs and fears of the other party. Thus, the key to an effective influence strategy based on the exchange of positive incentives is responsiveness to the other party's concerns, that is, actively exploring ways in which each party can help to meet the other's needs and allay its fears as well as ways in which the parties can help each other to overcome the constraints within their respective societies against taking the actions that each wants the other to take. The advantage of a strategy of responsiveness is that it allows each party to exert influence on the other through positive steps (not threats) that are within its own capacity to take. The process is greatly facilitated by communication between the parties to identify actions that are politically feasible for each party and yet are likely to have an impact on the other.

A key element in an influence strategy based on responsiveness is mutual reassurance, which is particularly critical in any effort to resolve an existential conflict. The negotiation literature suggests that parties are often driven to the table by what Zartman in 1989 described as "a mutually hurting stalemate," which makes negotiations more attractive than continuing the conflict. But parties in existential conflicts are afraid of negotiations, even when the status quo has become increasingly painful and they recognize that a negotiated agreement is in their interest. To advance the negotiating process under such circumstances, it is at least as important to reduce the parties' fears as it is to increase their pain.

Mutual reassurance can take the form of acknowledgments, symbolic gestures, or confidence-building measures. To be maximally effective, such steps need to address the other party's central needs and fears as directly as possible. When Sadat spoke to the Israeli Knesset during his dramatic visit to Jerusalem in 1977, he clearly acknowledged Egypt's past hostility toward Israel and, thus, validated Israelis' own experiences. In so doing, he greatly enhanced the credibility of the change in course that he was announcing. At the opening of this visit, Sadat's symbolic gesture of engaging in a round of cordial handshakes with the Israeli officials who had come to greet him broke a long-standing taboo. By signaling the beginning of a new relationship, it had an electrifying effect on the Israeli public. In deep-rooted conflicts, acknowledgment of what was heretofore denied—in the form of recognition of each other's humanity, nationhood, rights, grievances, and interpretation of history—is an important source of reassurance that the other party may indeed be ready to negotiate an agreement that addresses one's own

fundamental concerns. By signaling acceptance of the other side's legitimacy, each party reassures the other that negotiations and concessions no longer constitute mortal threats to its security and national existence. By confirming the other side's narrative, each party reassures the other that a compromise does not represent an abandonment of its identity.

An influence strategy based on responsiveness to each other's needs and fears, and the resulting search for ways in which to reassure and benefit each other, has important advantages from a long-term point of view. It does not merely elicit specific desired behaviors from the other party but can contribute to a creative redefinition of the conflict, joint discovery of mutually satisfactory solutions, and transformation of the relationship between the parties.

#### **2.4. Conflict as an Interactive Process with an Escalatory, Self-Perpetuating Dynamic**

The influence strategies employed in a conflict relationship take on special significance in light of the fourth proposition, which views international conflict as an interactive process and not merely a sequence of action and reaction by stable actors. In intense conflict relationships, the natural course of interaction between the parties tends to reinforce and deepen the conflict rather than reduce and resolve it. The interaction is governed by a set of norms and is guided by a set of images that create an escalatory, self-perpetuating dynamic. This dynamic can be reversed through skillful diplomacy, imaginative leadership, third-party intervention, and institutionalized mechanisms for managing and resolving conflict. But in the absence of such deliberate efforts, the spontaneous interaction between the parties is more likely than not to increase distrust, hostility, and the sense of grievance.

The needs and fears of parties engaged in intense conflict impose perceptual and cognitive constraints on their processing of new information, with the resulting tendency to underestimate the occurrence and the possibility of change. The ability to take the role of the other is severely impaired. Dehumanization of the enemy makes it even more difficult to acknowledge and access the perspective of the other party. The inaccessibility of the other's perspective contributes significantly to some of the psychological barriers to conflict resolution described by Ross and Ward in a 1995 review. The dynamics of conflict interaction tend

to entrench the parties in their own perspectives on history and justice. Conflicting parties display particularly strong tendencies to find evidence that confirms their negative images of each other and to resist evidence that would seem to disconfirm these images. Thus, interaction not only fails to contribute to a revision of the enemy image but actually helps to reinforce and perpetuate it. Interaction guided by mirror images of a demonic enemy and a virtuous self (a concept introduced by Bronfenbrenner in 1961 and elaborated by White in 1965 and elsewhere) creates self-fulfilling prophecies by inducing the parties to engage in the hostile actions they expect from one another.

Self-fulfilling prophecies are also generated by the conflict norms that typically govern the interaction between parties engaged in an intense conflict. Expressions of hostility and distrust toward the enemy are not just spontaneous manifestations of the conflict; they are normatively prescribed behaviors. The assumption of political leaders that their publics' evaluations of them depends on their adherence to these norms influences their tactical and strategic decisions, their approach to negotiations, their public pronouncements, and (ultimately) the ways in which they educate their own publics. For the publics, in turn, adherence to these norms is often taken as an indicator of group loyalty. Thus, the discourse in deep-rooted conflicts is marked by mutual delegitimization and dehumanization. Interaction governed by this set of norms—at both the microlevel and macrolevel—contributes to escalation and perpetuation of the conflict. Parties that systematically treat each other with hostility and distrust are likely to become increasingly hostile and untrustworthy.

The dynamics of conflict interaction create a high probability that opportunities for conflict resolution will be missed. Parties whose interaction is shaped by the norms and images rooted in the history of the conflict are systematically constrained in their capacity to respond to the occurrence and possibility of change. Each party finds it difficult to communicate the changes that have occurred on its own side, to notice the changes that have occurred on the other side, and to explore the possibilities for change that would serve both sides' interests. Therefore, conflict resolution efforts require promotion of a different kind of interaction that is capable of reversing the escalatory and self-perpetuating dynamics of conflict—an interaction conducive to sharing perspectives, differentiating the enemy image, and developing a language of mutual reassurance and a new discourse based on the norms of responsiveness and reciprocity.

### **3. PSYCHOLOGICAL PROCESSES PROMOTING CONFLICT**

Social-psychological analysis can be particularly helpful in explaining why and how, once a conflict has started, normative and perceptual processes that promote its escalation and perpetuation are set into motion and create or intensify barriers to conflict resolution. By the same token, social-psychological analysis, in helping to identify and understand these barriers, can also suggest ways in which to overcome them.

#### **3.1. Normative Processes**

A variety of interaction processes at the mass and elite levels of conflicting societies that influence the evolving course of the conflict are governed by a set of powerful social norms that tend to encourage actions and attitudes conducive to the generation, escalation, and perpetuation of conflict and that tend to inhibit the perception and occurrence of change in the direction of tension reduction and conflict resolution.

##### **3.1.1. Formation of Collective Moods**

With periodic shifts in collective mood, public opinion can act as both a resource and a constraint for political leaders in the foreign policy process. In principle, public opinion can provide support for either aggressive or conciliatory policies, but under the prevailing norms in an intense protracted conflict, leaders are more likely to expect—and mobilize—public support for aggressive policies than for conciliatory policies. Apart from transitory moods, certain pervasive states of consciousness underlie public opinion in a society engulfed in a deep-rooted conflict, reflecting the existential concerns and central national narratives widely shared within the population. In many cases, such as Serbia, Northern Ireland, and the Middle East, historical traumas serve as the points of reference for current events. Although these memories may be manipulated by demagogic leaders, they are part of the people's consciousness and are available for manipulation, as is the associated sense of injustice, abandonment, and vulnerability. The effect of such collective moods is to bring to the fore powerful social norms that support escalatory actions and inhibit moves toward compromise and accommodation. When fundamental concerns about survival and identity are tapped, national leaders, with full expectation of public support, are far

more ready to risk war than to take risks for peace, in line with the proposition derived from prospect theory (as applied to international relations by Levy and others) that people are more reluctant to take risks to achieve gains than to avoid losses. Any change in the established view of the enemy and of the imperatives of national defense comes to be seen as a threat to the nation's very existence.

### **3.1.2. Mobilization of Group Loyalties**

Public support is an essential resource for political leaders engaged in a conflict relationship, both in ensuring the public's readiness to accept the costs that their policies may entail and in enhancing the credibility of their threats and promises to the other side. The primary means of gaining public support is the mobilization of group loyalties. Arousal of nationalist and patriotic sentiments, particularly in a context of national security and survival, is a powerful tool in garnering public support. It may evoke automatic endorsement of the policies that the leadership defines as necessary and a willingness to make sacrifices that cannot be entirely understood in terms of rational calculations of costs and benefits. The nation generates such powerful identifications and loyalties because it brings together two central psychological dispositions: the need for self-protection and the need for self-transcendence.

Group loyalties can potentially be mobilized in support of conciliatory policies. Political leaders may promote painful compromises and concessions to the adversary on the grounds that the security, well-being, integrity, and survival of the nation require such actions. Indeed, leaders with impeccable nationalist credentials, such as Charles de Gaulle, Yitzhak Rabin, and F. W. de Klerk, are often most effective in leading their populations toward peaceful resolutions of conflicts once they have decided that this approach best serves their national interests. In general, however, group loyalties are more readily available to mobilize support for aggressive policies than for conciliatory ones. Proposals for aggressive actions can more easily rely on the vocabulary of nationalism, which characteristically marks off the in-group from the out-group to the detriment of the latter. An appeal to defend the nation against an imminent attack, in particular, is more compelling than an appeal to seize a promising opportunity, as prospect theory might predict. Also, such an appeal elicits a nearly unanimous response among members of the population, whereas an appeal to take advantage of an opportunity

for peace holds no attraction to that segment of the population that equates peace with surrender.

Processes of group loyalty create barriers to change in a conflict relationship. Group loyalty requires adherence to the group's norms, which call for a militant, unyielding, and suspicious attitude toward the enemy in an intense conflict. Militancy and intransigence thus become the measures of loyalty. Hence, particularly in situations of perceived national crisis, the militants exercise disproportionate power and often a veto over official actions and policies. They impose severe constraints on the ability of leaders to explore peaceful options. Dissent from the dominant conflict norms becomes defined as an act of disloyalty and is suppressed, further undermining the exploration of peaceful alternatives.

### **3.1.3. Decision-Making Processes**

The way in which decisions are made in a conflict situation tends to inhibit the search for alternatives and the exploration of new possibilities, particularly when decision makers are operating in an atmosphere of crisis. These tendencies are by no means inevitable, and there are historical instances of creative decision making in dangerous crisis situations, for example, the Cuban missile crisis (as analyzed by Allison in 1971). However, conflict norms do impose serious burdens on the decision-making process.

A major source of reluctance to explore new options is the domestic constraints under which decision makers labor. In an intense conflict situation, adherence to the conflict norms tends to be seen as the safest course of action. Cautious decision makers assume that they are less vulnerable domestically if they stay with the conflict's status quo, adhere to a discourse of hostility and distrust vis-à-vis the other side, and/or threaten escalatory actions than if they take steps toward accommodation and compromise. The search for alternatives in response to changing realities is also inhibited by institutionalized rigidities in the decision-making apparatus. Decision makers and their bureaucracies operate within a framework of assumptions about available choices, effective strategies, and constituency expectations—shaped by the prevailing conflict norms—that may make them unaware of the occurrence and possibility of change. Furthermore, they often rely on established procedures and technologies, which are more likely to be geared toward pursuing the conflict—by military and other means—than toward resolving it.

The microprocesses of action and interaction in crisis decision making further inhibit the exploration of new options. At the level of individual decision makers, the stress they experience in situations of crisis, when consequential decisions must be made under severe time pressures, limits the number of alternatives they consider and impels them to settle quickly on the dominant response. In intense conflicts, this dominant response is likely to be aggressive and escalatory, as shown in books by Holsti in 1972 and Lebow in 1987. At the level of decision-making groups, crisis decision making often leads to what Janis termed “groupthink,” that is, a concurrence-seeking tendency designed to maintain the cohesiveness of the groups. Decision making under these circumstances is much more likely to produce policies and actions that perpetuate and escalate the conflict than to produce innovative ideas for conflict resolution.

### **3.1.4. Negotiation and Bargaining Processes**

The norms governing negotiation and bargaining between parties involved in long-standing conflict strongly encourage zero-sum thinking, which equates the enemy’s loss with one’s own gain. Negotiation, even distributive bargaining in its narrowest form, is possible only when the parties define the situation—at least at some level—as a win-win, mixed-motive game in which they have both competitive and cooperative goals. Each party, while pursuing its own interests, must actively seek out ways in which the adversary can also win and appear to be winning. But this is precisely the kind of effort that is discouraged by the conflict norms.

At the microlevel, negotiators in an intense conflict tend to evaluate their performance by the forcefulness with which they present their own case and by their effectiveness in resisting compromise. To listen to what the other side needs and to help the other side achieve its goals would violate the conflict norms and might subject the negotiators to criticism from their own constituencies, particularly from their hard-line domestic opposition. At the macrolevel, the parties tend to pursue an overall outcome that strengthens their own strategic position and weakens that of the adversary, even when they recognize their common interest in negotiating certain specific issues. Such a strategy reduces the other party’s incentive for concluding an agreement and its ability to mobilize public support for whatever agreement is negotiated. Zero-sum thinking at both levels undermines the negotiating process, causing delays, setbacks, and repeated failures.

### **3.1.5. Structural and Psychological Commitments**

Finally, conflict creates certain commitments that take on a life of their own and contribute to structural changes conducive to the escalation and perpetuation of the conflict, as described in Rubin and colleagues’ *Social Conflict* in 1994. Most obviously, in a conflict of long standing, various individuals, groups, and organizations (e.g., military, political, industrial, scholarly) develop a vested interest in maintaining the conflict as a source of profit, power, status, and/or *raison d’être*. Others, although not benefiting from the conflict as such, may have a strong interest in forestalling a compromise solution because it would not address their particular grievances or fulfill their particular aspirations. Vested interests do not necessarily manifest themselves in deliberate attempts to undermine efforts at conflict resolution. They may take indirect and subtle forms such as interpreting ambiguous realities and choosing between uncertain policy alternatives in ways that favor continuation of the conflict.

Vested interests and similar structural commitments to the conflict are bolstered by psychological commitments. People involved in a long-standing and deep-rooted conflict tend to develop a worldview that is built around the conflict and would be threatened by an end to the conflict. Resistance to change is likely to be more pronounced when the cognitive structure or ideology in which the view of the conflict is embedded is more elaborate, because changing this view would have wider ramifications. In an intense conflict, the image of the enemy is often a particularly important part of people’s worldview, with implications for their national identity, view of their own society, and interpretation of history. This is one reason why images of the enemy are highly resistant to change and contribute to the escalatory and self-perpetuating dynamic of conflict.

## **3.2. Perceptual Processes**

Perceptual and cognitive processes, that is, the ways in which people interpret and organize conflict-related information, play a major role in the escalation and perpetuation of conflict and create barriers to redefining and resolving the conflict despite changing realities and interests. Two perceptual processes that characterize mutual images of parties in conflict can account for this effect: the formation of mirror images and the resistance of images to contradictory information.

### **3.2.1. Mirror Image Formation**

As noted earlier, Bronfenbrenner and White, social psychologists writing about U.S.–Soviet relations, first noted the formation of mirror images as a characteristic of many conflict relationships. Both parties tend to develop parallel images of the self and the other, except with the values reversed. The core content of mirror images is captured by the good–bad dimension. Each side sees itself as virtuous and peaceful, arming only for defensive reasons and prepared to compromise. The enemy, in contrast, is seen as evil and hostile, arming for aggressive reasons and responsive only to the language of force.

A typical corollary of the good–bad images in protracted conflicts is the view that the other party's aggressiveness is inherent in its nature (e.g., ideology, religion, national character, political system), whereas any signs of aggressiveness on one's own part are entirely reactive and defensive. In the language of attribution theory, the enemy's aggression is explained in dispositional terms, whereas one's own aggression is explained in situational terms. Another common corollary of the good–bad image, one that derives from the virtuous self-image, is the assumption on each side that the enemy knows very well that "we" are not threatening them. Each side's own basic decency and peacefulness, and the provocation to which it has been subjected, are so obvious to one's own party that they must also be obvious to the other side. Apart from such generic features of mirror images, which arise from the dynamics of intergroup conflict across the board, mirror images in any given case may reflect the dynamics of the specific conflict. Thus, ethnic conflicts may be characterized by mutual denial of the other side's national identity accompanied by efforts to delegitimize the other's national movement and claim to nationhood, by mutual fear of national and personal annihilation, by a mutual sense of victimization by the other side, and/or by a mutual view of the other side as a source of one's own humiliation and vulnerability.

The mirror image concept implies that certain symmetries in the parties' reactions arise from the very nature of conflict interaction and that they play an important role in escalating the conflict. There is no assumption that all images of the self and the enemy are mirror images, that images on the two sides are equally inaccurate, or that there is empirical symmetry in the two sides' historical experiences and current situation or moral equivalence in their positions. However, the dynamics of the conflict relationship produce a

degree of parallelism in some of the images developed by both participants in that relationship, arising out of the motivational and cognitive contexts in which they operate. At the level of motivation, each side is concerned with "looking good" when blame for the conflict events is being apportioned. Therefore, political leaders feel a strong need to persuade themselves, their own people, the rest of the world, and future historians that the blame rests with the enemy. Cognitively, each side views the conflict from its own perspective and is convinced that it is acting defensively and with the best intentions (painfully aware of its own needs, fears, historical traumas, grievances, suspicions, and political constraints) and that this is so self-evident that it must be equally clear to the enemy.

Mirror images produce a spiraling effect (exemplified by the classical pattern of an arms race) because each side interprets any hostile action by the other as an indication of aggressive intent against which it must defend itself, yet that side's own reactions—the defensive nature of which presumably should be obvious to the enemy—are taken by the other side as signs of aggressive intent. The effect of mirror images is accentuated insofar as the enemy's ideology or national character is perceived to be inherently aggressive and expansionist. In addition to the escalatory effect of mirror images, they tend to make conflicts more intractable because the sharp contrast between the innocent self and the aggressive other makes it difficult to break out of a zero-sum conception of the conflict. However, the concept of mirror images may be a useful tool in conflict resolution. For example, in problem-solving workshops, the parties' discovery that their own actions are perceived differently by the other side than by themselves may open them up to the possibility that the reverse may be true as well. Thus, they may gain access to each other's perspective, insight into the escalatory effects of such two-directional differences in perception, and awareness of the need for mutual reassurance to set a deescalatory process in motion.

### **3.2.2. Resistance to Contradictory Information**

The second feature of conflict images, their high propensity to resisting contradictory information, inhibits the perception of change and the expectation of future change. A great deal of social–psychological theorizing and research has addressed the general phenomenon of the persistence of attitudes and beliefs in the face of new information that, from an outside point of view, challenges their validity but is somehow neutralized or

ignored. Research has focused on several types of mechanisms that account for resistance to contradictory information, including selectivity, consistency, attribution, and the self-fulfilling prophecy. The concepts of selective exposure, selective perception, and selective recall all point to the fact that people's attitudes help to determine the kind of information that is available to them. People are more likely to seek out and be exposed to information that confirms their existing attitudes and to perceive and remember new information in ways that fit into their preexisting cognitive framework. The various models of cognitive consistency (e.g., Heider's theory of cognitive balance, Festinger's theory of cognitive dissonance) suggest that, in the interest of maintaining consistency, people tend to screen out information that is incongruent with their existing beliefs and attitudes. Although inconsistent information may also instigate attitude change, it is more likely to be resisted when the existing attitudes are strongly held and have wide ramifications, as is the case with enemy images. Attribution mechanisms, linked to what Ross called "the fundamental attribution error" in 1977, also promote confirmation of the original enemy image. Hostile actions by the enemy tend to be attributed dispositionally, providing further evidence of the enemy's inherently aggressive, implacable character, whereas conciliatory actions are explained away as reactions to situational forces, requiring no revision of the original image (a phenomenon observed in research by Rosenberg and Wolfsfeld in 1977, by Heradstveit in 1981, and by Rouhana in 1997). Finally, interactions between conflicting parties tend to create self-fulfilling prophecies by causing one's adversaries to behave in line with one's expectations—to take on the roles in which they have been cast by the other side—thereby confirming the parties' original attitudes.

The mechanisms that account for resistance to disconfirming information are particularly powerful in a conflict relationship for several reasons. First, images of the enemy and conflict-related self-images are central aspects of the national consensus; therefore, resistance to disconfirming information is reinforced by strong normative pressures. Second, in a conflict relationship, the opportunities and capacity for taking the perspective of the other side are limited, and this reduces the impact of potentially new information about the varieties, changes, and signs of flexibility in the other side's views. Third, the resistance of enemy images to disconfirmation is magnified by strong beliefs about the unchangeability of the enemy and is reinforced by the view that it is dangerous or even

treasonous to propose that the enemy has changed or will change.

Despite all of the reasons why conflict images are particularly resistant to contradictory information, they are not immutable. Social-psychological evidence suggests that they can change, and historical evidence shows that they do change. The challenge for scholars and practitioners of international conflict resolution is to devise the means to overcome their resistance to change.

#### 4. IMPLICATIONS FOR CONFLICT RESOLUTION

Social-psychological analysis can contribute to international peacemaking by identifying the psychological and social processes that generate and escalate violent conflicts and impede their peaceful resolution as well as by identifying the conditions and procedures required for breaking and reversing the conflict cycle and setting into motion a process of change in the direction of conflict resolution and reconciliation. Social-psychological principles have also informed the development and application of various unofficial microlevel efforts at conflict resolution, designed to complement official diplomacy in a larger multidimensional peace process. Efforts along these lines were pioneered by Burton, who described and conceptualized them in his 1969 book and in numerous subsequent publications. Fisher's *Interactive Conflict Resolution* in 1997 summarized and integrated the range of psychologically based models for intervening in protracted conflicts between identity groups. Kelman's model, interactive problem solving, is described in a 1998 article and a series of other publications, as are the problem-solving workshops through which the interactive problem-solving model is operationalized.

The implications of social-psychological analysis for the macroprocess of conflict resolution, as well as for the microprocess of problem-solving workshops and related activities, can be summarized by returning to the four propositions about the nature of international conflict spelled out earlier in this article.

The view of conflict as a process driven by collective needs and fears serves as a reminder that a conflict cannot be genuinely resolved until these needs and fears of the parties are addressed. Only solutions that satisfy the fundamental needs and allay the deepest fears of each party can serve as the basis of a stable, long-term peace and a cooperative, mutually enhancing

relationship. Unofficial interactive approaches to conflict resolution, such as problem-solving workshops, can contribute to resolving contentious issues in ways that meet the basic needs of one party without raising the fears of the other. They do so by enabling unofficial but politically involved representatives of the two parties to engage in exploratory noncommittal communication in which they can enter into each other's perspective, gain insight into the other's (and indeed their own) main concerns, and jointly shape creative new ideas for mutually satisfactory solutions to the conflict. The insights and ideas generated by this process can then be injected into the political debate and the decision-making process within the two societies.

The view of conflict as an intersocietal process points to the limits of political agreements signed by governments, often under the pressure, or with the mediation, of outside powers or international organizations. At least in the case of a protracted existential conflict between identity groups, a durable peace requires movement beyond settlement of the conflict to its resolution. The agreement must address both parties' basic concerns, must be widely accepted within the two societies, and must reflect and further promote structural and attitude change conducive to a transformation of the relationship between the societies. Another implication of the intersocietal nature of conflict is a view of diplomacy as a complex array of complementary official and unofficial processes. Unofficial interactions make it possible to explore potential openings for mutual accommodation and help to create a political atmosphere that is conducive to negotiation. Furthermore, awareness of the intersocietal nature of conflict helps to counteract the monolithic image that parties in conflict tend to have of each other by encouraging them to attend to what is happening within the other society and to the diversity of tendencies that it encompasses.

Problem-solving workshops and related forums for interactive conflict resolution can be seen as coalitions across conflict lines, enabling pro-negotiation elements on the two sides to find common ground in the pursuit of their respective interests. They engage in a process of direct interaction designed to generate ideas for resolving the conflict that address the needs of both parties and to which both parties can feel committed—essential conditions for agreements that will be widely accepted within the two societies and, thus, conducive to a durable peace. Binding agreements can be reached only at the official level, but unofficial interactions can contribute to enhancing the probability and quality of

official agreements. They can also promote and model the new relationship between the conflicting societies on which a long-term peace must ultimately rest.

The third proposition, positing conflict as a multifaceted process of mutual influence, suggests which strategies and tactics of influence between conflicting parties are most conducive to conflict resolution and to the development of a long-term peaceful relationship. Coercive strategies, which are the dominant mode of influence in conflict relationships, entail serious costs and risks and tend to produce only short-term effects. Therefore, there is a need for a shift of emphasis from the use or threat of force to the use of positive incentives. Moreover, positive inducements are most effective when they are deliberately chosen to be responsive to the needs, fears, and constraints of the other party. A key element of an influence strategy based on responsiveness is mutual reassurance, which can take the form of acknowledgments, symbolic gestures, and/or confidence-building measures. A strategy of responsiveness and reassurance is most effective when it addresses the other party's most basic concerns about identity and security and when it adheres to the principle of reciprocity.

The unofficial, noncommittal setting of problem-solving workshops can be particularly useful in identifying and formulating acknowledgments, gestures, and confidence-building measures that each side can offer to the other. Through a process of joint thinking, the parties can help each other to devise steps of mutual reassurance that would be meaningful to the recipient without entailing unacceptable costs to the actor. This process often involves some informal negotiation, including "negotiation" of national identities and national narratives. Parties cannot be expected to abandon their identities or their national narratives, but each party may be able to adjust its own identity so that it is no longer contingent on denial of the other party's identity and to accommodate its own narrative to the narrative of the other.

Finally, the view of conflict as an interactive process with an escalatory, self-perpetuating dynamic suggests that conflict resolution efforts must be designed to counteract and reverse this conflict dynamic. A conflict relationship generates images and norms that entrench the conflict and create barriers to change that inhibit conflict resolution. Therefore, conflict resolution efforts must be geared to discovering the possibilities for change, identifying the conditions for change, and overcoming the resistances to change. Openness to change and reversal of the conflict dynamic depend on the

establishment of a new discourse among the parties characterized by a shift in emphasis from power politics and threat of coercion to mutual responsiveness, reciprocity, and invitation to a new relationship.

At the microlevel, problem-solving workshops and similar approaches to conflict resolution are designed to contribute to the reversal of the escalatory dynamic of conflict interaction by promoting a different kind of interaction characterized by a deescalatory dynamic. To this end, workshops encourage the parties to penetrate each other's perspective, to differentiate their images of the enemy, to develop a deescalatory language and ideas for mutual reassurance, and to engage in joint problem solving designed to generate ideas for resolving the conflict that are responsive to the fundamental needs and fears of both sides.

### See Also the Following Articles

Decision Making ■ Groupthink ■ Human Rights

### Further Reading

- Allison, G. T. (1971). *Essence of decision: Explaining the Cuban missile crisis*. Boston: Little, Brown.
- Bronfenbrenner, U. (1961). The mirror image in Soviet-American relations: A social psychologist's report. *Journal of Social Issues*, 17(3), 45–56.
- Burton, J. W. (1969). *Conflict and communication: The use of controlled communication in international relations*. London: Macmillan.
- Burton, J. W. (Ed.). (1990). *Conflict: Human needs theory*. New York: St. Martin's.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Fisher, R. J. (1997). *Interactive conflict resolution*. Syracuse, NY: Syracuse University Press.
- George, A. L., & Smoke, R. (1974). *Deterrence in American foreign policy: Theory and practice*. New York: Columbia University Press.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.
- Heradstveit, D. (1981). *The Arab-Israeli conflict: Psychological obstacles to peace* (2nd ed.). Oslo, Norway: Universitetsforlaget.
- Holsti, O. R. (1972). *Crisis, escalation, war*. Montreal, Canada: McGill-Queen's University Press.
- Janis, I. L. (1982). *Groupthink* (2nd ed.). Boston: Houghton Mifflin.
- Jervis, R. (1976). *Perceptions and misperceptions in international politics*. Princeton, NJ: Princeton University Press.
- Jervis, R., Lebow, R. N., & Stein, J. G. (1985). *Psychology and deterrence*. Baltimore, MD: Johns Hopkins University Press.
- Kelman, H. C. (Ed.). (1965). *International behavior: A social-psychological analysis*. New York: Holt, Rinehart & Winston.
- Kelman, H. C. (1991). A behavioral-science perspective on the study of war and peace. In R. Jessor (Ed.), *Perspectives on behavioral science: The Colorado Lectures* (pp. 245–275). Boulder, CO: Westview.
- Kelman, H. C. (1993). Coalitions across conflict lines: The interplay of conflicts within and between the Israeli and Palestinian communities. In S. Worchel, & J. Simpson (Eds.), *Conflict between people and groups* (pp. 236–258). Chicago: Nelson-Hall.
- Kelman, H. C. (1998). Social-psychological contributions to peacemaking and peacebuilding in the Middle East. *Applied Psychology: An International Review*, 47, 5–28.
- Kelman, H. C., & Bloom, A. H. (1973). Assumptive frameworks in international politics. In J. N. Knutson (Ed.), *Handbook of political psychology* (pp. 261–295). San Francisco: Jossey-Bass.
- Kelman, H. C., & Fisher, R. J. (2003). Conflict analysis and resolution. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 315–353). New York: Oxford University Press.
- Lebow, R. N. (1987). *Nuclear crisis management: A dangerous illusion*. Ithaca, NY: Cornell University Press.
- Levy, J. S. (1992). Prospect theory and international relations: Theoretical applications and analytical problems. *Political Psychology*, 13, 283–310.
- Osgood, C. E. (1962). *An alternative to war or surrender*. Urbana: University of Illinois Press.
- Rosenberg, S. W., & Wolfsfeld, G. (1977). International conflict and the problem of attribution. *Journal of Conflict Resolution*, 21, 75–103.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.
- Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 27, pp. 255–304). San Diego, CA: Academic Press.
- Rouhana, N. N. (1997). *Palestinian citizens in an ethnic Jewish state: Identities in conflict*. New Haven, CT: Yale University Press.
- Rubin, J. Z., Pruitt, D. G., & Kim, S. H. (1994). *Social conflict: Escalation, stalemate, and settlement* (2nd ed.). New York: McGraw-Hill.
- Schelling, T. C. (1963). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

- Tetlock, P. E., Husbands, J. L., Jervis, R., Stern, P. C., & Tilly, C. (Eds.). (1989). *Behavior, society, and nuclear war* (Vol. 1). New York: Oxford University Press.
- Tetlock, P. E., Husbands, J. L., Jervis, R., Stern, P. C., & Tilly, C. (Eds.). (1991). *Behavior, society, and nuclear war* (Vol. 2). New York: Oxford University Press.
- White, R. K. (1965). Images in the context of international conflict: Soviet perceptions of the U.S. and the U.S.S.R. In H. C. Kelman (Ed.), *International behavior: A social-psychological analysis* (pp. 238-276). New York: Holt, Rinehart & Winston.
- White, R. K. (1984). *Fearful warriors: A psychological profile of U.S.-Soviet relations*. New York: Free Press.
- Zartman, I. W. (1989). *Ripe for resolution* (2nd ed.). New York: Oxford University Press.
- Zartman, I. W., & Rasmussen, J. L. (Eds.). (1997). *Peacemaking in international conflict: Methods and techniques*. Washington, DC: U.S. Institute of Peace Press.



# Internet Counseling

Azy Barak

University of Haifa, Haifa, Israel

1. Introduction
  2. Historical Development
  3. The Practice of Internet Counseling
  4. Research on Internet Counseling
  5. Ethical, Professional, and Legal Considerations
  6. Limitations and Problems
  7. Summary
- Further Reading

## GLOSSARY

**blog** (*Web log*) A publicly accessible Web page, updated regularly and often, that serves as a personal journal of the site owner.

**chat** Text-based synchronous (real-time) conversation between two Internet users through special software.

**chat room** A virtual environment in which numerous Internet users may “meet” and converse with one another in synchronous communication through typed text.

**cyberspace** The virtual space that people enter and experience when connected to the Internet, following an idea used in Gibson’s *Neuromancer*.

**e-mail** Abbreviated term for electronic mail; a typed letter sent to a virtual mailbox over the Internet.

**e-mail list** (*listserv*) Electronic mailing list where e-mails are sent by subscribers to all other subscribers on the list.

**emoticon** A cluster of punctuation or a small graphical sign or picture that may be used in online communication to compensate for the lack of nonverbal cues to indicate a facial expression or gesture.

**encryption** Coding of text through special software so that only the intended receiver of the message can decipher it (i.e., decryption).

**forum** (*message board or discussion board*) A Web site designed to allow asynchronous group communication.

**instant messaging** Software that allows text-based synchronous communication, with possible time delays (i.e., users control level of synchronicity), between two Internet users.

**post** Publishing a message on the Internet, including sending a message to an e-mail list or publishing a message in a forum.

**synchronous** Coinciding in time of communication that is carried out with all parties present at the same time (e.g., chat).

**thread** A multiple-part virtual “conversation” on a given topic in an e-mail list or a forum composed by a leading (first) message and responses.

**virtual communication** Communication between two or more anonymous individuals through the mediation of computers where the looks, appearance, and nature of these individuals are basically the products of their imaginations.

Internet counseling is a general term referring to a wide spectrum of approaches and techniques that exploit the Internet’s special capabilities in facilitating counseling and psychotherapy online. Growing evidence suggests that despite the lack of in-person contact and the consequent lack of nonverbal cues, professionals do provide effective interventions online. However, in addition to focused training in Internet counseling,

special precautions are necessary, especially regarding unique ethical and legal issues.

## 1. INTRODUCTION

Internet counseling is a generic term that refers to various methods used by psychologists to provide professional services through, or with the help of, the Internet. Common terms that have been used include online (psycho)therapy (or counseling), e-therapy, cybertherapy, and virtual therapy. Although attempts have been made to differentiate among these terms, it has become clear that they refer to similar, if not identical, processes and techniques. Internet counseling involves a variety of online communication technologies that generally fit into  $2 \times 2$  dimensions: synchronous versus asynchronous communication and individual versus group communication. Individual synchronous communication is conducted by means of personal chat or instant messaging, individual asynchronous communication is conducted by e-mail exchanges, group synchronous communication is carried out through a chat room, and group asynchronous communication is managed by a forum (discussion board) or an e-mail list (listserv). Internet counseling may also include other online methods and tools such as Web site-accompanied therapy and online personal journals. Although currently less prevalent, Internet counseling may also use computer sound and camera (Webcam or videoconferencing) facilities.

Internet counseling is usually text based. Consequently, it is characterized by the invisibility of counselors and clients, lack of nonverbal and physical cues, and absence of eye contact—three factors that significantly influence social communication and interpersonal relations. These characteristics, in addition to other factors such as (possible) anonymity and aloneness by the computer, create the “online disinhibition effect” that brings about greater and faster openness and closeness, on the one hand, and less controlled behavioral expressions (often aggression), on the other. Thus, the therapeutic process accelerates more intensively and rapidly in comparison with face-to-face therapy, resulting in faster disclosures and greater exposure of intimate contents. In addition, unique features of Internet use—availability, flexibility of use (in terms of time and place), affordability, anonymity, acceptability, aloneness, interactivity, hypertextuality, updatibility, multimedia capability, easy departure (escape), and richness of contents—have gradually caused Internet counseling to

become a convenient accepted channel through which people seek psychological consultation and help.

## 2. HISTORICAL DEVELOPMENT

Although not documented, it seems that serious professional attempts to harness the Internet for therapeutic purposes were first made during the mid-1990s. At that time, e-mail was used as the primary technology to communicate between practitioners and clients, either individually or in groups. Despite positive experiences reported by therapists, traditional mental health practitioners voiced their opposition to the new method, citing its limitations and possible risks. Five major problems were highlighted: (a) counselors' inability to accurately and correctly perceive and evaluate clients' full emotional states due to the lack of visual and physical cues (e.g., tears, smiles, gestures, body movements); (b) therapists' difficulty in communicating warmth and empathy owing to a similar lack of nonverbal cues; (c) the unclear, intangible, easy-to-break boundaries between therapists and patients; (d) the possibility of erroneously dealing with clients from other countries and cultures without being aware of indigenous customs, rules, norms, or other pertinent information; and (e) the risk of dealing with severe pathological cases, especially suicidal clients, brought on by problems relating to identifying and locating them at critical moments. Criticisms also were made in regard to problems of privacy, confidentiality, and security; the possible undesirable effects of clients' individual disadvantages, such as those related to computer and verbal skills, on the therapeutic process and outcome; discrimination against economically distressed potential clients (what has been termed “the digital divide”); and the potential for impersonation, lying, and other abuses of the system. Several professional associations, in addition to individual professionals, published warnings to prevent (or minimize) online counseling or at least to try restricting its possible dangers to clients.

However, with the help of continuously upgraded hardware and software, as well as growing awareness of and experience, training, and pioneering research in this developing field, many of the criticisms have been answered and gradually been overcome. Moreover, arguments began to be advanced that the opposition reflected mainly resistance to change in general and ignored (or downplayed) a growing body of evidence that supported and pointed to the success of Internet counseling.

Evidently, Internet counseling has now been established and is generally accepted as legitimate professional conduct. During recent years, a growing number of publications—theoretically, practically, and empirically oriented—have emerged in professional outlets as well as in professional conferences. A major portal ([www.metanoia.org](http://www.metanoia.org)) lists many relevant issues relating to psychotherapy through the Internet. Quite a few professional and scientific books have been published on the subject. A major professional association, the International Society for Mental Health Online, has been formed in an attempt to provide a common organization for professionals interested in this emerging area and to promote the understanding, use, and development of mental health online.

### 3. THE PRACTICE OF INTERNET COUNSELING

Internet counseling is diverse, heterogeneous, and multifaceted. The following 10 points describe its characteristics and provide some distinctions from traditional counseling. First, unlike traditional face-to-face therapy, Internet counseling can be practiced through various alternative communication channels. These channels differ from one another in terms of synchronicity (usually synchronous or asynchronous communication, whereas instant messaging allows a changing degree of synchronicity), mode (individual or group communication), means of communication (usually typed text, but pictures and voice communication are also possible), and the degree of on-site human involvement (direct human interaction, self-help by using information and instructions published on a Web site, or online interactive software). Thus, clients have a wide range of choices. Counselors may specialize and offer a specific service modality and provide it through numerous channels. Likewise, clients may receive service through one or more channels of communication.

Second, Internet counseling may be practiced as a complementary, adjacent, or adjunctive process to face-to-face counseling or as an independent process of therapy. As an accompanying procedure, clients may use online communication with counselors in between their face-to-face sessions (according to preestablished agreed-on rules); manage online diaries (i.e., specific behavior monitoring and recording) relevant to the therapeutic process that may be inspected and reviewed by themselves and their counselors at a time and place of choice; deliver materials relevant to the therapy

process such as articles, important links, and necessary or urgently needed documents; and publish and maintain a Web site in which consequences or the offspring thoughts of face-to-face sessions may be expressed or implemented (i.e., a blog). Cyberspace may also be used to search for materials and information found to be relevant in the face-to-face meetings or for exercises and practice activities suggested in face-to-face sessions. In all of these procedures, counselors attempt to integrate offline and online therapeutic experiences at different rates and to different degrees to promote clients' change. Independently, clients and counselors may meet online (usually, but not necessarily, through a designated Web site) and interact virtually through an online channel(s) of communication. In using this procedure for therapy, counselors and clients may be distant from each other geographically, interact in different time zones, and become involved with the therapeutic interaction at a time of choice (if asynchronous communication is used). To overcome the lack of standard, normal, and familiar therapeutic procedures in these cases, such as checking and evaluating therapists' credentials, payment procedures (often done through third-party payment), emergency contacts, and termination of the therapy, a set of instructions and rules is usually established and agreed on between counselors and clients at a preliminary stage. Internet counseling methods, rates of sessions, and lengths of interactions vary among counselors according to their approaches, their availability, the clients' particular problems, communication channels, costs, and other professional considerations.

Third, Internet counseling generally relies on text-based communication between counselors and clients. Writing can often be healing. For many people, venting their personal concerns, expressing their deepest feelings, and sharing their intimate thoughts—all while alone and writing as though talking to themselves, free of therapists' scrutinizing eyes, feeling more secure and protected in their own habitats than in counselors' offices, having the ability to preserve a full record of counseling dialogues, and experiencing less intimidation in developing relationships with virtual (but trustworthy) figures—contribute to creating a unique therapeutic experience, much different from traditional “talking and listening” counseling. The ability of clients to read back, cite, and quote themselves or counselors adds a special value to the counseling dialogue. For many people, sharing their writings—even from the past—with trusted individuals gives them a feeling of relief and sometimes even of elation.

Fourth, Internet counseling may be provided to individuals for either long-term or short-term interventions as well as for single consultations. Although continuous counseling process enables the establishment of relationships between counselors and clients as well as the development of trust and greater exposure, a one-time question-and-answer psychological service is quite common on the Internet. Similar to advice columns in newspapers and magazines, clients present (usually either through sending e-mails or posting messages on online bulletin boards) personal issues and expect counselors to provide advice or solutions. However, there are several differences. First, online consultations are not necessarily public given that clients may discretely send e-mails to counselors and receive private answers from counselors who offer such a service (through Web sites or other means of communication). Second, accumulated questions and answers can be (and actually are) edited and posted on the Internet, classified by problem area, target population, and the like, with the help of an index and an internal search engine for the benefit of many visitors. Third, such an online service is convenient and relatively fast, and it enables people in distress to consult anonymously with experts during times of need for quick assistance.

Fifth, Internet counseling enables consultation and treatment through various theoretical therapeutic approaches, methods, and techniques. For example, psychodynamic, cognitive, and behavioral approaches have been used online, with each one finding the Internet to be a convenient—or even an ideal—platform for interacting with clients according to counselors' own conceptions and principles. For psychodynamic counselors (and even more so for psychoanalytic counselors), cyberspace represents a ready desirable environment for psychotherapeutic encounters because it is characterized by ambiguity. Thus, dynamic processes activated by projection, transference, and object relations usually prevail. In other words, the perceived disadvantages of Internet relationships, such as invisibility and blurred boundaries, become advantageous to counselors who hold to the psychodynamic approach. Therefore, analyzing clients' personality dynamics becomes easier, generally faster, and perhaps more valid than traditional therapy. For cognitive counselors, the straightforward use of clients' verbalization of thoughts and feelings obviously represents a major advantage. Compared with traditional therapy, cognitively oriented counselors are more able to assess and identify thought patterns that may be responsible for negative emotions and consequent problematic behaviors.

Because all online verbal interactions with clients are recorded (when using the common text-based therapy), clients' statements are easy to cite, compare, analyze, and manipulate. In this context, it is important to mention that therapists who apply bibliotherapy and therapists who claim that writing and reading can heal also find online, text-based therapy to be a highly effective therapeutic means. Likewise, users of narrative psychotherapy can exploit written interactions with patients to effectively help them construct and reconstruct personal "stories." Yet behaviorally focused counselors also find Internet counseling to be ideal medium in many senses in that it allows a structured, programmed, straightforward, easy to construct and maintain set of instructions to be communicated to clients, whether through personal contact or through Web sites. Moreover, clients' continuous recording of behavioral data and their delivery to counselors are simple and efficient. Counselors may guide, supervise, and monitor clients' behavior management from a distance and use in vivo simulations of various behavior tasks, accompanied by immediate feedback. Various theoretical approaches, then, can be applied through Internet counseling, as can any eclectic combination desired by counselors. It is important to note, however, that the unique psychological environment created in cyberspace might eventually result in a unique or new adaptation of traditional therapeutic approaches that previously were used solely in the face-to-face physical environment.

Sixth, not only does Internet counseling allow more than a single, standard, and invariable way of communication, its alternative modalities also enable a variety of interactions and activities that may enrich the counseling experience and contribute to client change. In addition to the obvious advantage of Internet counseling in terms of flexibility of location and time, several other features provide added value to the interactions. Asynchronous personal communication through e-mail enables the occurrence of several psychological processes that may improve the counseling experience. For clients, the ability to express feelings and thoughts at times of their choosing—especially at times of personal distress—and not at periodic predetermined meetings with counselors, is a significant advantage. The ability to reread counselors' messages as often as desired, and at any time, is also important. Patients' ability to reflect, associate, consider, recall, and process these interactions in between e-mail exchanges contributes to psychologically desirable incubation and reflection processes. Moreover, the ability to go back and compare notes—the clients' own notes or their counselors'

notes—is very useful during the therapy process. In addition, the simplicity of attaching other documents to e-mail messages, including drawings and links, possibly makes e-mail interaction more effective than mere talk. For counselors, e-mail interaction enables referring to clients at times of their choosing and convenience, providing more opportunities to analyze, compare, recall (i.e., retrieve previous messages), use quotes, and enrich response messages with important relevant links and documents. Likewise, counseling through asynchronous group communication—by using e-mail lists or forums—enables all group participants to benefit from the same advantages. An additional special ingredient that exists in asynchronous online groups is the ease of members' interacting individually with one another through a back channel of communication, a factor that might contribute to groups' dynamics and change so long as there are certain rules that participants must follow. Synchronous individual counseling, executed through a chat or an instant messaging program (e.g., ICQ), enables real-time communication between patients and their therapists from nearly any location. Unlike telephone conversations under similar circumstances, the use of computers allows conversations to be saved for future recall, analysis, and/or quoting. The disinhibition effect of online communication brings about a rapid exposure of significant and essential material for counseling. In many instances, clients open up in minutes about matters that would have taken weeks or even months to reveal in face-to-face interaction with therapists. Moreover, by experiencing less intimidation (and, consequently, being less defensive) and feeling more egalitarian and equal in status, patients have a greater chance of constructively processing and adopting therapists' messages. One of the unique features of synchronous communication that is available to counselors only on the Internet is "multiconversing," that is, conversing privately with several individuals in parallel. Although the technology requires certain technical and cognitive abilities and skills that might be acquired and promoted through training and experience, multiconversing allows a counselor to make use of the time required for one client to read the counselor's message, process it, and type a response to communicate with other clients. Although special concentration and refocusing abilities are obviously required, possible benefits of this activity are clear, especially when providing counseling services to a population through an Internet-based support center. However, synchronous online group therapy (through a chat room) is quite difficult to manage due to the lack of nonverbal

communication and the difficulty in controlling participants' behavior. If practiced, it is important to have a strict, well thought out set of rules and a small number of participants (usually not more than seven).

Seventh, the practice of Internet counseling allows nonhuman interventions; that is, providing interventions through psychoeducational Websites, as well as through interactive software. Advanced sophisticated Web sites employ interactive software that communicates with users in an attempt to assess and provide tailored intervention in specific problem areas. This is done through various alternative methods, including interactive conversations with robotic therapists, adaptive and tailored instructions in dealing with a specific problems, meditating and relaxation suggestions, and interactive decision-making assistance (e.g., about a career direction or school major). More recently, attempts have been made to adapt virtual reality software in the field to the Internet, thereby allowing mass use and continuous central upgrades. Therefore, the ability of advanced (and continuously developing) software to deliver professional information and interventions in an efficient, attractive, rich, and friendly way makes this a prevalent and pervasive type of communication with clients. Regardless of the areas of distress or problems, whether quitting smoking, obtaining sexuality education, experiencing posttraumatic stress disorder following a sexual assault, facing the death of a loved one, choosing a major in school, or going through a divorce, a well-designed, thoughtful, updated, interactive, sophisticatedly linked (to other Web resources), and stimulating Web site can provide effective psychological interventions for people in need. As in any process of self-help, users' commitment, on the one hand, and counselors' monitoring, on the other, may be problematic. However, the cost-benefit advantage, in addition to factors mentioned earlier (e.g., anonymity), has turned this into a common and successful approach.

Eighth, free Internet counseling, by means of information, advice, and support, has flourished. Despite the clear problem of potential abuse of the system by nonprofessionals or even criminals (e.g., practicing without a license), the use of the Internet has enabled people in need to reach out for psychological help when, for whatever reasons, they could not receive it in standard ways. People in remote locations, the physically disabled, the elderly, the sick, those with financial limitations, and those who are too shy, introverted, or afraid of exposure (e.g., celebrities and their families) all fall into this category. Online support services for a variety of areas of distress are operated by governments,

organizations (e.g., schools), professional associations, and private people for free. Usually, e-mail consultation is offered to individual referrals. One of the main activities of such free counseling services, however, is the provision of online support groups, in which participation is usually open. This type of Internet counseling, with or without the active participation of trained counselors, is widespread and considered to be successful.

Ninth, Internet counseling is effectively and conveniently used as a vehicle for communicating with clients before beginning and after terminating face-to-face counseling. Information provided through Web sites regarding counseling approaches, counseling methods, the counseling process, specific counselors, costs, insurance coverage, and technicalities equip clients with important and relevant knowledge before starting counseling. Questions and unclear matters are easily dealt with through e-mail. This preparatory stage not only saves time and effort but also helps to build realistic constructive expectations about therapy. After termination, e-mail is conveniently used to contact former clients at any point in time, either for professional follow-up or for research purposes (e.g., questionnaires may be attached). On termination, Internet counseling clients are often urged to make e-mail contact with their counselors when needed. This step provides not only a convenient way in which to communicate but also reassurance for patients departing from close, and sometimes dependent, therapeutic relationships. It should be noted, however, that this option might hinder clients from working through successful terminations.

Finally, Internet counseling may include and coincide with online psychological assessment procedures. Online psychological tests, usually in a multiple-choice format, are very common on the Internet, although they do not necessarily produce valid assessments due to the nonprofessional construction and publication of these tests. However, much effort has been made to develop online tests and questionnaires that meet rigorous psychometric standards and make valid assessments. In addition, other assessment procedures, such as online interviewing through videoconferencing, have been implemented. Counselors who operate on the Internet may use professionally established online diagnostic tests that integrate with the counseling process, thereby avoiding cumbersome paper-and-pencil testing of clients with whom they interact online. Several methods have been developed to ensure test quality and to verify clients' identities if necessary. It should be kept in mind, however, that whereas

online tests produce test-based assessment, counselors miss the opportunity to diagnose clients in relation to their test-performing behaviors (e.g., nervousness).

These 10 principles characterize Internet counseling and make it unique and different from any previously practiced psychological interventions. These principles represent a significant revolution in the traditional conception of psychotherapy. Perhaps this is the reason why many have voiced concern about its implementation and growing acceptance by practitioners.

#### **4. RESEARCH ON INTERNET COUNSELING**

As a relatively new area of conduct, Internet counseling has been subjected to empirical study in only a limited scope. Yet published research reports, or those presented at public conferences, can shed some light on intriguing questions regarding therapeutic process and outcome as well as on other important subjects (e.g., training of counselors). It is important to note that the methods used for research on Internet behavior and Internet-related interventions are somewhat different from those used for research on offline counseling. In Internet counseling, individuals often are anonymous and unidentifiable, people's behaviors often parallel their written scripts (and not their observed behaviors), technical difficulties might interfere with the durability of data, imposters and fakers among participants are possible, dropout rates are higher than usual, and so forth. Research has been conducted on various questions that are important for Internet counseling, including some issues relating to the process and outcome of a variety of online counseling interventions, the impact of online support groups, and the quality of measurement of online testing. In addition to research, quite a few case studies have been presented in publications and at conferences.

An overview of the empirical research on Internet counseling provides support of an emerging approach. Process-related research has found that clients of online therapy actually build therapeutic relationships, communicate and express themselves effectively, experience therapeutic alliances, and find relationships with therapists to be satisfactory. Outcome research in numerous areas of distress, such as depression, several types of anxiety, recurrent headaches, eating disorders, pathological grief, panic disorder, and general concerns, often provides positive results, usually in comparing effects with those of parallel offline treatment. Most of this

research has found that Internet counseling leads to results that are similar to, and sometimes even better than, those of traditional counseling. Feedback from Internet counseling users frequently highlights their appreciation of flexibility of use in terms of place and time, the value of experiencing sensitive (and sometimes threatening) relationships and avoiding physical proximity and eye contact, the importance of written records of conversations, and the reassuring feeling due to the options of anonymity and easy termination.

More research has been published on online support groups than on online psychotherapy. Studies have reported that online group dynamics are generally similar to those of offline groups in terms of therapeutic factors and group processes, although the weight and impact of these factors may vary. Outcome research on online support groups has consistently found a great impact on participants in terms of emotional relief and symptom reduction. Users of online support groups repeatedly praise this activity and highlight the ability to find a place where other people go through difficulties similar to their own and to share feelings and thoughts with them, receive and provide advice, and experience nonjudgmental acceptance.

Research on Internet-based assessment, especially online testing, has generally been very supportive. In most published research, it has been found that when online diagnostic tests are professionally constructed and managed, they offer psychometric qualities similar to those of their offline equivalents. Moreover, most test takers find this activity to be convenient and satisfactory. However, research does show that test parameters may change significantly from offline to online versions. Thus, a new standardization is suggested, as is caution in interpreting raw scores derived from the Internet.

In contrast to those findings, research on self-help Web sites has consistently found the validity of information provided to users to be questionable. Problems have been found regarding wrong or even misleading information, superficial and nonprofessional information, obsolete information and the slow pace of updates, and biased and selective information (in many cases reflecting Web site owners' agendas). Not only may these problems result in not helping users' behavioral and/or emotional states, they might even cause more harm in extreme situations.

Very little research has been published on software-managed Internet counseling. Although experimental data from laboratory studies seem to be promising, there is still much to be investigated in terms of actual effectiveness and the impact on real online clients.

## 5. ETHICAL, PROFESSIONAL, AND LEGAL CONSIDERATIONS

The special challenges introduced by online provision of psychological applications, especially treatment interventions, have generated debate, especially on ethical issues. It seems that two significant differences from the traditional counseling encounter—the invisibility of therapists and patients, on the one hand, and the communication from a distance, on the other—have triggered much criticism, at times even leading to a point of total debarment. This outcome comes on top of the reliance on new and developing technologies and unstable communication technology, many people's uncertainty and anxiety over technology in general, the lack of computer knowledge and skills, the discriminating digital gap, and the fluidity of professional boundaries, to mention just a few factors. All of these have created the need for thorough ethical and professional considerations and subsequent remedies. Several professional organizations (e.g., American Psychological Association, Clinical Social Work Federation) have published statements emphasizing the concern and need for special caution. Other professional organizations (e.g., American Board of Certified Counselors, American Counseling Association, International Society for Mental Health On-line) have endorsed special ethical codes regarding the provision of Internet counseling.

The ethical standards according to which professionals are advised to practice Internet counseling correspond to, and coincide with, general ethical guidelines for counseling practice. These standards also complement their special online environment and connection with clients. Several major principles form the center of the common ethical codes. First, because online communication is essentially different from regular face-to-face communication, clients ought to be oriented about online limitations in communicating (e.g., lack of nonverbal cues, technological failures) and how to deal with them. Second, counselors must reveal some of their personal details (e.g., full names) and share their professional qualifications with clients. Third, counselors are expected to explain to clients the potential benefits as well as the potential risks of Internet counseling. Fourth, counselors are advised to create and come to an agreement with clients on rules according to which counseling will take place (e.g., frequency of and reaction time to e-mails). Fifth, counselors ought to practice special safeguards in protecting the secrecy of clients' records and correspondence (e.g., use encrypted e-mail, chat through a secure site,

protect computer records by means of a password). Sixth, counselors should discuss emergency situations with clients, including alternative channels of communication and local backup service.

There are quite a few other ethics-related, professional, and legal issues and considerations of which counselors must be aware. The easy access to and reception of Internet counseling by minors necessitates special attention to clients' ages and acquiring guardians' legal consent (where the law requires it) in cases where minors are referred to counseling. Obviously, this is not an easy task to deal with given that many young referrals want to maintain secrecy—often from their parents—and preserve anonymity. The ability to provide counseling services to clients from other states, provinces, and countries introduces complicated legal problems that reflect issues of jurisdiction, legal accountability, and the like, leading to possible lawsuits. In addition, the ability and ease of being referred to counselors from other cultures—a very common phenomenon—necessitates a special awareness of language limitations and intercultural differences that might significantly intervene in the counseling process and outcome. Standard malpractice insurance might not cover online professional practice; hence, counselors must take special cautionary steps.

Another issue deals with providing counseling to anonymous clients. As indicated previously, this factor has a significant impact on the special value of Internet counseling; however, it also raises concerns of a possible legal (e.g., parental consent for minors, duty to report of crime) and professional (e.g., emergency action at a time of potential suicide or violence) nature. It should be noted here that special attention has been given to suicidal referrals and to people who are in emergency situations. Perhaps related to the harboring of traditional conceptions and habits—assuming that physical proximity and face-to-face visibility have special value in helping people and in saving lives in extreme situations—is the fact that many Internet counselors prefer not to accept clients in such situations for therapy and instead prefer to refer them to face-to-face services or to telephone hotlines. It should be noted, however, that the support provided through online channels to suicidal people and to people who were extremely disturbed emotionally has proven to be as successful as that provided through offline traditional services. Owing to anonymity, ease of termination, and the lack of eye contact, it seems that people in these situations may be drawn to online counseling more easily than to offline counseling, thereby giving them a better chance of being helped.

To deal with the identity and authenticity of online counselors, a system of credentials checking has been suggested. Counselors who desire this service would supply full details and documents relating to their identities and qualifications (i.e., professional training and experience). These would be checked and verified with the authorizing bodies and institutions. When confirmed, counselors would receive special icons to be posted on their Web sites. Concomitantly, the system would educate potential users to apply only to counselors whose sites carry the permit. Although this idea might solve a major problem that currently characterizes Internet counseling, its application is still in the developmental stage.

Payment for Internet counseling is another professional issue. Although it was originally thought that fees for online counseling would be lower than those for face-to-face counseling, the fees (where counseling is not provided free of charge) are actually similar to those for in-person counseling. It seems that the special expenses for technology and continuous upgrades, the cost of required security measures, and the extensive time needed for counselors to type (as opposed to talk) made the preliminary expectations concerning reduced fees premature.

## 6. LIMITATIONS AND PROBLEMS

Internet counseling is not flawless or simple to use and is vulnerable to quite a few problems. The heavy dependence on technology, including electricity, the numerous components of personal computers, and complicated Internet communication, makes the whole system vulnerable and fragile. Working with sensitive clients in shaky situations under these circumstances imparts an uncertainty, or even risk, to Internet counseling. Internet counseling also requires that both counselors and clients be skilled in the use of computers and the Internet because a lack of skills or limited skills directly affects the counseling process. Unskilled counselors might, for example, lose contact with clients in the midst of highly sensitive sessions without preparing the clients for such incidents, where the consequences might be detrimental. Moreover, this type of counseling, suits those who find Internet communication to be easy and convenient, who prefer reading and writing over listening and speaking, and who do not find computer technology and Internet surfing to be intimidating. Still another problem has to do with the “digital divide,” that is, discrimination

against “economically challenged” individuals. Although free Internet services are available in some public places, this certainly cannot be considered to be a solution to a problem that prevents many from receiving counseling.

The existence of sophisticated hackers and computer-savvy individuals who might infiltrate private computer zones and copy, alter, or eavesdrop on confidential materials represents another severe problem. Although developing legislation worldwide and changing the policy and priorities of the police regarding hackers might seem to lessen this problem in the future, and although offices and/or file cabinets certainly are not impenetrable themselves, Internet counseling is still considered to be less secure than face-to-face counseling. However, the growing use of encrypted online correspondence reduces the severity of this problem. Finally, a major problem has to do with untrained professionals who practice online. The practice of Internet counseling requires special incremental training, on top of standard counseling education, that focuses on the specific knowledge needed to practice counseling online. Professionals who offer Internet counseling without appropriate training might not provide optimal counseling to clients, at best, and might cause harm to them, at worst.

## 7. SUMMARY

Internet counseling harnesses advanced computer and communication technology so that professionals can provide consultation and therapeutic services to people who can and want to use the Internet for these purposes. Despite the criticisms voiced by individuals and organizations alike regarding problems and disadvantages of the new approach compared with traditional face-to-face counseling, this movement has been growing in size and scope. With proper awareness of the risks and limitations, appropriate professional training, strict and well-communicated ethical codes, public education, and perhaps legal actions, Internet counseling might prove not only to be useful and efficient but also to exceed traditional counseling in many ways. The possibility of reaching remote populations that do not receive counseling services regularly and of providing convenient access to disadvantaged, shy, and disabled people as well as people who want to avoid exposure, the importance of automatically stored records for reflection and further psychological analysis, and the great opportunities for education, supervision, and research all make

Internet counseling unique and useful. In the current trend, it seems that professional associations are cautiously and gradually endorsing Internet counseling as a legitimate, but not problem-free, method of providing counseling.

## See Also the Following Articles

Cyberpsychology

## Further Reading

- Barak, A. (1999). Psychological applications on the Internet: A discipline on the threshold of a new millennium. *Applied and Preventive Psychology*, 8, 231–246.
- Barak, A., & English, N. (2002). Prospects and limitations of psychological testing on the Internet. *Journal of Technology in Human Services*, 19(2/3), 65–89.
- Bloom, J. W., & Walz, R. G. (Eds.). (2000). *Cybercounseling and cyberlearning: Strategies and resources for the new millennium*. Alexandria, VA: American Counseling Association.
- Boer, P. M. (2001). *Career counseling over the Internet*. Mahwah, NJ: Lawrence Erlbaum.
- Fenichel, M., Suler, J., Barak, A., Zelvin, E., Jones, G., Munro, K., Meunier, V., & Walker-Schmucker, W. (2002). Myths and realities of online clinical work. *CyberPsychology & Behavior*, 5, 481–497.
- Fink, J. (Ed.). (1999). *How to use computers and cyberspace in the clinical practice of psychotherapy*. Northvale, NJ: Jason Aronson.
- Gackenbach, J. (Ed.). (1998). *Psychology and the Internet*. San Diego: Academic Press.
- Goss, S., & Anthony, K. (2003) (Eds.). *Technology in counseling and psychotherapy: A practitioner's guide*. London: Palgrave Macmillan.
- Hsiung, R. C. (Ed.). (2002). *E-therapy: Case studies, guiding principles, and the clinical potential of the Internet*. New York: Norton.
- Joinson, A. (2002). *Understanding the psychology of Internet behaviour*. Basingstoke, UK: Palgrave Macmillan.
- Kraus, R., Zack, J., & Stricker, G. (Eds.). (2003). *Online counseling: A handbook for mental health professionals*. San Diego: Academic Press.
- Ritterband, L. M., Gonder-Frederick, L. A., Cox, D. J., Clifton, A. D., West, R. W., & Borowitz, S. M. (2003). Internet interventions: In review, in use, and into the future. *Professional Psychology: Research and Practice*, 34, 527–534.
- Riva, G., & Galimberti, C. (Eds.). (2001). *Towards cyberpsychology: Mind, cognition, and society in the Internet age*. Amsterdam: IOS Press.
- Rochlen, A. B., Zack, J. S., & Speyer, C. (2004). Online therapy: Review of relevant definitions, debates, and

- current empirical support. *Journal of Clinical Psychology*, 60, 269–283.
- Stofle, G. S. (2001). *Choosing an online therapist: A step-by-step guide to finding professional help on the Web*. Harrisburg, PA: White Hat Communications.
- Suler, J. (1996–2004). The psychology of cyberspace. Retrieved on April 20, 2004, from <http://www.rider.edu/~suler/psycyber/psycyber.html>.
- Suler, J., Barak, A., Chechele, P., Fenichel, M., Hsiung, R., Maguire, J., Meunier, V., Stofle, G., Tucker-Ladd, C., Vardell, M., & Walker-Schmucker, W. (2001). Assessing a person's suitability for online therapy. *CyberPsychology & Behavior*, 4, 675–679. (see also correction, *CyberPsychology & Behavior*, 5, 93).
- Wallace, P. (1999). *The psychology of the Internet*. New York: Cambridge University Press.



# Interpersonal Attraction

Carlos Yela

Complutense University, Madrid, Spain

1. Introduction
2. What Is IA and What Generates It?
3. Causes and Explanations for IA: Theories of Social Psychology and Levels of Analysis
4. Applications of the Psychology of IA  
Further Reading

because of their association with a releaser or unconditioned stimulus).

**socialization** A lifelong process through which every person learns cultural, social, and group norms, both explicit and implicit, through the influence of parents, siblings, peers, teachers, the media, and social organizations.

**socially desirable characteristics** Characteristics of a subject that are perceived as positive for the majority of a society.

## GLOSSARY

**arousal** Physiological activation level of a person at a given moment, which varies according to many situational and psychological variables.

**complimentary of needs** An interpersonal state in which one's needs tend to be satisfied by another's behavior, and one's behavior tends to satisfy the other's needs.

**personally desirable characteristics** Characteristics of a subject that are perceived as positive by another, regardless of whether they are socially desirable or not.

**pheromones** Biochemical hormone-like substances that seem to be involved in eliciting sexual attraction and perhaps also interpersonal attraction (and a promising and intriguing future area of research).

**propinquity** Physical proximity between two (or more) persons, which increases the possibility of interaction and therefore familiarity, with a tendency to increase interpersonal attraction.

**releasers** In sociobiological language, stimuli that tend to elicit an automatic, unconditioned attraction response (in contrast to conditioned stimuli, which raise attraction

Interpersonal attraction (IA) plays a major role in human lives, improving both personal well-being and interpersonal relationships, or on the contrary, resulting in adaptive, interpersonal, and psychological problems if one fails to attract others or to find others attractive. The main factors that provoke IA are socially and personally desirable characteristics (including physical attractiveness), similarity, reciprocity of attraction, propinquity, familiarity, high arousal level, complimentary of needs and, presumably, pheromones. People feel attracted to both unconditioned and conditioned stimuli, thus not only are biological processes relevant, but also sociocultural, interpersonal and psychological factors all play a role. Sociopsychological knowledge can be applied to try to improve IA so as to construct better interpersonal relationships and improve quality of life. This can be done in a variety of fields and contexts, such as within organizations, sports teams, school, the family, the couple, therapy, politics, and international and intercultural relationships.

## 1. INTRODUCTION

All human beings are heavily influenced by interpersonal attraction (IA) in their everyday lives, and this has been the case throughout the history of humankind. In this threatening and problematic world, IA may be one of the keys to harmonious relationships, and its absence leaves the door open to negative feelings and negative interpersonal behavior patterns. Clearly, it would be blithe to suggest that social problems can be solved through increased IA; rather, attention should be focused on the major role IA can play in improving personal well-being and interpersonal and social relationships.

Although it contrasts in important ways, IA is firmly linked to other basic feelings and behavior patterns essential for personal well-being and everyday social interaction, such as the need for affiliation (and group processes), fondness, friendship, and love itself. Failure to attract others or to find others attractive often results in adaptive, interpersonal, and psychological problems.

## 2. WHAT IS IA AND WHAT GENERATES IT?

IA is usually described as a positive attitude toward someone, or an attitude or evaluation of other people in terms of how much one likes or dislikes them. Being an attitude, IA has an affective, a cognitive, and a motivational behavior component, and is mainly learned via the socialization process and in everyday interaction. Like all other attitudes, IA influences (and is influenced by) people's behavior, cognition, and feelings.

In summary, IA is based on the need for affiliation of he or she who is attracted, and its origins can be found in a feature or features of the attractor as they are perceived by the attracted. This article does not provide an exhaustive list of different factors and processes that psychological research has shown to generate IA, nor an attempt to identify those that are strictly necessary, but rather a general overview of such complex phenomena is presented, based on the main findings of empirical psychological research, whose general conclusions, despite exceptions, lead to a better overall understanding of IA. The characteristics that social psychological research has shown to generate IA, then, are the following, as shown in [Table I](#).

**TABLE I**  
Main Factors That Generate IA from A to B

• Socially desirable characteristics of B (perceived by A)		
• Physical attractiveness	• Social skills	• Intelligence
• Competence	• Affection	• Kindness
• Sense of humor	• Generosity	
• Personal characteristics of B (perceived by A) specially valuable for A		
• Similarity between A and B (perceived by A)		
• Values	• Attitudes	• Interests
• Customs	• Opinions	• Likes (and dislikes)
• Reciprocity of attraction between A and B (perceived by A)		
• Propinquity between A and B		
• Familiarity between A and B		
• Complementarity of needs between A and B (e.g., talkative and a good listener, protective and vulnerable, etc.)		
• High arousal level of A		
• Pheromones of B		

1. Socially desirable characteristics of the subject of attraction, such as social skills, intelligence, competence, affection, kindness, a sense of humor, and generosity. It is important not to underestimate the influence of physical attractiveness, which has been shown to have stronger links with IA and with other positive attributions than most would like to acknowledge. Whether this is due to the halo effect, to the learned association of "what is beautiful is good," or to both, the fact of the matter is that although it may seem unjust, there are clear indications of this effect.

2. Personally desirable characteristics of the subject of attraction: personally desirable for the attracted, meaning that some features not universally considered as positive are specially relevant for the attracted. Examples of such traits include self-sufficiency, modesty, seriousness, being easygoing, eroticism, femininity, masculinity, leadership, sociocultural status or prestige, knowledge, or even a particular hobby or interest, a reminder that "beauty is in the eyes of the beholder." This is one of the more interesting questions regarding IA: what leads a person to feel attracted toward one kind of person and not another? Where does attraction to specific traits come from? In this section, this question is addressed, calling on the biological, cultural, social, interpersonal,

and finally personal (psychological) forces that exert their influence on the individual's feelings of attraction toward some particular trait or person. In simple terms, the socialization process, life experiences, history of reinforcements, and everyday interaction that contribute to one's personality all offer some of the most coherent answers to these often-asked questions.

3. Similarity between attractor and attracted, not only in core values, but also in many other areas: attitudes, interests, customs, opinions, likes and dislikes, etc. This is one of the most firmly established and empirically supported findings in social psychology, although it is not always accepted by the general public. And whereas similarity generates attraction, dissimilarity tends to generate rejection.

4. Reciprocity of attraction: it is also well known that being liked by another is a good predictor for being attracted toward him or her, because people certainly like to be liked (just as being rejected is a good predictor for rejecting someone). The strength of this effect is such that the mere perception of reciprocity is sufficient to trigger IA, regardless of whether this perceived reciprocity is real.

There are some other factors that have been related to IA, such as propinquity (although the curious phenomenon of IA across the Internet is becoming increasingly pertinent); familiarity, in accordance with the "mere exposure" effect; a high arousal level, such as the one evoked by a new, risky, uncertain, or ambiguous situation, or a situation in which our senses are overloaded; or complementarity of needs, something that leads people to the often mistaken conclusion that "opposites attract." Alongside these other factors, some authors have outlined the role of biochemical substances named pheromones, which seem to be related to sexual attraction and probably also to IA. Thus, popular references to the "chemistry" of a relationship may have some real foundations. And there is one last factor that may be just as crucial to IA as it is in the rest of human lives: mere chance (though perhaps somewhat difficult for some scientists to accept). To be in the same place at the same time, with the same aims and attitudes, and to be introduced, or simply informed of the other's existence, clearly introduces a note of randomness to the incidence of attraction. IA, therefore, is one of the bases for friendship and love, but these other interpersonal relationships imply other factors outside the scope of this section.

As with any interpersonal process, IA is not a static but a dynamic phenomenon. It begins at a certain moment, develops, and can be maintained (through

intermittent reinforcement, by establishing a good balance between costs and rewards, or through cognitive processes like attribution or dissonance, for example) or diminished (through the learning laws of habituation and satiation, or a bad costs-rewards balance). It is just as important to focus, then, on factors that tend to maintain or diminish IA as well as those that generate it.

### 3. CAUSES AND EXPLANATIONS FOR IA: THEORIES OF SOCIAL PSYCHOLOGY AND LEVELS OF ANALYSIS

There is already an overwhelming body of literature on IA, ranging from classic studies to the most recent work. In the last few decades, there has been a constant stream of research in the field, and hundreds of papers have been published. Some research is serious and systematic; some, quite frankly, is not. With reference to the former, many theories have been offered regarding IA, to explain why humans feel particularly attracted to some individuals. Some of those theories propose explanations in terms of reinforcement, others focus on cognitive processes, others still on the concepts of social exchange. Taking the findings as a whole, it could be said that an individual will probably feel IA for those who match one or more of the following conditions: those who reinforce his or her behavior, those who produce reinforcement expectations in the individual, those who show similar attitudes to him or her, those who satisfy his or her needs, those who decrease everyday worries and negative feelings, those who provide more benefits than costs to the individual, those who provide him or her approximately the same benefits as costs, those who provide just the type of resources the individual desires, those who the individual chooses of his or her own "free will," those who are attracted by the individual, those who the individual feels he or she should not feel IA for, those whose behavior is reinforced by him or her, those who reinforce the individual "freely," and/or those who give him or her positive information.

Each one of these different conditions is upheld by one or more of the many different theories of social psychology that have been applied to explain IA. Many of them are classic, well-known theories, such as theories of reinforcement (and its application to IA, such as Byrne's famous attraction paradigm), of balance, of cognitive dissonance, of social comparison, of reactance, of self-perception, of social exchange, of

**TABLE II**  
**Factors from Different Levels of Analysis That Influence IA from A to B**

<i>Level of analysis</i>	<i>Factors influencing IA</i>
Species	Biological factors, pheromones, releasers
Culture	Esthetic criteria, cultural trends, culturally valued traits
Society	Social norms of IA, social influences, social pressures, socialization agents, and organizations
Interpersonal	Everyday interaction, group processes (conformity, modeling, etc.), propinquity, familiarity, reciprocity, complementarity
Individual	Personality (specific characteristics, values, attitudes, attraction to certain specific characteristics), physical attractiveness, needs, arousal level (generated by interaction with the situation)

gain–loss, of interdependence, and of equity. However, the process of IA, as with many other interpersonal processes, is deeply influenced by a wide range of factors, which can be understood more easily if some different levels of analysis are considered, as shown in Table II.

### 3.1. Species

As a member of the human species, we are often physically attracted to physical stimuli called “releasers,” which tend to elicit an automatic response of attraction. Thus, sociobiological researchers have found that as a general rule, despite usually quite significant cross-cultural differences, men tend to feel more attracted by women with neotonous (child-like) facial characteristics, large and firm breasts, narrow waists, fine eyebrows, moderately wide hips, a slim and curvaceous figure (commonly reinforced by high-heeled shoes in some cultures), long and slender legs (exaggerated by the use of short skirts), a slightly protruding backside, and a lack of body hair. According to the same kind of studies, women, meanwhile, tend to feel attracted by a strong jaw, rough skin, bushy eyebrows, a firm and compact backside, a robust frame, tall stature, large eyes, and a flattish stomach. Furthermore, the role of pheromones,

mentioned previously, is another biological factor related to IA. Of course, IA is not the same as physical attraction, but many social psychology researchers have demonstrated a very strong link between the two phenomena, such that people often feel IA for those who seem physically attractive to them, just as they tend to perceive those whom they feel IA for as more physically attractive than they supposedly are. Moreover, as was previously mentioned, neither IA nor physical attraction are evoked only by unconditional (automatic, instinctive) stimuli, but rather by conditioned (learned) ones, which each individual has learned to associate with that response throughout socialization processes. This is exemplified, for example, by changes in the concept of “beauty” through the ages.

### 3.2. Culture

IA is also strongly influenced by aesthetic criteria and cultural trends and values, which vary from one culture to another, which is to say that they also vary from one historical moment to another. In fact, one of the main factors that often precedes IA, the presence of socially desirable characteristics in the attractor, is something that depends mainly on the norms of each culture. One culture (at a given moment in time) may value the capacity to defend oneself in the face of adversity, while another may value the capacity to ask for help and community support, to give one simple example. Cross-cultural social psychology has shown us, for example, the variability between individualist versus collectivist cultures in terms of the norms and rules that prevails in interpersonal relationships.

### 3.3. Society

Social norms, continuously constructed and reconstructed in everyday interactions and transmitted by the socialization agents (family, school, social and work organizations, church, mass media, etc.), play an important role in the process of IA. Social factors influence what is socially desirable, as well as the configuration of an individual’s personality, which in turn influences what each person considers personally desirable in others. Thus IA, as well as physical attraction and love, is not simply a psychological process, as it is commonly perceived by laypeople, but more precisely is a sociopsychological (in fact, biosociopsychological) process.

### 3.4. Interpersonal Interactions

Social pressures and social influence are usually exerted through interpersonal relationships, especially within the groups in which people interact every day. Group attitudes and group norms regarding other groups and other people (according to their ages, race, gender, attitudes, etc.) have a major influence on the factors that are likely to elicit and to maintain IA. Conformity, polarization, or even group thinking may all have a role to play. Readers should be aware that the behavior (actions, feelings, and cognitions) of the people with whom a person interacts every day has an important effect on that person's attitudes and behavior, including the IA he or she feels toward other people, and including the decisive role of reciprocity of attraction. Others' opinions and the pressure they exert are often quite relevant to people's feelings of IA toward someone (whether they lead to conform to their standards or to reject them by means of a "reactance process").

### 3.5. Individual

Finally, of course, the individual is also important. For some authors this is the main factor in any case; for others, the individual is less relevant than people would like to admit, being strongly shackled by biological and social forces. People's behavior in general, and more specifically feelings of attraction toward others, are constrained by biological, cultural, social, and interpersonal processes and factors but in truth are "acted out" or felt by a specific individual with specific psychological characteristics. Psychological studies have revealed that some features are more relevant to IA, such as physical attractiveness, social skills, or personality variables that happen to be socially recognized as valuable. In addition to these, and as mentioned previously, each person will feel attracted to certain specific features that are especially valued by the individual in question. Furthermore, one's own psychological characteristics are of considerable importance, because their similarity to the psychological characteristics of the other is one of the main factors (if not the main one) responsible for IA. In addition to these, complementarity of needs (between attractor and attracted) and personal need for affiliation are two further relevant variables affecting the IA process. Clearly, there are too many factors here to ignore, underlining the crucial role that the individual characteristics of each person play in establishing his or her feelings of IA toward one person rather than another. Nevertheless, psychological researchers are still far from

establishing any kind of simple (and probably simplistic) taxonomy of types of personalities associated with tendencies to feel IA for other specific types of personalities.

## 4. APPLICATIONS OF THE PSYCHOLOGY OF IA

Psychological knowledge of IA not only leads to a better understanding of the human being, but also offers opportunities to apply that knowledge, and thus solve some personal and interpersonal problems, increasing subjective well-being and quality of life.

As was pointed out earlier, the absence of feelings of IA to others and/or not being the object of attraction for others will probably result in severe psychological difficulties. So, psychological knowledge about what generates IA can be used to assess and train people in these characteristics, at least to some extent. Using sociometry, social-skills training, role-playing, and other systematic and well-established psychological techniques, it is possible to assess IA and to teach people ways to make themselves more attractive, not only physically but also personally. Sociopsychological knowledge of IA can be used to give guidelines for attracting others and initiating and establishing relationships, with a reasonable degree of success, by taking into account factors such as propinquity, context, physical appearance, verbal and nonverbal communication (conversational and listening techniques, attractive body language etc.), overcoming fear of rejection, self-presentation strategies, socially desirable characteristics, accurate rhythm of self-disclosure, and so on. More constructively, psychologists can design programs conceived to increase tolerance of differences, so as not to diminish the attraction felt for someone other than oneself.

However, there is far more basic research than there are application programs for IA, and the majority of these applications are centered on the couple as the fundamental context of application. Furthermore, it is physical attraction, much more than IA, that has been applied in various fields. And, in addition, normal "applications" of IA are frequently no more than research studies concerning IA applied to a specific area (e.g., the study of the role of IA on the personnel selection process, sports team performance, jury decisions, or consumer behavior), much more than the true application of IA knowledge on solving personal and interpersonal problems and increasing quality of life.

Nonetheless, it is possible to find examples of studies concerning IA applied to certain contexts, in which those studies (and future ones) have at least a potential field of application (see Table III).

There are some well-known sociopsychological (and unconscious) processes that have been verified in some applied contexts, such as the halo effect (the influence of first impression on psychological inferences about a person), the physical attractiveness bias (the attribution of positive characteristics to attractive people and not-so-positive ones to unattractive people), and the relevance of IA on persuasion processes (IA as one of the main factors of communication agents related to success in persuasive communication). An IA halo effect has been detected in teacher expectations about his or her students (the Pygmalion effect), in falling in love, and even in contexts such as personnel selection processes and jury deliberations. A physical attractiveness bias (in some ways related to the halo effect, but not exactly the same) has been noted almost everywhere, from work organizations to everyday interactions. The impact of IA on persuasive communication is especially relevant in areas in which the capacity for influence should rely on sources other than IA, e.g., advertising, politics, or therapy.

**TABLE III**  
**Real and Potential Contexts of Application of Social Psychology of IA**

<i>Context of application</i>	<i>Type of relationships</i>
Work organizations	Horizontal relationships, vertical relationships, personnel selection process, promotion process
Education	Teacher–students, among students
Family	Parents, parents–children, siblings
Intimate relationships	Intimate friendships, loving relationships
Therapy	Therapist–client
Sports	Manager–players, among members of a team
Courtroom	Among all figures: jury, judge, defendant, lawyer, etc.
Politics	Politicians–citizens behavior (e.g., voting)
Advertising	Salesmen and models–consumer behavior (e.g., buying)
Health	Social models—healthy/unhealthy behavior (e.g., smoking)
Everyday life	Interracial and intercultural relationships

One promising and highly relevant context of application is IA in work organizations. There is some evidence to support the influence of IA in both vertical and horizontal interactions, including the personnel selection process, and therefore in productivity and the quality of one's working life. For example, in general, data indicate a positive relationship between IA and interview outcomes, taking into account measures of personal attraction, physical attraction, and even some physical measures such as height and weight; in addition, IA moderates interactions in recruitment interviews by gender. Along the same lines, other results indicate that physically attractive applicants in jobs requiring high (but not low) levels of social competence were recommended for hire more often, were expected to perform at a higher level, and were perceived as more representative of successful employees than were unattractive applicants. It is well established that professionals are as susceptible to this bias as are college students and laypeople.

Other fields of application include the following relationships:

- between teacher and students, in which the relevance of IA has been shown to be a determinant of first impressions (and thus, as is generally accepted, on attributions and behavior), not only among students but also on the teacher's expectations and behavior toward his or her students;
- between parents and children, a field in which IA is no doubt more important than physical attraction;
- between lovers, an area in which there are real applications, such as modeling strategies to become (or to seem) more attractive to another (and, for single people seeking a relationship, psychological knowledge of IA can be applied as a serious means to form relationships);
- between therapists and clients, which might seem intriguing, given its supposed irrelevance;
- between the members of a sports team, and between players and coaches; here, IA has an effect in terms of improving results, cohesion, and satisfaction;
- between judge, jury, and defendant, with potentially dramatic consequences—for example, the fact that, in general, people tend to look more favorably on victims and defendants who are good looking;
- between politicians and citizens, IA influencing, for example, voting intention and even voting behavior, to some extent;
- between the seller (or its advertising models) and consumer, since it is well established that the IA of the

seller (or the model who advertises a product) has a major role on the success of the persuasion process for people to buy the product (here physical attraction plays a part as well as IA in general);

- between social and TV models and common people, knowing that the IA of social models (mainly perceived via TV) influences the extent to which people imitate their behaviors. This is relevant, for example, in relation to good and poor health (smoking, anorexia, etc.), violence/nonviolence, social/unsocial behavior;

- and finally, of increasing importance nowadays, there are international, interracial, and intercultural relationships in which IA is especially difficult to raise (for all of the aforementioned reasons), but at the same time is especially important in today's global environment.

### See Also the Following Articles

Emotion ■ Interpersonal Behavior and Culture ■ Interpersonal Perception ■ Work and Family, Relationship between

### Further Reading

Berscheid, E., & Walster, E. (1978). *Interpersonal attraction*. Reading, MA: Addison-Wesley.  
 Buss, D. M. (1994). *The evolution of desire: Strategies on human mating*. New York: Basic Books.

Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.  
 Cook, M., & Wilson, G. (Eds.). (1979). *Love and attraction*. Oxford: Pergamon Press.  
 Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality & Social Psychology*, 24, 285–290.  
 Duck, S., & Gilmour, R. (Eds.). (1981). *Personal relationships*. London: Academic Press.  
 Etcoff, N. (2000). *Survival of the prettiest: The science of beauty*. New York: Anchor Books.  
 Fletcher, G. (2002). *The new science of intimate relationships*. Malden, MA: Blackwell.  
 Gaines, S. O., Buriel, R., Liu, J. H., & Ríos, D. I. (1997). *Culture, ethnicity and personal relationships processes*. New York: Routledge.  
 Hatfield, E., and Rapson, R. L. (1996). *Love and sex: Cross-cultural perspectives*. Boston, MA: Allyn & Bacon.  
 Huston, T. L. (Ed.). (1974). *Foundations on interpersonal attraction*. New York: Academic Press.  
 Rhodes, G., and Zebrowitz, L. A. (2001). *Facial attractiveness: Evolutionary, cognitive, and social perspectives*. Westport, CT: Ablex.  
 Sangrador, J. L., & Yela, C. (2000). What is beautiful is loved: Physical attractiveness in love relationships in a representative sample. *Social Behavior and Personality*, 28, 207–218.  
 Wilson, G., & Nias, D. (1976). *Love's mysteries*. Glasgow: Collins Sons & Co.  
 Yela, C. (2002). *El amor desde la Psicología Social: ni tan libres ni tan racionales* ("Love from social psychology: Not so free nor so rational") (2nd ed.). Madrid: Pirámide.



# Interpersonal Behavior and Culture

John Adamopoulos

Grand Valley State University, Allendale, Michigan, USA

1. The Interface of Culture and Interpersonal Behavior
2. Theoretical Approaches to Culture and Interpersonal Behavior
3. Thematic Approaches
4. Applications: Cross-Cultural Interaction
5. Conclusion  
Further Reading

## GLOSSARY

**circumplex models** Models of the structure of personality traits, behavior, or emotions that are characterized by a circular arrangement of their components.

**cultural syndromes** Patterns of behavior, cognitions, and emotions that characterize whole cultures or groups of cultures and account for psychological diversity across cultures; the most popular cultural syndromes are individualism (which emphasizes individual autonomy) and collectivism (which focuses on interdependence among group members).

**distal versus proximal determinants of behavior** Causes of behavior that are remote (other causes may intervene between them and the behavior) versus causes that are immediately connected to behavior.

**ecocultural framework** A framework for cross-cultural psychology that views human behavior both as the result of shared psychological processes and as a series of adaptations to ecological challenges.

**in-group** A group of people who feel strong emotional and psychological ties and who perceive that they share a common fate.

**internal (dispositional) versus external (situational) attributions** Explanations of one's own behavior, or of the behavior of others, that focus on the internal personal characteristics of the individual engaging in the behavior as opposed to external demands and social pressure.

**interpersonal structure** The psychological dimensions that describe the meaning of interpersonal relations or behavior.

**resource exchange theories** Theories of social behavior that view interaction as the exchange of physical and psychological resources needed for individual survival and well-being.

**utilitarian models** Explanatory models of human action that emphasize the purpose, function, and expected consequences of the action.

Psychological theories about the determinants of interpersonal behavior traditionally have either postulated intraindividual causes (e.g., attitudes, personality traits) or looked at the immediate social situation (e.g., group pressure, influence of the social or physical environment). However, a number of recent approaches have emphasized the role of culture in the production of behavior. In these approaches, culture has been conceptualized as a system of adaptations that transcends the individual human individual or as a subjective meaning system (or, in some instances, as both). Interpersonal behavior, then, can be viewed in part as a consequence of cultural influences or as part of a set of constructs that constitute culture. Either perspective provides considerable theoretical insight

into the meaning, origins, and function of interpersonal behavior and facilitates a fresh examination of established social psychological phenomena.

## **1. THE INTERFACE OF CULTURE AND INTERPERSONAL BEHAVIOR**

There have been numerous attempts to provide a psychologically relevant definition of culture, but there is no single definition that has met with widespread agreement. One difficulty stems from disagreements between those who view culture as a purely subjective and ideational construct and those who emphasize its observable characteristics (e.g., behavioral processes). A second disagreement stems from an ontological question: Is culture a set of invented convenient categories for the classification of certain observations, or does culture have an existence of its own (i.e., independent of the individual person)?

Whatever their definition of choice, many psychologists accept the notion that culture can be viewed as a series of adaptations to environmental challenges that enhance the lives of a group of people and, therefore, are transmitted from one generation to the next. In this context, culture provides a functional basis for the analysis of interpersonal behavior. In other words, interpersonal behavior is understood as serving a purpose and as having some sort of utility in the lives of people. In a broader sense, it should be clear that any definition of culture may have profound implications for the psychological explanation of interpersonal behavior.

## **2. THEORETICAL APPROACHES TO CULTURE AND INTERPERSONAL BEHAVIOR**

During the past two decades or so, we have witnessed a significant growth in the number of psychological theories of interpersonal behavior that are framed within a cultural perspective. Although not exhaustive, the following classification provides a sense of the direction of these efforts.

### **2.1. Structural Models**

The pioneering work of Osgood on the measurement of connotative meaning across cultures has been an important influence on the development of structural

models of social behavior. Osgood showed that three psychological dimensions—evaluation (*E*), potency (*P*), and activity (*A*)—appear to be universal (culture-common) structures involved in the understanding of the implicative meaning of things. In other words, Osgood proposed that, regardless of specific linguistic or cultural practices, humans understand the meaning of things by judging, among other attributes, how good or bad (*E*), strong or weak (*P*), and fast or slow (*A*) they appear to them.

Inspired by this work, a number of psychologists explored the structure of the interpersonal domain. The fundamental aim underlying much of this research has been the identification of a set of psychological dimensions that remain (relatively) invariant across cultures and that reveal how humans understand and communicate the meaning of interpersonal behavior. Particularly noteworthy in this effort has been Triandis's approach, which significantly influenced cross-cultural psychological research during the 1970s and later years. Triandis developed methods that could be used to study what he termed "subjective culture," that is, a group's characteristic way of understanding its social environment. This process often involved exploring belief and value systems, norms, roles, attitudes, and expectations in various cultural groups.

Central to the approach was the structural analysis of social behavior within and across cultures using a variation of Osgood's "semantic differential" technique, which Triandis called the "behavioral differential." This instrument consisted of scales gauging the perceived likelihood that people engage in different behaviors in various social contexts. Data reduction techniques were then employed to extract both culture-common and culture-specific dimensions that captured the variety of meanings of interpersonal behavior in very different cultures.

As in the case of connotative meaning, structural analyses of interpersonal behavioral intentions observed in many cultures showed an impressive convergence of findings. It appears that social behavior, regardless of what other local features may also characterize it, is understood across a wide range of cultures in terms of a few fundamental dimensions: (a) association–dissociation, (b) superordination–subordination, and (c) intimacy–formality. Specifically, the first dimension appears to gauge the extent to which an interpersonal behavior involves affiliating, approaching (as opposed to avoiding) another person, and generally giving (as opposed to denying) resources to others. The second dimension involves the notion of dominance and

control over one's social environment and gauges the extent to which a particular behavior communicates the exercise of authority and status (as opposed to submissiveness and acceptance of another's authority). Finally, the third dimension captures the extent to which a behavior involves social and physical proximity and seems to be independent of the rewarding or negative characterization of the specific behavior.

This dimensional structure, part of which has been confirmed by independent research exploring the interpersonal domain of personality, has been quite useful in summarizing the array of behaviors exhibited in daily activity and in focusing theoretical analyses of social interaction. Most significant theories in social psychology explore the behavioral domains covered by the three dimensions. For example, work on prosocial behavior or aggression clearly falls within the domain of the first dimension (association–dissociation), research on obedience and conformity falls within that of the second dimension (superordination–subordination), and the study of personal relationships falls within the domain of the third dimension (intimacy–formality).

The psychological comparison of any two or more cultures based on the three interpersonal dimensions can offer a very broad understanding of fundamental similarities and differences between or among the cultures. For example, Triandis and colleagues have shown that although research participants in the United States, India, and Japan use the three dimensions in understanding behavioral intentions in interpersonal contexts, the meaning of particular behaviors may be quite different in each of the three cultures. For example, excluding a person from one's neighborhood may imply dissociation and dominance in the United States and Japan, whereas it may be considered primarily a formal behavior in India, where the caste system plays an important role in determining the range of personal interactions.

A final question regarding this approach concerns the origins of the dimensions as psychological universals. It is generally assumed that judgments of the social environment based on the three structures offer a significant survival advantage to humans and, thus, led to their emergence over time and across various cultures. Alternative views, also based on utilitarian notions, are discussed later in this article.

## 2.2. Cultural Syndromes

Early work on culture and interpersonal behavior was successful in identifying a large number of cross-cultural differences (and similarities) in interpersonal behavior.

For example, rates of conformity in an Asch-type situation, where group members who report an incorrect perceptual judgment induce naive research participants into also making incorrect perceptual judgments, may differ across seemingly similar cultures. However, this work was relatively unsuccessful in explaining such differences or similarities in a systematic manner. A basic problem was the relative absence of broad cultural theories appropriate for psychological research.

This situation began to change during the 1980s when a number of dimensions that could be used to describe behavioral patterns within and across cultural groups began to emerge. Hofstede's examination of work-related values in a large number of cultures provided the impetus for the development of the notion of "cultural syndromes," that is, patterns of behavior, beliefs, values, attributions, judgments, and emotions that are characteristic of whole cultures (and often groups of cultures). These patterns can be used to describe, in a systematic and predictable fashion, cross-cultural similarities and differences in interpersonal behavior. The approach has been one of the most important directions of research in the area during the past 20 years or so and has been particularly emphasized in the work of Triandis.

Whereas a number of cultural syndromes have been identified, the most successful and popular ones by far have been those of individualism and collectivism. As applied to the analysis of interpersonal behavior, individualism is a cultural pattern that is characterized by a preference for control of the social environment, personal achievement and fulfillment, self-reliance, competition, and (perhaps most important) a desire for autonomy and independence in thought, action, and emotion. Collectivism, on the other hand, is characterized by a preference for interpersonal harmony, in-group cohesiveness, and an emphasis on the interdependent nature of human relations, thoughts, and emotional experiences.

The success of these two constructs, which had been available in the social sciences for a long time before psychologists began to employ them in their work, centers entirely on their usefulness in making predictions about cultural differences in interpersonal behavior and social judgment. For example, the finding that individualism is characteristic of cultures found mainly in North America and Europe (particularly Western Europe), whereas collectivism is characteristic of Asian cultures, leads to the rather obvious prediction that social values and attitudes, as well as interpersonal behavior, will be more similar among European and

North American cultures than among European and Asian cultures.

Results of research tend to support these predictions, but not without occasional contradictory findings. In general, however, agency and independence are the building blocks of Western philosophical, value, and moral systems, whereas responsibility to community, duty, and interdependence are often the foundations of morality in collectivist cultures (e.g., India, China). Individualism–collectivism differences have also been described regarding patterns of and preferences for competition (individualism), cooperation (collectivism), and conformity (collectivism), although the particular situation (e.g., whether an interaction takes place within or outside the context of an in-group) often is critical in determining the direction of the difference. In a relevant finding in this area, collectivists have been shown to exhibit greater subordination and conformity toward in-group members, but also greater dissociation and hostility toward members of out-groups, than have individualists.

Other major differences between individualist and collectivist cultures have been identified in attributional processes. For example, collectivists are not as likely to commit what has been (mis)labeled in Western social psychology as the “fundamental attribution error,” that is, the tendency of observers of behavior to rely overwhelmingly on internal and dispositional explanations for the behavior of others. Obviously, differences in attributional style can lead to differences in behavior toward others because others’ behavior is perceived to be motivated by very different causes. In other words, our reactions to what others do toward us will be substantially shaped by how we tend to interpret their actions and by whether we tend to focus on personality-based or situational explanations for their behavior.

The usefulness of the constructs of individualism and collectivism has been demonstrated repeatedly during the past two decades or so. With increased reliance on these constructs, however, comes the danger of confusing description with explanation. Critics of the approach have pointed to the problem that often what is predicted as a consequence of individualism or collectivism (e.g., a behavioral difference between people of two cultures) is actually nothing more than a constitutive part of the construct itself, rendering a seemingly theoretical prediction nothing more than a tautology. With care, such problems can be avoided and these constructs can continue to provide a rich theoretical basis for the discussion of cross-cultural differences in interpersonal behavior, cognition, and the like.

## **2.3. Utilitarian Approaches**

An implicit theme underlying much cross-cultural research over the past half-century has been the notion that much of what is observed and identified as either a difference or a similarity among cultures is functional for members of a particular group or for humans at large. This theme has either been articulated explicitly or been given a central role in several theories of interpersonal behavior.

### **2.3.1. Behavioral Prediction Models**

Concern with the determinants of social behavior, particularly with the attitude–behavior relationship, during the 1960s and 1970s led to the development of several models that boasted fairly good behavioral prediction. The models’ predictive success has been demonstrated in a variety of applied settings, from advertising to fertility-related practices, in a number of different cultures. These models, particularly those formulated by Fishbein and Triandis, have a number of common features. For example, they all treat attitudes and other intrapsychic variables as major causes of interpersonal behavior, but they also emphasize the influence of the social environment in determining behavior. Invariably, the models assume that this latter influence is filtered through internal processes. For example, the effect of others’ opinions is thought to lead to the formation of personal norms of appropriate behavior. In turn, the norms are expected to affect the formation of behavioral intentions or self-instructions to engage in particular social behaviors.

These models have an explicitly utilitarian character because they consider expectations of the consequences (negative or positive) of a behavior to be among its most important determinants. The models are not always in agreement regarding the precise mechanism through which perceived behavioral consequences affect behavioral outcomes. For example, Triandis’s model treats this influence as an independent process, whereas Fishbein’s model embeds the influence within the attitude formation process. In either case, however, a utilitarian component of human social behavior is explicitly taken into consideration. In other words, these models rely on a combination of utility and rationality to offer explanations of interpersonal behavior.

Cultural influences are not considered explicitly in such models. However, there is a clear assumption that the set of beliefs, expectations, and personal norms that

function as proximal determinants of behavior are themselves influenced substantially by cultural inputs. Thus, culture is viewed as an important but distal behavioral determinant. For example, cultural beliefs about the role of a woman in the family, coupled with local religious or moral values about contraception and conception, may lead to the formation of personal norms that influence a woman's decisions about the size of her completed family. It would be expected, of course, that such norms would play a far more significant role in determining the behavior of collectivists, whereas personal wishes and attitudes would be more influential determinants of the behavior of individualists. Indeed, there is some empirical evidence that cultural norms about appropriate behavior tend to become internalized among collectivists, presumably having an even greater effect on their behavior.

### 2.3.2. Resource Exchange Theories

Another class of utilitarian models has been developed to account for systematic relationships among various dimensions of interpersonal behavior. These models typically are based on the assumption that all human social interaction involves—indeed, is designed for—the exchange of physical and psychological resources. Foa and colleagues established that any social behavior appears to involve the exchange of one or more of six classes of resources: (a) love, (b) services, (c) goods, (d) money, (e) information, and (f) status. These resource classes form a circumplex model defined by the dimensions of abstractness–concreteness and universalism–particularism. The first dimension refers to the material (concreteness) as opposed to the symbolic (abstractness) nature of a behavior, whereas the second dimension refers to the extent that the relationship of the actor to the target of the behavior plays a significant (particularism) or trivial (universalism) role in the satisfactory completion of the interaction.

Extensions of this work have led to formulations that attempt to explain the emergence of the culture-common psychological dimensions of social behavior reviewed earlier and of cultural syndromes such as individualism and collectivism. Adamopoulos has argued that the dimensions of concreteness and particularism, along with the question of whether a resource is given or denied, form fundamental constraints on human interaction and are components of a differentiation process that yields uniquely defined features of interpersonal meaning. For example, superordination is defined as the denial of abstract and particularistic

resources. In this fashion, the interpersonal dimensions discussed earlier may be conceptualized as emerging from this process of constraint differentiation over long periods of time. Adamopoulos extended this idea in 1999 to account for the emergence of individualism and collectivism as cultural patterns of interpersonal behavior. In this case, an additional constraint, the orientation of a particular behavior to secure resources for the self as opposed to another person, was included in the theoretical model.

In general, these resource exchange models attempt to account for two major problems: (a) the fundamental structure of interpersonal behavior and the interrelationships among its major dimensions and (b) the emergence of interpersonal structure as a function of basic social constraints. This orientation is necessarily of a universalistic nature. In other words, these models tend to focus on themes that transcend cultural boundaries. At the same time, resource exchange models can be sensitive to cultural differences by taking into consideration the extent to which various cultures expose their individual members to different constraints. It is assumed that such differential exposure leads to behavior systems with different interpersonal emphases. For example, if ample symbolic resources are available for exchange to members of a particular culture, they are more likely to engage in behaviors that involve power and superordination than are members of cultures that provide primarily concrete resources in interpersonal exchanges. The systematic description of resource availability in various cultures is, unfortunately, a task of enormous complexity and one that is likely to take a very long time to complete.

### 2.3.3. Sociality

Fiske's theory of sociality proposes that there are four basic types (elemental forms) of social relationships: (a) communal sharing, where individuals within an in-group share pooled resources; (b) authority ranking, where powerful leaders control and dispense necessary resources as needed; (c) equality matching, where individuals of equal status and rank enjoy comparable resources; and (d) market pricing, where interpersonal relations are a function of market demands and expectations of profit.

Fiske has argued that these four types of relationships account for most of the variability in interpersonal exchanges across very different social situations. In the context of this theoretical model, the role of a culture is to shape and present particular forms of the

four types of relationships to its individual members in ways that are relevant to their lives. For example, authority ranking may be manifested very differently in the context of a military organization than in that of a cult or gang. At the same time, the theory facilitates connections among seemingly unrelated contexts.

The relative simplicity of this theoretical model makes it quite attractive. However, critics have pointed out that, in fact, the four types of sociality deal with only a portion of the range of interpersonal relations. Specifically, the four types can easily be classified along the dimensions of association–dissociation and superordination–subordination. This, of course, indicates that this model cannot explain other aspects of interpersonal behavior.

#### **2.4. Ecocultural Framework**

The view of culture as a series of individual and group adaptations to the environment has a long history in the social sciences. During recent years, many psychologists have used Berry's "ecocultural framework" as a means of embedding such notions about adaptation into psychological research across cultures. This framework rests on two fundamental assumptions. First, humans share a number of basic psychological processes in thought and action, although such "psychological universals" appear with nearly infinite variation across the world's cultures. Second, human behavior is adaptive in that it rises in response to ecological, cultural, and situational challenges that account for the tremendous diversity evident in human life.

As elaborated by Berry and colleagues in 2002, the ecocultural framework is a dynamic system of population- and individual-level variables with multiple sources of influence. Despite its dynamic nature, however, the emphasis in the framework is clearly on the manner in which ecological and sociopolitical systems of variables affect psychological processes and behavior at the group or individual level. This is accomplished through a series of biological and cultural adaptations, although once again the emphasis is overwhelmingly on the process through which cultural adaptations are transmitted at the group level and, thus, determine the psychological functioning and behavior of individual humans.

The framework has been primarily a useful guide to research on cognitive and perceptual processes across cultures, but it has also been used to account for certain cross-cultural differences in interpersonal behavior. In general, the advantage of the ecocultural framework is

that it can facilitate the exploration of behavioral phenomena by transcending standard explanations in experimental psychology. Such explanations typically revolve around intrapsychic or, at best, interindividual processes. For example, personality traits, attitudes, and the presence and behavior of others (i.e., the immediate social situation) are the most typical determinants of behavior found in psychological theory. The ecocultural framework focuses attention on the role of physical resources and subsistence economies in determining the range of human activities. Often, such accounts can be quite insightful and add considerable depth to psychological explanation.

### **3. THEMATIC APPROACHES**

Traditionally, the investigation of interpersonal processes and behavior has been organized topically due to the large number of theoretical perspectives available and the inability of any single theoretical model to account for the variety of the observed phenomena. A comprehensive review of cross-cultural research on interpersonal behavior is beyond the scope of this article. Instead, this section outlines four major areas of relevant research and, in each case, provides examples of how cultural variables may contribute to better understanding the behavior of interest.

#### **3.1. Social Norms and Conformity**

Norms, or rules for appropriate behavior in particular social contexts, demonstrate the influence of culture on interpersonal behavior because they reflect the opinions and actions of others. Norms can be internalized and become embedded in individual belief systems, particularly among collectivists. Thus, it is expected that conformity to norms varies significantly across cultures.

Considerable research, much of it done in the context of the ecocultural framework, has demonstrated the validity of this hypothesis. Specifically, consistent differences in conformity to group norms have been found between cultures with highly structured and rigid interpersonal relationships (high conformity) and cultures with considerable autonomy in how individuals fulfill their role obligations (low conformity).

A plausible, but partially speculative, explanation of such differences rests on fundamental ecological and structural differences between the two types of cultures. Cultures in which the ecology favors the emergence of agricultural economies must develop vertical

authority systems in which the presence of powerful and respected leaders facilitates the coordination of human activities involved in agricultural production. In turn, this process necessitates the development of socialization practices that encourage conformity to authority figures and social norms. In contrast, ecologies that favor the development of, for example, subsistence economies based on hunting (an activity that requires much less cooperation among many people) may lead to a relatively greater emphasis on individual autonomy during socialization and, thus, to less pressure to conform to group norms in social life.

### 3.2. Prosocial Behavior

Cross-cultural differences in prosocial or positive forms of interpersonal behavior have been demonstrated consistently. These differences range from being willing to mail a letter for someone else, to asking for directions, to helping a disabled person cross the street. In fact, the variety of helping behaviors studied over the past 30 years or so has been limited only by the ingenuity of the investigators and implementation constraints.

Research findings in this area are not easy to interpret because the specific cultural characteristics that may account for prosocial behavior patterns are not always well understood. However, a consensus appears to be emerging that the cognitive load associated with everyday life, and the resulting demands on people's attention, may explain the relatively higher likelihood of helping in rural (as opposed to urban) and poor (as opposed to affluent) areas.

Other cultural characteristics may also account for observed rates of helping in interpersonal contexts. For example, cultures differ in their traditions regarding how one behaves toward strangers or foreigners. Thus, it is possible that in some cultures the influence of norms for helping others may be moderated by the relationship between the person requesting help and the person offering it. In some collectivist cultures, where social interaction is guided by norms of interdependence and where people often rely on each other for the acquisition of resources necessary for survival, interpersonal behavior follows particular cultural scripts. Such scripts inform people's notions about responsibility and duty, and they lead to expectations that others will be of help if the circumstances demand it.

### 3.3. Aggression

Aggression is generally defined as behavior intended to harm another person physically or psychologically.

The study of aggression has been of great interest to social and biological scientists of diverse persuasions and has resulted in a number of universal findings. Two of the best-known findings are as follows. First, aggression and violence are truly ubiquitous phenomena but are manifested in very different ways and occur at different rates around the world. Second, with very few exceptions, boys and men tend to behave more aggressively across cultures than do girls and women, respectively.

The precise processes that explain these findings are not clearly understood. Biological mechanisms have been proposed on occasion, but the supporting evidence is typically not very strong. Most psychologists have instead relied on a variety of sociocultural processes. For example, child-rearing practices, socialization, and social learning have often been evoked as explanations for both male-female and cross-cultural differences in the incidence of aggressive and violent acts. It appears that, in many cultures, men not only learn to engage in more aggressive behavior than do women but also learn to interpret aggressive behavior differently than do women. Furthermore, as documented in recent work on "cultures of honor," the incidence of violence may be higher in some cultures because their members are socialized to protect themselves and their possessions. This results in increased tolerance for aggressive behavior because such behavior is viewed as serving a self-protective function.

The preceding processes all explain aggression in terms of normative influences. A contrasting perspective may also be formulated, although it is one that receives scant attention from psychologists. According to this view, aggression may be more likely to occur in those domains of interpersonal interaction for which there are no strong sociocultural norms of appropriate behavior. In such contexts, aggression becomes instrumental for individuals in acquiring certain classes of resources. The investigation of this intriguing proposal necessitates a fundamental shift in research orientation away from social influence processes and toward the manner in which humans attempt to establish agency within a complex and often threatening social environment.

### 3.4. Interpersonal Relationships

The study of relationships between people has been the focus of research primarily in European and American psychology. The impetus behind this work has been a growing concern with intimacy, personal relationships,

attraction, love and marriage, and dissolution processes in relationships. In all of these areas, culture is thought to play a significant role, but in many instances, actual empirical research is lacking. For example, we may understand the mythology that sustains (and often destroys) American and European personal relationships, but similar research in non-Western cultures is practically nonexistent.

Intimacy has been traditionally defined as self-disclosure that is largely a rewarding experience for two persons. The negative aspects of intimacy as a psychological phenomenon have generally been ignored or considered only in the context of relationship dissolution processes. An alternative view of intimacy is that it is the exchange of particularistic and concrete resources. In this context, intimacy is considered a universal phenomenon whose manifested form and functions are shaped and modified by culture. For example, collectivists tend to have few but highly intimate relationships, whereas the reverse is true for individualists. It is possible that this is because the greater availability of social norms in collectivist cultures makes intimate interactions more concrete experiences. In contrast, individualists must search for appropriate information to clarify the social situation. Thus, developing intimate relationships may require more work for individualists and may explain why individualists tend to consider a romantic component critical for many intimate relationships. The strong kinship and community ties that often bind collectivists may diminish the importance of romantic love in personal relationships.

Interactions involving self-disclosure that have traditionally dominated Western conceptualizations of intimacy are, by definition, based on the rather abstract process of searching for information to better define the situation. Such conceptualizations are clearly inadequate in light of the better-defined and predictable nature of interpersonal exchanges in collectivist cultures. Consequently, alternative views of intimacy that also take into account systems of social interaction from non-Western cultures must be developed.

#### **4. APPLICATIONS: CROSS-CULTURAL INTERACTION**

The theoretical approaches to the analysis of interpersonal behavior reviewed earlier typically provide explanations not only for similarities but also for

differences in the actions in which people engage during their everyday lives. Thus, these approaches can offer considerable insight into understanding a variety of situations in which people with different cultural backgrounds interact with each other.

Psychologists have studied cross-cultural interaction extensively and from diverse perspectives that have emphasized both its positive and negative consequences. For example, Berry and colleagues have developed models of psychological acculturation and the effects of cross-cultural contact on individual psychological processes. These models specify some of the conditions under which persons who come into contact with another culture may select to replace their cultural identity with that of the new culture (assimilation), may choose some form of biculturalism (integration), or may reject either the new cultural identity or both the new and the old cultural identities.

Other analyses of cross-cultural interaction have focused on the concept of adjustment, that is, the ability to develop and maintain psychological well-being even in the context of interpersonal exchanges with considerable conflict and communication difficulties. Research has established that successful adjustment to changing conditions following cross-cultural contact improves performance on a variety of tasks and, of course, decreases the amount of stress typically associated with cross-cultural experiences.

Cross-cultural interaction often involves exchanges between people who rely on very different value systems to interpret each other's behavior. As mentioned earlier, there appear to be some fundamental differences in the manner in which individualists and collectivists make attributions about social behavior. Specifically, individualists seek the causes of behavior in people's personalities, attitudes, and belief systems (i.e., internal attributions), whereas collectivists tend to prefer attributions that involve the experience of social pressures, obligations, duty, and group membership (i.e., external attributions).

Interaction between members of these two distinct types of culture may be problematic and may lead to substantial misunderstandings. For example, an individualist is likely to attribute a perceived lie in an interpersonal exchange to some flaw in the actor's personality, whereas the (collectivist) actor may attribute the lie to the constraints of the social situation (e.g., an obligation to be polite and respectful). The need to resolve conflicts of this sort—quite common in cross-cultural contact—has led to the

development of a number of instruments and programs for cross-cultural training during recent years. One of them, the culture assimilator, has been shown to increase the chances that people from different cultures will make similar attributions in various social situations.

## 5. CONCLUSION

The influence of culture on psychological processes is nowhere more evident than in the study of the generation and specific manifestation of interpersonal behavior. At the same time, the task of explicating the role that culture plays in behavioral processes is enormous given the range of behaviors that humans produce and the variety of processes necessary to account for them.

A number of theoretical models are available to accomplish this task, each with its unique strengths and weaknesses. Often, these models make very different assumptions about the manner in which culture is involved in the production of interpersonal behavior. This diversity of assumptions may actually be helpful at this stage given that we still know relatively little about the culture–behavior link. What is needed is a cultural theory appropriate for psychology with considerable power for description, prediction, and

explanation. Such a theoretical system does not exist at this point.

## See Also the Following Articles

Aggression and Culture ■ Cross-Cultural Psychology, Overview ■ Cultural Syndromes ■ Prosocial Behavior, Development of

## Further Reading

- Adamopoulos, J., & Kashima, Y. (Eds.). (1999). *Social psychology and cultural context*. Thousand Oaks, CA: Sage.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications* (2nd ed.). New York: Cambridge University Press.
- Fiske, A. P. (1990). *Structures of social life: The four elementary forms of human relations*. New York: Free Press.
- Segall, M. H., Ember, C. R., & Ember, M. (1997). Aggression, crime, and warfare. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology*, Vol. 3: *Social behavior and applications* (2nd ed., pp. 213–254). Boston: Allyn & Bacon.
- Smith, P. B., & Bond, M. H. (1999). *Social psychology across cultures* (2nd ed.). Boston: Allyn & Bacon.
- Triandis, H. C. (1994). *Culture and social behavior*. New York: McGraw–Hill.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.



# Interpersonal Conflict

Panos S. Kordoutis

Aristotle University of Thessaloniki, Thessaloniki, Greece

1. Understanding and Managing Conflict in Interpersonal Relationships
2. Relationship Features and Psychological Processes Involved in Conflict
3. Overview of the Conflict Mechanism: Directives and Keys to Managing Conflict
4. Managing Conflict
5. Conclusion  
Further Reading

## GLOSSARY

**correspondence of outcomes** The degree to which the outcomes of one partner in an interdependence relationship can be attained without preventing the other partner from attaining his or her outcomes; perfectly correspondent outcomes lead to pure coordination, whereas perfectly noncorrespondent outcomes lead to pure conflict.

**dispositional attributions** Identifying as causes of behavior internal factors and processes such as intentions, motives, and personality traits.

**interdependence or outcome interdependence** A social situation involving two or more individuals who depend on one another for achieving individually valued outcomes and goals.

**interference** Describes Partner A's actions in obstructing Partner B's effort to bring about a personally desirable outcome; interference may be actual or perceived and may be intentional or unintentional.

**interpersonal skills training** Developing skills that are pertinent to constructive conflict resolution such as building problem-solving relationships with other disputants,

seeking multiple solutions to conflict situations, and viewing conflict issues from an "outsider's" perspective.

**maximum minimum** The best possible of the worst outcomes that one is willing to accept in a negotiation setting, where agreements can be attained only by lowering both disputants' outcome levels.

**mediation** The intervention of an independent third party in a dispute to facilitate disputant interaction, identify sources of conflict, and promote constructive communication toward a mutually satisfying resolution.

**minimum maximum** The lesser among the best outcomes that one is willing to accept in a negotiation setting where disputants cannot attain their optimal outcomes.

**outcome** A general term employed to indicate how drives and motives shape preferences and goals worth pursuing in social situations of interdependence; outcomes may be material possessions (e.g., money, a house, a car) or abstract states (e.g., social status, security, happiness).

**relative deprivation** A perception that outcomes in some realm that is valued and important to self are inferior to a reasonable standard; relative deprivation is a potential source of conflict.

**social comparison** Using others as a comparison standard in evaluating own qualities such as efficacy, ability, judgments, and attitudes.

**social norm** A consensually accepted rule about behaving in social settings that has been internalized by individual society members; breeching of a social norm by an individual is sanctioned but may also cause pangs of conscience.

Interpersonal conflict is a process triggered when one of the parties in a relationship perceives that the other adversely interferes with the attainment of own

outcomes and goals. It may also be a consequence of perceiving own outcomes as severely inferior to those of the other in the relationship and to a commonly accepted standard.

## **1. UNDERSTANDING AND MANAGING CONFLICT IN INTERPERSONAL RELATIONSHIPS**

Interpersonal conflict is a by-product of individuals' need to establish and maintain relationships for their intrinsic wellness and survival value. Relationships ensure that personal goals, such as a comfortable life, a rewarding professional career, security, fulfillment, and love, can be attained to a qualitative and quantitative level that either would have been impossible outside a relationship (as in the case of love) or would have required much more effort, time, and resources. Despite their motivation to uphold relationships, individuals will often come into conflict with their partners as the realization of their respective goals clash, impede one another, or become the medium of an unfavorable comparison with the other's achievements. Conflict can be a painful process, with destructive consequences for the well-being of individuals and for the potential of relationships to yield valuable outcomes for their members and society. Of course, many conflicts cause little rancor and are resolved peacefully. Intimate couples, friends, roommates, and office colleagues can often find mutually satisfying ways in which to eliminate their differences. In professional settings, managers and employees can frequently settle their disputes through simple discussions or, when matters get more complicated, through mutual demands with official labor-management negotiations. However, there are times when people will resort to overt or subdued hostility, or to verbal or physical aggression, so that they can force the resolution of differences in their favor. Hostility in all forms will also occur as a means of expressing displeasure about what is perceived as an unfair resolution of relationship differences. In either case, the ambience of personal or professional relationships is severely disturbed. Simple human interaction becomes unproductive and harmful not only for the disputants but also for those directly or indirectly related to them. For example, consequences of persistent conflict often involved in divorce are numerous and distressing for the personal development of the disputants and their

children. Unfortunately, divorce rates involving intense conflict in the United States, in European Union member countries, and in other Western countries continue to be high. In professional settings, even subdued conflict between fellow employees can undermine the productivity and creativity of an entire work unit, adversely affecting the quality and quantity of products and services. Considering how threatening conflict can be to the stability and productivity of relationships, individual well-being, and the cohesiveness of society, it is not surprising that it has attracted a lot of scientific theory and research.

Social psychology, in particular, has dealt with conflict in an effort to understand it and identify ways in which to manage it. Because managing conflict requires knowledge and understanding of conflict, social psychologists have tried to develop a theory that would allow the prediction and control of conflict. The agenda of such a theory includes questions such as "What is conflict?," "What causes conflict?," "Under what circumstances does conflict occur?," "How is conflict expressed?," and "What are the social and psychological consequences of conflict?" If one could have answers to at least some of these questions, one would be able to perhaps predict, for instance, when conflict would be manifested and with what intensity. As such, one would be able to manipulate the timing and circumstances of its manifestation so as to divert or even prevent it. Unfortunately, despite efforts, there is no true comprehensive conflict theory that can permit concise prediction and control of conflict behavior. Nevertheless, there are numerous theoretical approaches that have contributed importantly to a fair understanding of the variables involved in conflict. Thus, one might not be able to totally predict and control conflict, but one knows that its basic source is the motivational structure of the relationship in which the conflict parties are involved, namely, the fact that they are, to a certain degree, interdependent for achieving their valued outcomes. One also knows that other, more qualitative features of the relationship will influence how conflict will be expressed and that attributions about the causes of conflict to the other's internal dispositions and character will escalate conflict expression. The major psychological processes implicated by social-psychological theory in conflict production are "interference" and "relative deprivation." Interference is based on the perception that the significant other in the relationship interferes with own goals, whereas relative deprivation is based on perceiving that own outcomes are relatively inferior to those of the other and to a commonly accepted outcome

standard. By means of another intermediary process, social comparison, the first two may work to undermine self-esteem, and this in turn generates conflict to preserve, restore, and enhance self-esteem. This loose mechanism of conflict production, put together for the purposes of the current discussion by what is known about conflict so far, provides directives and keys with which to understand the development of conflict resolution and management methodologies. For example, the role of self-evaluation and self-esteem in the conflict mechanism strongly suggests that the aim of conflict resolution should not be limited to removing the “realistic” difference between disputants but rather should also ensure that the self-esteem of the conflicting parties is preserved, restored, and enhanced (if possible). The importance of knowing conflict in developing conflict management is also illustrated by the way in which conflict management skills have used the research finding that dispositional attributions escalate conflict. Conflict mediators and disputants are trained to eschew from articulating inferences about others’ intentions in the relationship and personality characteristics.

The theoretical understanding of conflict not only has contributed to the development of conflict resolution methodologies but also has been conducive to diluting two lay misperceptions. First, most people commonly perceive conflict as a negative event of destructive consequences without paying sufficient attention to its more subtle positive effects. This is due mainly to the saliency of destructive consequences and to the survival value intrinsic in paying more attention to negative events. Second, people believe that disputants themselves cannot handle conflict and that conflict management skills cannot be learned. These misperceptions are perhaps even more destructive than conflict itself because they discourage people from training in conflict resolution and encourage people to avoid conflict at all costs, even in cases where conflict is called for and potentially beneficial to their personal interests and relationships. Indeed, research has indicated that people can be trained to handle conflict effectively, making the best out of it for their self-interests, self-respect, and self-expansion. Conflict can also be a creative and constructive process, helping in relationship functionality and development toward greater productivity and fruitfulness for both partners.

This article first explicates, by reference to theory, research, and real-life examples, the role of relationship features and psychological processes in yielding conflict. By pointing to how this knowledge is put into practice, the article then turns to analyzing major

conflict management methodologies. It concludes by underscoring the main issues underlying conflict escalation and conflict resolution.

## **2. RELATIONSHIP FEATURES AND PSYCHOLOGICAL PROCESSES INVOLVED IN CONFLICT**

The basic relationship feature involved in generating conflict has to do with motivational structure, that is, with the fact that the partners’ goals are interdependent. This so-called structural feature of outcome interdependence forces people to compare their outcomes, triggering social comparison. Thus, the next subsection discusses relationship structure along with the psychological process of social comparison. Social comparison, with its self-evaluative results (e.g., “I am better off than others,” “I am worse off than others”), leads to two other psychological processes: interference and relative deprivation. These latter processes, being different in nature, are discussed in separate subsections. Nevertheless, their aim is the same, that is, to determine the extent to which the relationship partner is the culprit of negative self-evaluation. This section concludes with a discussion of how conflict experience and conflict expression are regulated by the qualitative characteristics of relationships.

### **2.1. Relationship Structure and Social Comparison**

Interpersonal conflict springs from interdependence, an intrinsic characteristic of interpersonal relationships. Relationship partners depend on one another for achieving their wishes, goals, and preferred end states, generally called outcomes. Outcomes refer to material possessions (e.g., money, a house, a car) or to abstract states (e.g., social status and recognition, security, happiness, love). In addition to being attractive for satisfying specific needs, outcomes carry desirability and status value assigned by society. Thus, they contribute to identity definition and self-evaluation by means of social comparison. By comparing material and abstract possessions with those of others, people come to an understanding of where they stand in relation to others. In this way, relationship outcomes surreptitiously influence self-worth and establish potential for conflict within the relationship.

Relationship outcomes cannot be attained to the same qualitative and quantitative levels with the same ease and economy (in time, effort, and energy) outside a relationship. Hence, people are strongly motivated to enter relationships for optimizing their individual outcomes. That is not to say that individuals are indifferent to the common relationship outcomes or ignorant of the fact that others enter relationships motivated by the same individual outcome optimization goals. To the contrary, they are usually aware that no interdependent parties can ignore others in pursuing their goals and that all can influence one another's experiences, motives, preferences, behavior, and outcomes. As Lewin put it in his classical work on social conflict, relationship partners know that their "locomotion" toward attaining their goals may cross or be crossed by that of others. A lot of cognitive work needs to be carried out in planning the pursuit of own goals in an interdependence situation because others' plans must also be understood and predicted. Possibly, individuals will have to cooperate with others to maximize joint outcomes or compete for their own on the basis of some commonly agreed rules. Of course, rather than putting that cognitive effort into coordination and negotiation with others, direct and immediate access to desirable objectives may be demanded, and then conflict and fighting might arise. Success and failure will occur in predicting and planning, cooperating, and competing in conflict and fighting. In either case, there will be consequences for own outcomes as well as psychological repercussions for both self-worth and the continuing interaction of actors in the relationship. Others may facilitate, accelerate, and promote the attainment of own outcomes or, in contrast, may obstruct, delay, block, or simply interfere with their attainment by merely unilaterally pursuing their own goals in a relationship. The strong motivation of people to stay in a relationship and benefit from the advantages of outcome interdependence explains why they are willing to tolerate the nuisance of others' interference with personal outcomes.

## **2.2. Interference**

It is reasonable to expect that people sharing the same social milieu may also see commonly desirable outcomes as worthwhile to pursue. Unfortunately, outcomes, whether material or abstract, may be limited; they might not be available to all, at all times, and in the same quality or quantity. Actually, some may be depleting fast as more people claim them for

themselves and as more people claim more of them. Realizing this fact, social norms have been instituted in nearly all organized societies about coordinating, cooperating, and competing in claiming commonly valued outcomes. People also hold general implicit expectations, known as lay theories, about fair and accepted practices in pursuing such outcomes.

If partners sharing a relationship decide to pursue similar depleting outcomes, their ways are likely to cross. In that case, they might perceive each other as interfering with own outcomes, and conflict might erupt to remove interference and restore unhindered and direct access to desired outcomes. Conflict is most probable when each of the two partners insists on unilaterally claiming for self the maximum of the desired objective and eschews coordination, cooperation, or competition along norms and lay theories.

Money is an obvious example of a material outcome that depletes quickly as the needs to be satisfied in an intimate relationship increase. For instance, consider a married couple whose members must decide how to allocate their budget to satisfy common and individual objectives. Couple members have to reach an agreement on prioritizing objectives, and their first option likely would be to accept a norm about distribution of resources between the spouses. One such norm is that maximizing of joint outcomes must come first. Alternatively, they may devise their own rule of prioritizing, give up on some of their personal goals, and take turns in having money allocated to each member at different times. For some couples, particularly those whose members maintain a relatively rich profile of activities outside the relationship, budget allocation agreements are not easy to devise without minor or major conflict.

Another example of a more abstract outcome that depletes quickly and happens to be in great demand in interpersonal relationships is time. How much time should friends spend with each other, how often, and in what quality (e.g., leisurely time in doing things together, discussions about each other's plans and dreams, simple chatting on the phone)? How does a person decide how much time to spend with each friend and in what way? Social and cultural norms about this issue are variable and also depend on individual expectations about the quality of the specific friendship. Friend A's allocation of time to Friend B might not correspond to Friend B's willingness to allocate time for the friendship. Friend B might have expected more and higher quality time from Friend A and, thus, might interpret Friend A's decision as intentional devaluation of their friendship. In a way, although the valued commodity (i.e., time) seems

the same, it acquires different meanings and values for each friend because each appears to have different expectations about friendship. In reality, the friends' valued outcome called "friendship time" is noncorrespondent. If one friend could be satisfied in terms of his or her valued outcome (i.e., be given abundant time with friend), the other friend would have to be dissatisfied. Of course, this situation may give rise to displeasure, resentment, open conflict, or even a falling out.

Actors in an interpersonal setting may indeed come into conflict due to noncorrespondent outcomes. Correspondence describes the degree to which the outcomes of one partner in an interdependence relationship can be attained without preventing the other partner from attaining his or her outcomes. Noncorrespondence, then, refers to a situation where for one partner to attain his or her goals, the other partner must fail to do so. Perfect noncorrespondence of outcomes leads to pure conflict, whereas perfect correspondence leads to pure coordination. Noncorrespondence implies that interference and conflict may arise not only by the fact that the members of an interpersonal relationship pursue similar, albeit depleting, outcomes. They could just as well pursue dissimilar outcomes that happen to be incompatible. For instance, a married couple might not realize that when one spouse pursues a high-profile career while the other seeks a low-key family life, potential for conflict may be accumulating. One spouse's lifestyle generates outcomes that are noncorrespondent to those yielded by the other's lifestyle. Scarcity of time invested in home or family activities and an abundance of time spent with nonfamily make difficult the common pursuit of a low-key family life. Spouses must pursue their aspirations unilaterally, hoping that this will not negatively influence relationship maintenance. Alternatively, they could both change their lifestyles so as to be able to carry out more joint activities and enjoy common outcomes.

Other's interference with own goals may have cognitive and emotional ramifications because it makes other and own outcomes salient to self, thereby triggering social comparison. Self-evaluation may then ensue, typified by thoughts such as "Do I really merit this goal?" and "Am I able to achieve this target?" Social comparison may also yield perceptions of threat to self-worth projected onto the other such as "Her goal is to prove that I am ineffective" and "His aim is to humiliate me."

Either category of cognitions contributes to the multiplication of conflict manifestations and to conflict escalation. Removal of interference might not be sufficient to settle conflict because disputants are focusing

on the perceived threat to their self-esteem rather than on the interference per se. Other's interference may have already severely damaged perception of self-worth and efficacy, and conflict may persist with the aim of restoring them.

### 2.3. Relative Deprivation

Paradoxically, noninterference may also contribute to interpersonal conflict. This situation is referred to as relative deprivation. People experience relative deprivation when they realize that own outcomes are severely inferior to a commonly accepted standard and that reaching the standard by own efforts is infeasible. In a situation of this sort, significant others (who meet the standard) with whom people have relationships may unknowingly aggravate people's own predicaments. Social comparison again works to underscore both inferiority of outcomes and self-debasement. The likelihood of these effects is high when a person is comparing self to a significant other with respect to outcomes having a central role in self-definition of both parties. If the relationship is a stable personal one and the social norms encourage it, the person with the poorer outcomes is particularly likely to expect the other to interfere so as to alleviate the former's own predicament. Of course, the significant other might not agree that he or she should assist and might not be willing to do so. Noninterference or lack of assistance may cause envy and bitterness in the person with the poorer outcomes. Moreover, it might be interpreted as an active attempt by the significant other to denigrate him or her and to contemptuously seclude him or her from "what everyone has a right to have." Conflict is bound to follow. Relative deprivation as a cause of conflict is more common between social groups than between persons given that people generally either find themselves in social settings with people whose level of outcomes is similar to their own or tend to choose significant others with similar outcomes rather than dissimilar ones.

### 2.4. Experiencing and Expressing Conflict: The Role of Qualitative Relationship Features

Some manifestations of conflict remain discreetly in the background until conflict has fully erupted. However, most of them are not discreet at all and need no introduction. At times, all people have

experienced behavioral manifestations such as arguments, intense disagreements, antagonism, fights, and even physical violence against other people. They have experienced the emotions involved such as apprehension and mistrust as well as being irritated, hostile, angry, bitter, resentful, sullen, and ominously silent. All people have also ruminated on thoughts regarding the causes of other people's behavior, the ways in which other people interfered with their own wishes and goals, their intentions in so doing, and their predispositions and characters. People probably have experienced conflict in their relationships with acquaintances, friends, family members, spouses, and lovers as well as in their professional or more formal relationships.

Overt behavioral manifestation of conflict and the subjective experiences involved, emotions and cognitions, may differ both quantitatively and qualitatively depending on the relationship type. Relationships have been characterized in many ways. The previously discussed feature of outcome interdependence is primarily used to characterize the motivational structure of relationships because interdependence of outcomes is intertwined with the potential of developing conflict in relationships. Thus, relationships of high interdependence, such as intimate ones, are very likely to yield conflict. The more individuals depend on one another for more of their positive outcomes, the more likely they are to interfere with one another's positive outcomes. High interdependence provides a lot of opportunities for conflict while at the same time motivating people to insist on claiming their relationship outcomes because (a) they value the relationships in themselves (i.e., what they can derive from the relationships specifically), (b) they might not have alternative options (i.e., alternative relationships offering just as good outcomes), and/or (c) they might not be willing to invest in the costly procedure of seeking and constructing new relationships.

However, conflict expression per se seems to be influenced by three other qualitative (rather than structural) dimensions of relationship features: formality, stability, and personal/task orientation. More specifically, a relationship may be formal or informal, stable or transient, and personal or task oriented. A formal relationship is governed by strong social norms providing salient guidelines about appropriate interpersonal behavior. In contrast, although an informal relationship may also embed implicit behavioral guidelines, it is less directing and leaves more room for partners themselves to define interaction rules. A

transient relationship is temporary and tends to occur with irregular timing, often in different places. To the contrary, a stable relationship lasts longer, has relative time regularity, and tends to occupy identifiable spaces. A task-oriented relationship is established for the sake of accomplishing specific tasks, whereas a personal relationship is initiated mainly for its own sake and the outcomes are accrued through the relationship. Communication in the former focuses on tasks and actions, whereas communication in the latter focuses on the people and their dispositions.

An example of a relationship type that is highly interdependent for outcomes while being informal, stable, and personal is, of course, the married couple. Married couples may experience conflict over any number of issues, including budget allocation, division of household chores, family roles, failed mutual expectations, jealousy, neglect, and infidelity. In fact, it appears that the more a person knows another person, the closer the relationship is and the more likely the partners are to have had experienced a broader spectrum of all the preceding behaviors, emotions, and cognitions to different degrees. Consider the common example in married couples counseling of neglect that may give rise to particularly intense and expressively "rich" conflict. When one member of the relationship maintains a high level of rewarding life outside the relationship, allowing that area to absorb more and more of that person's interest, energy, and time, the other member may start perceiving the gap between past level of personal rewards from the relationship and present outcomes to grow. Failed expectations, deterioration of outcomes, and the inevitable social comparison with the significant other will soon trigger perceptions of interference with own relationship aspirations and a sense of relative deprivation. Feelings of frustration and self-deprecation may follow, and conflict behaviors on the part of the "underbenefited" member to restore level of outcomes and reclaim loss of self-esteem will be manifested. Because the focus in close relationships tends to be on the partners and their dispositions, attributions will be made about the "meaning" of the other's interference or lack of interference (i.e., relative deprivation). Representative thoughts could include "This is another sign that her feelings about me have changed," "He does not care about me," and "I am unable to attract her anymore." Dispositional attributions escalate conflict, and removal of interference might not be sufficient to settle it because the disputants are then focusing on perceived intentional threat to self-esteem rather than on the outcomes of their relationship. Conflict in such cases may be constructive

in the sense that a relationship member is painfully made aware of the fact that he or she is entangled in a relationship that does not accrue positive outcomes to him or her anymore. Restructuring or even leaving the relationship may be necessary to extinguish negative outcomes. Nevertheless, high interdependence, as well as the possible misperception that there are no alternative solutions, might not allow an individual to see conflict as a positive and constructive opportunity. Third-party assistance may be required in this situation.

The adage that the more people know one another, the more likely the conflict, in no way implies that conflict cannot erupt between strangers and acquaintances. More often than not, conflict will occur within a transient formal or informal relationship, that is, with people who one hardly knows or who one never actually meets, for example, the inconsiderate driver who will not allow one's car to overtake his car despite one's desperate signaling and horn honking that one is rushing to an emergency or the "high hat" assistant at the store in the posh side of the town who will answer one's naive questions with a polite yet reluctant and indignant smile. Expression of conflict in transient relationships may include a different, possibly more limited, and yet more extreme set of behaviors compared with those identified in other kinds of relationships.

The anonymity of informal transient relationships may diminish respect for social norms. The brevity and superficiality of the contact with the unknown others, the nonsaliency of social norms in such situations, and the expectation that the interactions will not be followed up all may result, by means of disregarding the human qualities of the others, in particularly acute and exceptional expressions of conflict behavior involving verbal aggression or even physical assault. The anonymous crowds at athletic events have often promulgated intense expressions of interpersonal conflict among spectating fans. Similarly, continuous movement and speed, and the impersonal indirect communication through signaling and horn honking, may precipitate antagonistic behavior between drivers on highways. Extreme behavior is not as likely in formal transient relationships, such as the encounter with the store assistant, because the environment is rich with salient norms of conduct dictating limits to appropriate "protest" behavior and sanctions for inappropriate behavior. There is less anonymity, and the likelihood that the interaction with the same person might be followed up is not negligible. As a consequence of these factors, conflict expression in formal transient relationships is usually subdued and restrained.

What about formal stable relationships such as work and professional relationships? These bear a characteristic that could breed conflict, that is, increased interdependence for outcomes. On the other hand, these relationships usually include a strong set of norms due to their formality. Because they are also stable, individuals cannot discount the future consequences of their negative behaviors. Finally, because these relationships have survival value for the individuals involved, they cannot be easily abandoned for other relationships and so members are particularly motivated to retain them.

Consider a small health care institution that hires two psychologists without a clearly delineated profile for each one's duties. In such a work setting, one is given the chance to define his or her exact job profile and to enjoy the work while gaining recognition and credit toward future promotion. Hence, the two psychologists are likely to compete in demonstrating effectiveness, diligence, and expertise—a productive state of affairs for both the institution and its patients. However, assuming that this small institution has only one higher status position for a psychologist, competition for recognition and credit may soon degrade to antagonism. In patient assessment multispecialist meetings, the two may try to undermine each other's input. They may selectively use scientific knowledge and data that counter each other's specific analyses just to raise generalized hard criticism of each other's approach. Alternatively, they may resort to side remarks and whispers to other members of the team or to personal comments about the rival colleague's personality and way of working. The result is, of course, detrimental to the productivity of the whole team and potentially threatening to patient care. The team must divert attention away from case analysis to the arguments of the two psychologists and must invest time and energy in deescalating conflict and refocusing on the task at hand.

### **3. OVERVIEW OF THE CONFLICT MECHANISM: DIRECTIVES AND KEYS TO MANAGING CONFLICT**

Interference and relative deprivation are the two basic psychological processes involved in conflict outbreak. Both processes implicate the self, particularly self-evaluation and self-esteem, because they activate social comparison with the significant other in the relationship and with a commonly accepted outcome standard. Thus, conflict is not solely about securing for self the

valued resources of interpersonal relationships but also is about enhancing, defending, and restoring self-esteem. Social comparison is a third intermediary process that acts as the interface of the other two to the self. Interference and relative deprivation lead to social comparison with the informational input that the self requires to understand where it stands in relation to others and to evaluate own abilities, achievements, social status, and more abstract personal end states such as satisfaction, happiness, and well-being. If the result of such social comparison is negative self-evaluation, conflict behavior is likely to arise because the self implicates the other in obstructing attainment of own outcomes (i.e., interference) or in nonintervening to assist in attaining better personal outcomes (i.e., relative deprivation). The saliency of self-evaluation, brought about by social comparison, diverts attention of relationship parties away from realistic conflict issues to dispositional attributions about each other's behavior. These attributions usually implicate the other in intentional threats to own self-esteem and are highly conducive to conflict escalation. All three processes are inconceivable outside a relationship that is desirable in its own outcomes by both parties. Underlying these processes is a relationship structure of outcome interdependence. The members of the interpersonal dyad depend on one another for their valued outcomes, and it is this very dependence that motivates them to remain in the relationship despite the hurdle of interference and the increased risk of relative deprivation. Therefore, a paradox results. Conflict behavior is more probable, and often more enduring, the greater the interdependence in the relationship. How conflict will be manifested, however, will be greatly influenced by the more qualitative features of the relationship. The specific repertoire of behaviors used to manifest conflict—their aggressiveness, intensity, persistence, and other behavioral characteristics—will be regulated by the formality or informality, stability or transience, and orientation (task vs personal) of the relationship.

The preceding overview of the most important concepts that have been implicated by research and theory in generating interpersonal conflict also attempts to assemble a loose framework for a general understanding of the psychological mechanism yielding conflict behavior. The framework, depicted in [Table I](#), connects relationship features to the psychological processes of interference and relative deprivation through social comparison. The core cohesive concept mobilizing the supposed psychological mechanism of conflict

production is self-esteem, that is, the need of the self to self-enhance by self-evaluating and comparing with others.

This loose mechanism of conflict production provides basic directives in conflict resolution and several keys to conflict management methodology. First, it suggests that in conflict resolution, the ultimate aim is to protect, preserve, restore, and (possibly) enhance the self-esteem of the parties involved in conflict. The "realistic" goal of eliminating the source of interference or relative deprivation should be viewed only as a step toward self-esteem preservation or enhancement. Attempting to eliminate interference and relative deprivation without taking into consideration that this action reflects on the self-esteem of the disputant parties will only temporarily appease the conflict. In conflict management methodology, skills development is inconceivable without taking into account the role of dispositional attributions in conflict escalation. Among the most important negotiating skills is avoiding allusions to the personality and intentions of disputant parties (e.g., "name calling," "reading between the lines"). A third party intervening to assist in resolving conflict cannot possibly proceed to realistic problem-solving propositions that satisfy both parties without a full analysis of the motivational structure in the relationship. The analysis should show the different ways in which each disputant's valued outcomes could be attained to a maximum possible level of individual desirability considering that the other wants to do the same. The intervening party should identify the maximum minimum (i.e., the best possible of the worst outcomes that one is willing to accept in a negotiation setting) and the minimum maximum (i.e., the minimum among the best outcomes that one is willing to accept in a negotiation setting). Furthermore, the mediator should have a good understanding of the dimensions used by conflicting parties to self-compare with the other in the relationship. Which outcomes are significant to each party, and why are they used as referents of comparison with the other and (possibly) with a common social standard? What are the consequences of such comparisons for a party's self-esteem? The nature of interference should also be clearly understood. How do the parties interfere with each other's outcomes, and with what costs and benefits for self and other? How will changes in the interference pattern change the cost-benefit ratio for self and other? Answering such questions can provide an indication of how motivated and willing the parties will be to shift their behavior to less interference or noninterference during negotiation.

TABLE I  
Relationship Features and Psychological Processes Involved in Interpersonal Conflict

Relationship features	Processes		
	Intermediary	Basic	
	Social comparison	Interference	Relative deprivation
<b>Structural–Motivational</b>		<b>Informational input</b>	
Outcome interdependence (high or low)	Self-evaluation becomes salient. Outcomes are compared with other's and with common standards. Result is input to "basic" processes	Perception of other as interfering with own outcomes Self and other's outcomes	Perception of self-outcomes as inferior to other's and to common standards
		<b>Mediating cognitive–affective output</b>	
		Self-attributions, self-devaluation (e.g., "Do I really merit this goal?")	Self-attributions and self-debasement
		Dispositional attributions about other (e.g., "His aim is to humiliate me")	Dispositional attributions about other (e.g., "She intends to deprive me of what everyone has a right to have")
<b>Qualitative</b>		<b>Behavioral output: Conflict expression</b>	
Stability–transience Formality–informality Task–personal orientation		Conflict expression is regulated in terms of persistence, intensity, restraint, and behavioral repertoire by the qualitative features of the relationship.	

#### 4. MANAGING CONFLICT

The fact that conflict is more or less natural to interpersonal relationships perhaps accounts for the common perception that disputants can do little to manage conflict themselves. However, research indicates that disputants trained in interpersonal skills may be able to contain, resolve, and make the best out of conflict, turning it into a constructive and creative process. Indeed, at the individual level, when a person has the skills to handle conflict, this otherwise destructive and painful process may work to mobilize and enhance individual abilities, resources, and ingenuity. It may further motivate self-knowledge, assert self-determination, clarify personal goals and desires, and encourage self-expansion. At the dyadic level, when each of the disputants possesses the skills to handle conflict (to a greater or lesser extent), the disputants may turn a poor interdependence relationship into a fruitful one for both parties. It appears that when

individuals appreciate their outcome interdependence, conflict may provide the thrust that pulls together resources, knowledge, and abilities into devising ideas and solutions that probably would not have been as innovating and productive without conflict.

Unfortunately, the notion of interpersonal skills training for managing conflict has not yet been met with wide social acceptance. For instance, educational systems have done little to integrate such training into their programs. Unfortunately, lack of skills in managing conflict continues to be widespread, and people often find themselves unable to understand how they got involved in escalating conflict and how to handle it. A solution often sought in these situations is intervention by a third party, usually someone with good training in interpersonal skills in conflict management. This section first reviews some of the well-known interpersonal skills for managing conflict. Then, it describes the basic principles and forms of third-party intervention.

## 4.1. Interpersonal Skills in Managing Conflict

Six basic interpersonal skills for managing conflict have been identified in the past. The first five are viewed as antecedents of the sixth. These basic interpersonal skills must be well acquired and put to effective use before any attempt is made to proceed on to learning how to practice the sixth, that is, problem solving. The latter is the most critical interpersonal skill involved in conflict resolution and, thus, is presented in more detail than the others.

### 4.1.1. Active Listening

This skill involves listening carefully to what the other person has to say and reflecting on the points he or she is trying to make. Feedback should be provided about what is understood. Points perceived to have ambiguous or double meanings should be clarified with the other by questions that do not include answers and do not assume interpretations.

### 4.1.2. Empathy

This skill refers to trying to see things from the perspective of the other person. It involves attempting to identify with expectations, fears, and concerns that seem to underlie the other's arguments and responses.

### 4.1.3. Assertion

This skill suggests that one's point of view, thoughts, feelings, and pursued outcomes are expressed clearly in concrete terms.

### 4.1.4. Avoiding Dispositional Attributions

Attributing events to the other person's personality, motives, and/or intentions should be avoided at all costs. Communication should be about tasks and issues, whereas affect expressed should not deviate from personal respect and recognition to the other's right to have claims in the relationship.

### 4.1.5. Constructive Feedback

This skill refers to identifying and rewarding positive behaviors toward resolution while not failing to mention or ignoring negative behaviors. Reasoning should

be provided as to why behaviors are positive or negative on the basis of their consequences to self, other, or both.

### 4.1.6. Problem Solving

Alternative solutions aimed at satisfying the goals and needs of both parties should be generated. This requires (a) each party developing separately a clear hierarchy of goals and needs, with his or her optimum goals placed on top and followed by the maximum minimum and the minimum of the maximum; (b) discussing proposed solutions, beginning with the target of fulfilling optimum goals for both parties and then, if discussions come to an impasse, discussing on the basis of satisfying maximum minimum goals for both parties and then, if this also fails, further discussing on the basis of minimum maximum goals; (c) examining whether maintaining the relationship and maximizing joint outcomes is indeed preferable to pursuing own optimum goals independently; and (d) repeating Steps b and c until both parties concede to aim for lower than their initial outcomes so that a mutually satisfying solution (referred to as "satisficing") can be reached. If necessary, the hierarchy list of preferred outcomes should be reviewed and restructured.

The conflict resolution process may last a long time and may require the control and manipulation of several parameters. Hence, it is important that disputants monitor their effectiveness in managing conflict. Based on Deutsch's 1994 analysis of conflict resolution training, four criteria captured by the following questions can be used to evaluate whether the preceding skills have been put to work effectively. First, has a cooperative, problem-solving relationship with the other person been initiated, and how well is it maintained? Second, do disputants discuss a reasonable range of alternative and mutually satisfying proposals as a result of problem solving? Third, do disputants seem to gradually adopt an "outsider's" perspective or an analytic perspective to the conflict? Fourth, are the various forms of conflict expression kept to a minimum during the process?

Influential applied research, such as Johnson and Johnson's 1994 research on constructive conflict in schools, has consistently documented the benefits of systematic interpersonal skills training in interpersonal educational as well as in other settings. However, as noted previously, educational systems have not yet integrated such training into their programs. Unfortunately, rather than preparing to personally handle the relatively frequent, inevitable, and potentially harmful life events of conflict, people seem to prefer resorting to the

intervention of a third party after conflict has already escalated and negative consequences are already difficult to bear. Although such persistent conflict may have caused a lot of damage or even irreparable harm, it is usually never too late to seek the help of a specialist.

## 4.2. Third-Party Intervention

A third (independent) party can often be invited by the disputants to resolve conflict. This is more likely when disputants are motivated to maintain their relationship despite its temporary negative outcomes or when they intend to exit the relationship with the maximum possible gain and minimum possible damage for self and other. Third-party intervention may also be offered when conflict has damaging consequences for others indirectly related to the disputants and when the functionality of their social setting is threatened. For instance, in the earlier example, negative consequences had resulted from the conflict between the two psychologists for the functionality of group multispecialist meetings, the work ambience, and the welfare of patients in the small health care institution. In this case, the chief administrator of the institution might have intervened in a number of informal ways. The administrator could have advised the psychologists individually to behave “professionally,” to avoid hostility, and to focus on patient care, or the administrator could have facilitated a constructive dialogue between the two. Alternatively, the administrator could have arranged so that the two psychologists never interact or interfere in each other’s duties. If these informal approaches had proven to be ineffective, the administrator could have employed a more formal approach such as mediation.

In general, intervention strategies differ in terms of their legalistic and nonlegalistic character as well as the intervening party’s interaction with the disputants.

### 4.2.1. Adjudication

This involves a judge or jury in a court setting. The interaction of these parties with the disputants is strictly formal and follows legally prescribed channels. Disputants are obliged to abide by the court’s decision.

### 4.2.2. Arbitration

This involves little interaction of the intervening party with the disputants. The arbitrator listens to both sides and decides what the solution should be. This

approach is also legalistic, to a certain degree, in that norms and prescribed rules must be taken into account in providing a solution. Disputants may have agreed before arbitration to abide by the arbitrator’s decision. However, they are not obliged to do so.

### 4.2.3. Fact Finding

This involves even less interaction with the disputants. The third party’s role is to gather information so as to arrive at an independent judgment regarding the dispute. This strategy could also be said to have a legalistic character in the sense described in the preceding subsection (on arbitration). The input of the “fact finder” is only advisory.

Conciliation and mediation, described next, are non-legalistic processes and tend to be more psychological in nature.

### 4.2.4. Conciliation

This involves a lot of interaction with the disputants, albeit limited to opening and maintaining communication channels and exchange of information. The conciliator does not get actively involved in disputants’ discussions and agreements.

### 4.2.5. Mediation

This involves an independent third party who must actively interact with the disputants and work with them in identifying the source of conflict and in reaching a mutually satisfying agreement. Mediation is largely preferred to other forms of third-party intervention in many types of interpersonal relationships. This may be due to its relatively lower cost to other, particularly legalistic, conflict resolution procedures or, more important, to the willingness and commitment of anyone choosing mediation to actually resolve conflict. The latter may, in part, account for the finding that mediation yields more effective, lasting, and satisfactory resolutions for disputants. It is also noteworthy that even when conflict is not resolved in a satisfactory way, the mediation process per se tends to improve disputant relations both within and outside the mediation framework.

Psychological research has dealt more with mediation than with any of the other intervention processes, focusing on understanding its mechanisms and its effectiveness. Following is a succinct description of the basic features of this procedure.

The mediator must be a person who is trained in interpersonal skills and has a good understanding of the theoretical underpinnings of conflict and experience in monitoring conflict. Mediation usually starts with preliminary contacts between the mediator and each disputant. The aim of these contacts is twofold: (a) to evaluate disputants' willingness and commitment to resolving conflict and (b) to identify the fundamental source of conflict and of minor issues causing friction. Setting the rules of subsequent interactions follows this stage. The mediator explicates the procedure, delineates his or her role as a neutral moderator, and states emphatically the modes of communication that will and will not be allowed during interaction, usually giving specific examples. For instance, interruptions, dispositional attributions (e.g., name calling), and generalizations and interpretations of the other's arguments should be avoided. Mediation proper then begins. Each disputant is asked to present his or her view of the conflict situation within a specific time limit. The mediator makes sure that no one is interrupted and that the rules are followed. Once both disputants have presented their views, the mediator may summarize each position and asks disputants for further clarifications. This is followed by the problem-solving phase, during which the two disputants and the mediator all collaborate in reviewing (a) each party's interests, goals, and desirables outcomes; (b) each party's sources of interference and nuisance; and (c) as many feasible and desirable solutions to conflict as is possible. Common discussion is followed by separate discussion with each disputant, during which the same issues are reviewed with an emphasis on discussing as many feasible and desirable solutions to the conflict as is possible. During these contacts, the mediator must reevaluate commitment to conflict resolution and exclude the possibility that mediation is being used by either disputant as a strategy to attain a hidden agenda. Furthermore, points of resistance, reluctance, and apprehension regarding the constructive process must be identified, and their causes must be discussed. If this stage yields a range of mutually acceptable alternatives, an agreement can be reached. Agreements are recorded, typed, and signed by both parties. Language used in documents should be specific, concrete, and unambiguous. Disputants are expected to honor their signatures. If the problem-solving stage does not produce a mutually acceptable solution, agreement cannot be reached and the entire procedure must be repeated. The stages described do not necessarily follow one another. An earlier stage might need to be rerun at any point during the

mediation procedure. Thus, mediation can be painstaking and time-consuming. Nevertheless, as has often been observed in relevant research, the procedure *per se* may improve the relationship of the conflicting parties, providing them with a thrust to proceed on to resolving conflict on their own.

## 5. CONCLUSION

Conflict can undermine the integrity of a relationship and generate negative consequences for the relationship members. However, such detrimental outcomes are due to uncontrollable escalation of conflict and not to conflict *per se*. In fact, conflict is an integral part of all interpersonal relationships. It is a product of the interdependence that is necessarily developed between two (or more) parties so that they can both (or all) enjoy valued outcomes that could not be enjoyed to the same degree, or with the same ease and quality, outside the specific relationship. Conflict can be managed so as to yield outcomes constructive for the relationship and positive for its individual members. Relationship members should be on guard for uncontrollable conflict escalation, which is usually due to (a) actual or perceived interference of the other with the attainment of own goals, (b) actual or perceived relative deprivation of own outcomes with respect to a commonly accepted standard and the standard attained by the significant other in the relationship, (c) perceiving interference or relative deprivation as a threat to self-esteem, (d) focusing on enhancing self-esteem rather than on specific conflict issues, (e) maximizing own outcomes unilaterally by ignoring the other's or at the expense of the other's outcomes, (f) manifesting and retaliating conflict behaviors, and/or (g) making dispositional attributions about conflict behaviors.

Training in interpersonal skills, such as active listening, empathy, assertion, constructive feedback, and problem solving, can deescalate conflict and turn it into a constructive process. Mediation by a third party is often required when people lack conflict resolution skills. The mediator attempts to facilitate interactions between disputants so that they can negotiate for themselves. Mediation requires analyzing facts, goals, outcomes, and conflict behaviors as reported by each party in individual and common sessions with the mediator. As a constructive process, conflict can regulate self-interest in a relationship. Moreover, it can maintain the relationship's productivity of positive outcomes to a maximum level for the members involved. It can work to clarify needs and goals and to delineate the

role of individual members. It can focus disputants on mutually satisfying solutions of objective conflict differences, moving them away from the persistent and futile effort to restore self-esteem by means of conflict behavior. Last, but not least, it can disentangle individuals from an unproductive relationship that perpetuates their negative outcomes by precipitating exit.

### See Also the Following Articles

Interpersonal Behavior and Culture ■ Interpersonal Perception

### Further Reading

- Deutsch, M. (1994). Constructive conflict resolution: Principles, training, and research. *Journal of Social Issues*, 25(2), 7–41.
- Holmes, J. G., & Murray, S. L. (1996). Conflict in close relationships. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 622–654). New York: Guilford.
- Johnson, D. W., & Johnson, R. T. (1994). Constructive conflict in the schools. *Journal of Social Issues*, 50(1), 117–138.
- Keashly, L. (1997). Conflict and conflict management. In S. W. Sadava, & D. R. McCreary (Eds.), *Applied social psychology* (pp. 248–273). Upper Saddle River, NJ: Prentice Hall.
- Kelley, H. H. (1979). *Personal relationships: Their structure and processes*. Hillsdale, NJ: Lawrence Erlbaum.
- Kressel, K., & Pruitt, D. G., & Associates (1989). *Mediation research*. San Francisco: Jossey-Bass.
- Lewin, K. (1948). The background of conflict in marriage. In G. W. Lewin (Ed.), *Resolving social conflicts: Selected papers in group dynamics* (pp. 84–102). New York: Harper.
- Pruitt, D. G. (1998). Social conflict. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 470–503). New York: McGraw-Hill.
- Rusbult, C. E., & Van Lange, P. A. M. (1996). Interdependence processes. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 564–596). New York: Guilford.
- Wilmot, W. W., & Hocker, J. L. (2001). *Interpersonal conflict*. New York: McGraw-Hill.



# Interpersonal Perception

David A. Kenny

University of Connecticut, Storrs, Connecticut, USA

1. Types of Perceptions
  2. Relations among Perceptions
  3. Statistical Analysis
- Further Reading

## GLOSSARY

**meta-accuracy** How well a person knows how she or he is viewed by others.

**metaperceptions** The way in which people believe others perceive them.

**reciprocity** The degree to which people view each other the same (or differently).

**self-perceptions** How individuals perceive themselves.

**self-other agreement** How much a person's self-perception is similar to the way in which others see him or her.

**target accuracy** How well a person really knows another person.

People, whether strangers or marriage partners, form perceptions of one another, and these perceptions have played an important role in social science research. For instance, leadership style was measured by Fiedler using an interpersonal perception measure called LPC, and Newcomb argued that the co-orientation of interpersonal perceptions is key to understanding the formation of friendships. This article considers the types of interpersonal perceptions, the relations among them, and their statistical analysis.

## 1. TYPES OF PERCEPTIONS

There are three major types of perceptions: perceptions of others, metaperceptions, and self-perceptions. Very often, researchers studying groups and relationships ask people to make judgments of one another. For instance, they may ask people in a work group how hard-working and friendly the other members of the group are. Sometimes, researchers ask observers or outsiders to form perceptions of group and dyad members. Thus, much of what is called an interpersonal "perception" is not so much a perception as an inference made about another person. Nonetheless, the term "interpersonal perception" continues to be used.

Besides asking for perceptions of others, researchers may ask people how they think others view them. These perceptions of perceptions are called metaperceptions or reflected appraisals. Finally, they may ask people how they perceive themselves (called self-perceptions). Although self-perceptions are not themselves interpersonal, they are thought to be closely tied to interpersonal perceptions.

In 1966, Laing and colleagues developed a simple notation for such perceptions. The perception that Person A has of Person B is denoted as A(B), the metaperception is denoted as A(B(A)), and the self-perception is denoted as A(A).

Other more complicated types of interpersonal perceptions are relatively less studied. Laing and colleagues discussed meta-metaperceptions (e.g., What does Person A think that Person B thinks that Person A

thinks?) and even meta-meta-metaperceptions. Others have studied triadic perceptions (e.g., What does Person A think that Person B thinks of Person C?).

The content of these perceptions is quite varied and is limited only by the imagination of the investigator. Among the topics studied are personality (e.g., Does Person A think that Person B is friendly?), liking (e.g., Does Person A think that Person B likes Person A?), interests and attitudes (e.g., Does Person A think that Person B is politically conservative?), and emotional states (e.g., Does Person A think that Person B is currently angry?). In general, researchers ask perceivers to make judgments about several variables. For instance, a person might be asked whether his or her roommate is friendly, intelligent, lazy, and/or organized.

## 2. RELATIONS AMONG PERCEPTIONS

Very often, researchers are interested in the relationship among different interpersonal perceptions. In 1994, Kenny, extending the pioneering work of Laing and colleagues, defined a series of questions that relate interpersonal perceptions to each other. Among the most common questions are the following:

- *Consensus*: Do two people agree in their perception of a common target?
- *Reciprocity*: Do two people view each other in the same or opposite ways?
- *Target accuracy*: Does a person really know what another person is actually like?
- *Meta-accuracy*: Does a person know how he or she is viewed by another person?
- *Self-other agreement*: Is a person's self-perception similar to how another person views him or her?

Very often, researchers and theorists use these relations among interpersonal perceptions. For instance, a key tenet of symbolic interactionism is that self-other agreement between a person and a significant other is due to the meta-accuracy of the person knowing how the significant other views him or her and then the incorporation of that view into the self-concept.

## 3. STATISTICAL ANALYSIS

Research in interpersonal perception can be either ideographic or nomothetic. In ideographic work, the

focus is on variation of relationships between perceptions by persons or relationships. An example of an ideographic question is that of whether married couples who know their partners better are more satisfied in their relationships. Most researchers who study interpersonal perception are interested in the ideographic level of analysis.

Alternatively, the focus might be nomothetic or general laws that are true of most persons. For instance, research might ask how much meta-accuracy there is. One advantage of a nomothetic analysis is that it is possible to decompose effects at the level of the individual and the dyad using the Social Relations Model. For instance, one can determine whether meta-accuracy is true of persons across relationships (e.g., Do people know how they are generally seen across relationships?) and whether meta-accuracy is true of particular relationships (e.g., Do people know how they are uniquely seen by particular others?).

There are several difficulties that arise in the analysis of interpersonal perceptions. For instance, researchers sometimes compute the absolute difference or discrepancy between two interpersonal perceptions. Following Cronbach's 1955 work, such scores lead to several interpretive issues. Other issues arise in the analysis of data from interpersonal perception. Despite difficult methodological and statistical issues, interpersonal perception and the relations between perceptions will continue to be important in social science theory and research.

### See Also the Following Articles

Perception and Culture ■ Reciprocity Norm

### Further Reading

- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, 52, 177-193.
- Fiedler, F. E. (1964). A contingency model of leadership effectiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1). New York: Academic Press.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.
- Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin*, 114, 145-161.
- Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142-182). San Diego: Academic Press.

- Kinch, J. W. (1963). A formalized theory of self-concept. *American Journal of Sociology*, 69, 481–486.
- Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and a method of research*. New York: Springer.
- McLeod, J., & Chaffee, S. H. (1966). Interpersonal approaches to communication research. *American Behavioral Scientist*, 16, 467–500.
- Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological Review*, 60, 393–404.



# Interrogation and Interviewing

Aldert Vrij

University of Portsmouth, Portsmouth, United Kingdom

1. Introduction
  2. The Problem: The Interviewee Does Not Provide Enough Information
  3. The Problem with Further Questioning: Suggestibility
  4. The Solution: Questioning Techniques to Minimize Suggestibility
  5. The Solution: Interview Techniques to Enhance Information Recall—The Cognitive Interview
  6. The Problem: Persuading Reluctant Interviewees to Talk
  7. False Confessions
  8. Detecting Deceit
  9. Common Pitfalls in Detecting Deceit
  10. Techniques to Improve People's Ability to Detect Deceit
  11. Conclusion
- Further Reading

## GLOSSARY

**cognitive interview** Forensic tool composed of a series of memory retrieval techniques designed to increase the amount of information that can be obtained from a witness.

**ethical interviewing** A police interview technique, based on three principles, designed to encourage a suspect to tell his or her side of the story; the use of deceit is forbidden, and psychological pressure is not used.

**Reid technique** A police interrogation technique, consisting of nine steps, designed to encourage reluctant suspects to talk during police interviews; the technique allows interrogators to use deceit and to build up psychological pressure.

**suggestibility** The act or process of impressing something (e.g., an idea, an attitude, a desired action) on the mind of another.

The purpose of a police interview or interrogation is to obtain information about a possible crime. Effective interviewing is difficult because interviewers risk eliciting inaccurate information that they may misconstrue as accurate.

## 1. INTRODUCTION

The purpose of a police interview with an alleged victim, witness, or suspect is to obtain information about a possible crime. During this information-gathering process, interviewers should attempt to obtain statements that are complete (i.e., include all the relevant information that the interviewee knows) and reliable (i.e., contain only accurate information). Those two aims are not always compatible. Certain interview techniques readily result in more information, but the information is often less reliable. Psychologists have developed interview techniques that increase the likelihood of obtaining both accurate and reliable information. However, these techniques might not work with interviewees who are reluctant to talk. For this reason, other more persuasive interview techniques have been developed, especially for interviewing uncooperative suspects. Those techniques have been criticized extensively by psychologists. A major problem is that they may induce false confessions, that is, suspects admitting to having participated in crimes when in fact they did not actually participate in those crimes. Obtaining unreliable information would be less problematic if

police officers could distinguish reliable information from unreliable information.

## **2. THE PROBLEM: THE INTERVIEWEE DOES NOT PROVIDE ENOUGH INFORMATION**

Obtaining information from witnesses, victims, and suspects is critical for effective police investigations. However, the answers to questions such as “Tell me exactly, and in as much detail as possible, what you witnessed” and “Tell me exactly, and in as much detail as possible, what happened to you” are typically incomplete. In nearly all cases, interviewees initially report less information than they actually hold in their memories. Routinely, interviewers want more information than is initially provided. It is not surprising that guilty suspects fail to report all of the information that they are able to recall. Why would they volunteer information that will indisputably link them with their crimes? However, witnesses and victims also rarely volunteer all of the information they know. There are several reasons for this. First, it is a matter of ignorance. Police interviews require witnesses and victims to describe people, objects, and events in more detail than is customary in casual conversation. Often, interviewees do not realize that the police want them to disclose all of the information they know rather than just the information that the interviewees believe is relevant. Second, information is often lacking or incomplete due to cognitive reasons. For example, young children’s first accounts of an event are particularly incomplete. Cognitive abilities and command of language develop throughout childhood, making it gradually more feasible to offer detailed accounts of what was witnessed. Older children and adults use more complex and successful retrieval strategies than do younger children, increasing the amount of information they can retrieve independently. People cannot verbally recall experiences that occurred before they learned to speak. Therefore, they cannot verbally relate events that they experienced before 2 years of age. Third, social factors contribute to incomplete accounts. For example, socially anxious people who feel discomfort in the presence of strangers spontaneously volunteer less information. Other people may be too embarrassed to talk freely about certain events that they have experienced or witnessed. For all of these reasons, interviewers must ask further specific questions to learn more about events. However, by asking a series of specific follow-up

questions, interviewers run the risk that the questions selected will reflect what the interviewers believe happened rather than what the interviewees actually witnessed or experienced. This risk is most pronounced when the interviewers’ questions are suggestive.

## **3. THE PROBLEM WITH FURTHER QUESTIONING: SUGGESTIBILITY**

Despite considerable controversy concerning the extent to which children and adults are susceptible to suggestion in forensic contexts, three conclusions can be drawn. First, there are individual differences in suggestibility. Second, there are age differences in suggestibility. Younger children are more suggestible than older children, and children are more suggestible than adults. Third, the manner in which an interview is conducted can affect suggestibility as well as the quality and extent of information elicited.

## **4. THE SOLUTION: QUESTIONING TECHNIQUES TO MINIMIZE SUGGESTIBILITY**

Research on questioning witnesses and victims in a forensic context has focused on ways in which to avoid suggestibility. Perhaps because children are particularly sensitive to suggestibility, much research has centered on interviewing children. This research shows that open-ended questions (e.g., “Tell me what happened”) are best for gathering information because the interviewee has the opportunity to give an open unrestricted answer and to control the flow of information. This type of question produces responses that are longer and richer in relevant details than are responses to other types of questions. Most “wh-” questions (e.g., what, where, when, why, who) can be classified as open-ended questions. These questions should be asked at the beginning of the interview. Facilitative responses by the interviewer, such as “okay” and “mm,” encourage the witness to continue the account. Focused questions can be asked later in the interview. These are questions that focus the witness’s attention on details or aspects of the event that the witness described previously. For example, “Tell me what the man looked like” is an example of a focused question after an interviewee mentioned a man but did not describe his appearance. Leading questions (e.g., “Was the man Black?”) are problematic because the interviewee may respond by

simply repeating the information conveyed in the question rather than by relying on his or her memory. Option-posing questions, or questions that offer alternatives (e.g., “Was the man Black or White?”), are preferable in that respect because they are less leading. However, there is still the chance that the interviewee’s preferred option might not be incorporated into the alternatives posed by the interviewer. An open-ended question (e.g., “What was the color of the man’s skin?”) is a less suggestive question format. Where leading questions are asked (e.g., “Did he touch you?”), they should be followed by open-ended questions designed to elicit further information about the topic (e.g., “Tell me everything about that”).

Few practitioners apply these principles in forensic interviews. For example, in 2002, Westcott and Brace reported an analysis of 119 interviews with children by police officers and social workers in England and Wales. The analysis revealed that open questions were rarely asked (6%) and that focused questions were most common (47%). Questions that included alternatives occurred more frequently (29%) than did facilitative responses (13%), and some questions were leading (5%). Similar studies conducted in Sweden, Israel, and the United States provided similar results.

### **5. THE SOLUTION: INTERVIEW TECHNIQUES TO ENHANCE INFORMATION RECALL—THE COGNITIVE INTERVIEW**

Another line of research has explored whether the quality and amount of information provided by cooperative interviewees can be improved by certain interview techniques. One such technique, the cognitive interview, has been adopted by police forces around the world, including Australia, Germany, the United Kingdom, and the United States.

The cognitive interview, developed by psychologists, has three core elements: social dynamics, memory/general cognition, and communication. Social dynamics and communication refer to interpersonal communication and ways in which to deal with anxious or inarticulate witnesses. The interviewing police officer has ostensible social authority and more control over the interview than does the witness and, therefore, may dominate the interview. To avoid counterproductive dynamics, when interviewing an cooperative witness, the police officer should explicitly ask the witness to take an active role in the interview.

Officers should communicate with witnesses in the style that the witnesses prefer by adopting witness-compatible questions. Most often, depictions will be verbal, but some witnesses find it easier to describe certain events nonverbally. For example, if the event that a witness experienced involved a tactile element such as brushing against a fabric, the witness might prefer to recall this experience in a similar mode by touching various fabrics.

The memory/general cognition element consists of four components. A “report everything instruction” is given to encourage witnesses to report everything they remember without any editing, even if the details seem trivial or their memory of a particular aspect of the event is partial or incomplete. Memory retrieval is most efficient when the context of the original event is mentally reinstated at the time of recall. Therefore, interviewees are instructed to reconstruct in their minds the environmental features (i.e., location where the event took place) and personal features (i.e., how they felt at the time) of the event. Different retrieval cues may access different aspects of a complex event. Therefore, after witnesses recount the event in their own order during free recall, they should be asked to recall the event using a variety of different temporal orders, for example, in reverse order (i.e., from the end of the event to the beginning of the event). People have a tendency to report events from their own perspective. An instruction to change perspective asks interviewees to recall the event from the perspective of another person who was present (e.g., “Go through the event again and tell me any further information witnessed by that other person you mentioned”).

Studies testing the efficacy of the cognitive interview have revealed that this technique elicits more accurate information than do conventional interview techniques. This technique also elicits more inaccurate information in that it increases the absolute number of errors; however, it does not increase the rate of inaccuracy. In other words, the percentage of information that is inaccurate is not higher using the cognitive interview than that obtained using conventional interview techniques.

### **6. THE PROBLEM: PERSUADING RELUCTANT INTERVIEWEES TO TALK**

The stereotypical view, often promoted in police manuals, is that suspects are reluctant to talk and that, therefore, tough interrogation techniques, rather than

interview techniques, are needed to persuade suspects to talk. The assumption that suspects are unwilling to talk is plausible because confessing to a crime dramatically increases the likelihood of a conviction. For example, Dutch research revealed that in cases where the suspects confessed, convictions were nearly inevitable (a 94% conviction rate was found), whereas without a confession, the conviction rate was only 44%. However, the view that suspects are generally uncooperative is not supported by research. Studies in England and Wales revealed that 80% of suspects are cooperative during interviews and that 55 to 60% of suspects confess. (Much research on this topic has been carried out in England and Wales because, by law in those countries, all police interviews with suspects must be audiotaped.) These studies also examined what prompts suspects to confess. The most important factor seems to be the perceived strength of evidence; suspects are more likely to confess if they perceive the evidence against them to be strong. Another influential factor is the seriousness of the crime; suspects are less inclined to confess when they are suspected of committing more serious offenses. Hence, suspects are reluctant to confess when they are suspected of serious crimes and/or the evidence against them is weak.

Several interrogation methods have been proposed to persuade reluctant suspects to talk. Probably the most well known is the “Reid technique” proposed by Inbau and colleagues. The back cover of the latest edition of their manual, published in 2001, proclaims that their technique is “the standard in the field” and that it has been used successfully by “thousands of criminal investigators.” The Reid technique is meant to apply to suspects who are likely to be guilty. Interviewers are advised to make preliminary assessment of the likelihood of guilt by observing suspects’ verbal and nonverbal responses. For example, according to Inbau and colleagues, verbal responses typical of untruthful suspects include giving delayed, evasive, and/or vague answers; displaying unusually poor memory; making statements against self-interest (e.g., “as crazy as it sounds”); and answering questions with questions (e.g., “Why would I do that?”). Assumed behavioral clues to deception include changing posture, maintaining a retreating posture (e.g., crossing arms in a tight fashion), making grooming gestures, placing a hand over the mouth or eyes when speaking, averting one’s gaze, and hiding hands (by sitting on them) or hiding feet (by pulling them under the chair).

### **6.1. The Reid Nine Steps of Interrogation**

The Reid technique strives to break the (guilty) suspect’s resistance to confess. It consists of the following nine steps:

1. Engaging in a positive confrontation
2. Developing a theme
3. Handling denials
4. Overcoming objections
5. Retaining the subject’s attention
6. Handling the suspect’s mood
7. Creating an opportunity to confess
8. Obtaining an oral confession
9. Converting an oral confession into a written one

Step 1 consists of a direct presentation of real or fictional evidence regarding the suspect’s involvement in the alleged crime (e.g., “Our investigation shows that you are the one who . . .”). This accusation is followed by a brief pause during which the suspect’s behavior is observed closely. The interrogator repeats the accusation. Suspects who fail to make a denial after the second direct confrontation are presumed to be guilty.

Step 2 deals with theme development tailored to emotional versus nonemotional suspects. In the case of an emotional suspect, the interrogator is advised to build rapport by offering the suspect a moral excuse for committing the offense. This can be achieved by rationalizing the situation, suggesting that anyone else in the same situation might have committed the same type of offense, minimizing the moral seriousness of the offense, suggesting a more morally acceptable motivation for the offence, condemning others, and so on.

Suspects who appear to listen attentively to these suggestions are presumed to be guilty. The following example, from Inbau and colleagues, illustrates this technique: “Joe, no woman should be on the street alone at night looking as sexy as she did. . . . It’s too much a temptation for any normal man. If she hadn’t gone around dressed like that, you wouldn’t be in this room now.” Critics of the technique label it the minimization approach and describe it as a technique in which the police interrogator tries to lull the suspect into a false sense of security by offering sympathy.

Nonemotional suspects are expected to perceive an interrogation as a “contest of endurance,” that is, as “pitting their willpower against the interrogator’s persistence.” Effective techniques for these cases are (a) to seek an admission of lying about some incidental aspect of the crime so that the interrogator can say

“You lied to me once and you will lie to me again” and (b) to encourage the suspect to associate himself or herself with the scene of the crime, pointing out that all of the evidence leads to the suspect’s guilt, that it is futile to resist telling the truth, and so on. Critics call this the maximization approach, a technique in which the interrogator tries to scare and intimidate the suspect into confessing by exaggerating the seriousness of the offense and the magnitude of the charges and, if necessary, by making false claims about the evidence.

Step 3 describes how to respond to denials of guilt. The goal is to curtail the suspect’s repetition or elaboration of the denial (using techniques described in Step 2) and is seen as a crucial step because the more frequently a guilty suspect repeats a lie, the harder it becomes for the interrogator to persuade the suspect to tell the truth. The theory is that an innocent suspect typically will persist in denying guilt, whereas a guilty suspect is more likely to allow the interrogator to return to the conversation theme. An innocent suspect is expected to maintain denial, and a guilty suspect is expected to offer reasons why the accusation is wrong (e.g., “I couldn’t have done that; I don’t own a gun”).

Step 4 consists of overcoming the suspect’s objections. For example, the interrogator is advised to show understanding and then return to the conversation theme (e.g., “That may be true, but the important thing is . . .”). Once the suspect observes that his or her objections are ineffective, the suspect is expected to become uncertain and display signs of withdrawal. The interrogator is advised to react by retaining the suspect’s attention (Step 5), for example, by moving physically closer to the suspect, leaning towards him or her, and maintaining eye contact. Step 6 involves strategies to manage the suspect’s passive mood. When the suspect appears to be attentive, the interrogator is directed to focus the suspect’s mind on possible reasons for committing the crime. The interrogator should exhibit signs of understanding and sympathy and should urge the suspect to tell the truth. Finally, the interrogator should attempt to elicit a remorseful mood, for example, by pointing out the negative consequences of the crime for the victim. In Step 7, alternative question forms are recommended to offer the suspect the opportunity to provide an explanation or excuse for the crime, and this will make an admission easier to achieve (e.g., “Was this your own idea or did someone talk you into it?”). Whichever alternative a suspect chooses, the net effect of an expressed choice is an admission. Following the selection of an alternative, Step 8 consists of the development of the initial

admission into a detailed confession that discloses the circumstances, motives, and details of the crime. Inbau and colleagues advised that the questions asked by the interrogator should be brief, clear, and designed so that they can be answered in a few words. Furthermore, they should not contain harsh or emotionally charged terminology. Step 9 deals with converting an oral confession into a written one. This step is seen as important because the suspect might deny ever making an oral confession.

## 6.2. Concerns with the Reid Technique

The Reid technique has been criticized extensively by academic researchers. Numerous assumptions presented in the manual, such as “two denials imply guilt” (Step 1), have never been tested empirically. Other assumptions, such as how liars behave and what they say, have been tested but not supported.

By asking (during the confession phase, i.e., Step 8) leading questions that can be answered by the suspect in a few words, the interrogator runs the risk that the confession will reflect more about what the interrogator believes happened than what actually transpired.

In addition, bluffing (i.e., making up evidence) may backfire. For example, if an interviewer lies and tells a burglar that his or her fingerprints have been found at the scene of the crime, a burglar who wore gloves now knows that this statement is a lie and will probably realize that the police do not actually have much evidence against him or her. Inbau and colleagues acknowledged this problem and recommended that this tactic be used only as a “last resort effort.” Researchers have demonstrated that North American police detectives can be very creative in devising examples of bluffing such as introducing nonexistent technology (e.g., “The Cobalt Blue test has proven that the fingerprints on the body of the victim belong to you”). Research has revealed that confronting a suspect with false evidence is quite common in the United States and happens in approximately 30% of custodial interviews. Although these sorts of tricks are allowed in the United States, they are illegal in several Western European countries. Police detectives who try to apply the Reid technique in those countries run the risk that subsequent confessions will be dismissed in criminal courts.

The Reid technique of psychologically manipulating people through trickery and deceit may lead to false confessions. One problem with tricks and deceit is that

both guilty and innocent suspects become more willing to confess. There is no strategy that motivates only guilty suspects to talk. As a result, the more successful a strategy is in eliciting confessions from guilty suspects, the more likely it is that this strategy will also produce false confessions from innocent suspects.

### **6.3. The Solution: Ethical Interviewing**

In the United Kingdom, an ethical framework of police interviewing is advocated. The British outcry for, and subsequent legislation of, less oppressive police investigative interview techniques has probably been heavily influenced by cases such as “The Guildford Four” and “The Birmingham Six,” well-known cases of British miscarriages of justice. Ethical interviewing is based on three principles: (a) to shift the police service from its traditional reliance in getting a suspect to confess to encourage a search for the truth, (b) to encourage officers to approach an investigation with an open mind, and (c) to encourage officers to be fair. Police training in this ethical framework emphasizes the importance of detailed information gathering at the beginning of the police investigation. It is argued that before the interview starts, the detectives should have been to the scene of crime and, therefore, know how it looks; should know all of the facts of the crime (e.g., evidence obtained at the scene of the crime, statements of witnesses, any peculiarities); and should know the suspect (e.g., background characteristics, family circumstances, possible addictions, possible diseases).

It is further stressed how to use open and closed questions in police interviews. An information-gathering strategy at the beginning of the interview is advocated because this increases the possibility of eliciting an account from the suspect. The use of open questions is generally preferable. They usually elicit longer answers and, therefore, more information. They are an invitation to the suspect to present his or her point of view and will increase the likelihood of the suspect believing that the interviewer is taking him or her seriously. Open questions encourage the suspect to talk and, therefore, facilitate the desired format of a police interview; that is, the suspect talks and the interviewer listens and asks for clarifications. An interviewer who is prepared to listen is more likely to be liked by the suspect, and this might make the suspect more willing to talk. Closed questions are also useful in interviews, particularly to obtain short factual answers

on specific points. However, these questions should be used sparingly at other times and rarely during the early stages of the interview.

Several strategic tactics are offered, including the “blocking escape routes” method, which is used to strengthen the evidence in a case. Evidence is sometimes thin and open to multiple interpretations. Presenting this evidence at an early stage of the interview may give the suspect the opportunity to “escape” by providing alternative explanations. The blocking escape routes method is designed to prevent this from happening. For example, suppose that the suspect’s car was noticed near the scene of the crime just after the crime took place. This might be a link, but the link is not strong. If the suspect is confronted with this piece of evidence at an early stage of the interview, the suspect might, for example, say that he or she used the car to go to a store. The evidence is much stronger if, before the evidence is presented, the suspect has told the interviewer (after being asked about this) that he or she did not use the car on that particular day, that he or she never lent the car to anyone else, and that nobody else has the keys to the car. After the escape routes have been blocked, the suspect could be confronted with his or her own statements and evidence (e.g., “You told me that you are the only one who uses that car, right? Well, a high-speed camera provided evidence that you drove with your car at high speed near the scene of the crime just after the crime took place. Could you explain that?”).

At first sight, the ethical interviewing approach appears to be “softer” than the Reid technique; however, this is not necessarily the case. Although the use of trickery and deceit is forbidden in an ethical interviewing framework, a “tough” interview is allowed. It might be that suspects who decided not to talk during their interviews will remain silent without the use of trickery and deceit, and in such cases the ethical approach will not be successful for the interviewing of uncooperative suspects. This poses a problem, especially if these uncooperative suspects are believed to have committed serious crimes. Ethical methods designed to make silent suspects more willing to cooperate have yet to be developed. The difficulty is that any such methods are likely to make both guilty and innocent suspects more willing to cooperate, and the latter could lead to false confessions. Inbau and colleagues acknowledged the problem of false confessions and offered the following solution:

A guideline that an interrogator may use in any case situation where he may be in doubt as to the

permissibility of any particular type of trickery and deceit, is to ask himself the following question: "Is what I am about to do, or say, apt to make an innocent person confess?" If the answer is "no," the interrogator should go ahead and do or say what was contemplated. On the other hand, if the answer is "yes," the interrogator should refrain from doing or saying what he had in mind.

The obvious problem, not addressed by Inbau and colleagues, is that the police detective never knows for sure whether he or she made a correct judgment.

## 7. FALSE CONFESSIONS

How often false confessions occur is unknown, but estimates range from 35 to 600 convictions on the basis of false confessions per year in the United States alone. It is often difficult to establish whether a confession is false. Police coercion during the interview is not determinative because a confession obtained using an oppressive interview style could be truthful. The retraction of a confession is another potential indicator. On the one hand, a confession may be truthful even though the suspect retracts the confession after the interview; on the other hand, suspects convicted on the basis of confessions do not necessarily appeal their convictions but still could have confessed falsely. Finally, one might consider the accuracy of the confession. On the one hand, a confession in which accuracy cannot be established beyond doubt (e.g., because independent case facts are not available) is not necessarily false; on the other hand, innocent suspects may at times give accurate accounts of what has happened because police detectives provided case information to the suspects, who subsequently incorporated the information into their statements.

Why suspects make false confessions is difficult to determine, but several factors are often present, for example, a combination of tricks employed by the police, a susceptible suspect, and interviewers determined to secure a confession.

### 7.1. Types of False Confessions

Psychologists distinguish among three psychologically distinct types of false confessions: voluntary false confessions, coerced compliant false confessions, and coerced internalized false confessions.

#### 7.1.1. Voluntary False Confessions

Voluntary false confessions are given without any external pressure from the police. Most often, people voluntarily go to the police station and inform the police that they have committed crimes reported on television or in newspapers. Voluntary false confessions occur for several reasons. One reason is a "morbid desire for notoriety," that is, a pathological need to become infamous and to enhance self-esteem, even if it entails the prospect of imprisonment. One serial false confessor, Henry Lee Lucas, is estimated to have confessed to more than 600 murders during the early 1980s. When asked why he confessed to all of these crimes, Lucas answered that prior to his arrests, he had no friends and nobody listened to him or took any interest in him; however, once he began to make false confessions, all that changed and he now had many friends and enjoyed his celebrity status. Other people may voluntarily falsely confess because they are unable to distinguish facts from fantasy. Schizophrenic people, in particular, are prone to this type of confession. Yet another reason why people may voluntarily falsely confess is that they want to protect the real criminals. For example, a father might falsely confess to a crime that he knows or believes was actually committed by his son. Finally, people may voluntarily falsely confess because they see no possible way in which to prove their innocence and hope that confessing will secure reduced punishment.

#### 7.1.2. Coerced Compliant Confessions

Coerced compliant confessions occur when, as a result of police interviews, suspects confess to acts they know are untrue. This type of confession results from the pressures of police interviews and from the strategies used by the police in the interviews. As a result, suspects believe that the benefits of confessing outweigh the costs, and so they confess to escape from the police interviews, which they consider stressful and intolerable. Alternatively, they confess due to police promises of rewards following their confessions, or they confess due to a combination of the two.

#### 7.1.3. Coerced Internalized False Confessions

Coerced internalized false confessions occur when people come to believe, during police interviewing, that they have committed the crimes of which they are accused, even though they have no actual memories of committing the crimes. One vivid example is

the case of Tom Sawyer, a socially anxious man and former alcoholic who came to believe that he committed a murder that he in fact never committed. In trying to engage Sawyer in conversation about the crime, the detectives asked him to help them create a scenario describing how the murder might have happened. Sawyer, who loved to watch detective shows on television, was eager to help and joined in. The police let Sawyer outline several scenarios and accused him of having committed the murder. The police claimed that Sawyer knew nine facts that only the killer could have known. Analysis of the interrogation transcripts showed that all of the crucial information was introduced into the interrogation by the police. Following the accusation, Sawyer strongly denied his guilt. The police obtained fingerprint and hair samples from Sawyer and suggested a polygraph examination. Sawyer believed that the polygraph would prove his innocence and agreed to the examination. After the test, the examiner told Sawyer that the test proved he was lying. (Subsequent rescoring of the test by a polygraph expert revealed that the test outcome was inconclusive.) Once told that the polygraph showed him to be lying, Sawyer's confidence began to erode. He was no longer able to express firm denials of guilt, and he could say only that he still did not believe he had done it. His main defense against the police demands to confess was his lack of memory of having committed the crime. The police replied that he was blocking his memory of the crime and that he had a blackout, just as he often did when he had been drinking. Sawyer still refused to fully accept that he had committed the crime, and he hoped that the fingerprint and hair tests would prove his innocence. The detectives decided to lie and told Sawyer that his hair samples matched hairs found on the victim's body. On receipt of this information, Sawyer's resistance collapsed, and he agreed that "all the evidence was in" and that he must have committed the crime.

The fact that police deception leads to false confessions might appear to be surprising. However, police suspects are often not average people. Typically, their intelligence quotients (IQs) are lower than average. A sample of 160 suspects in England had an average IQ of 82. People with lower IQs are more suggestible. Sawyer was also socially anxious and a former alcoholic. The combination of police deception and the vulnerability of the suspect may have led to the false confession.

Internalized false confessions do not imply that suspects become convinced that they have committed crimes. They occur because suspects form opinions

that it is more likely than not that they are guilty. They typically have no memory of committing the crimes, but tactics used by the police diminish the suspects' confidence in their memories, make them less certain about their innocence, and make them consider whether it is possible that they had committed the crimes.

## **7.2. How to Detect False Confessions**

The best way in which to distinguish between false and true confessions is to look for intimate knowledge and impossibilities. Intimate knowledge is present when a suspect provides information that only the guilty person could know (e.g., information about the location of a disposed body). Impossibilities refer to providing evidence that demonstrates conclusively that the person could not have committed the crime. For example, Lucas's confession could not be true because he was more than 1,300 miles from the scene of the crime on the day of the murder. The psychologist who conducted a psychological evaluation of Lucas considered the uncritical willingness to accept his confession as the most fundamental mistake made by the police. However, police officers claim that it is the responsibility of the courts, and not the police, to check the veracity of a confession. Courts tend to attribute impossibilities to errors in confessions rather than considering them as evidence of false confessions.

Obtaining unreliable information, such as false confessions and incorrect answers to leading questions, would be less problematic if police or jurors could distinguish reliable statements from unreliable ones. A distinction should be made between unreliable (inaccurate) information, which is given deliberately and which the presenter himself or herself knows is inaccurate, and inaccurate information, which is given unwittingly (e.g., because the person is mistaken or no longer knows for certain what happened). Virtually no research has been carried out examining whether people can detect these latter fortuitous truth distortions. The few findings that do exist are not encouraging. Considerable research efforts have examined whether observers can distinguish between truths and deliberate truth distortions (i.e., lies).

## **8. DETECTING DECEIT**

Research has demonstrated convincingly that catching liars is a difficult task and that people frequently

err when they attempt to detect deceit. In scientific studies on the detection of deception, observers are typically given videotaped statements by a number of people who are either lying or telling the truth. After each statement, observers (often college undergraduates) are asked to judge whether the statement is truthful or false. In such tasks, guessing whether someone is lying has a 50% chance of being correct. A review of 37 lie detection studies, in which undergraduates attempted to detect truths and lies told by people (adults) with whom they were unfamiliar, revealed a total accuracy rate (percentage of correct answers) of 56.6%, barely above the level of chance. People are shown to be capable of detecting truths to some extent (i.e., 67% accuracy in correctly judging that someone is telling the truth) but are particularly poor at detecting lies (i.e., 44% accuracy in correctly judging that someone is lying). In fact, 44% is below the level of chance. In other words, people would be more accurate at detecting lies if they simply guessed. The superior accuracy rate for truthful messages is the consequence of the truth bias. People are more likely to expect messages to be truthful than to be deceptive, and as a result, truthful messages are identified with more accuracy than are deceptive ones.

In other studies, people attempted to detect children's truths and lies using the same methodology except that the liars and truth tellers were children rather than adults. That research revealed that detecting children's lies is difficult as well, with accuracy rates typically in the 50 to 65% range.

Undergraduates are not habitually called on to detect deception, whereas police officers and customs officers have more experience in interviewing people. However, studies in which professionals such as police officers attempted to detect truths and lies told by adults revealed accuracy rates similar to those found in studies with undergraduates, mostly in the 45 to 60% range. This research suggests that professionals are no better than laypersons at detecting deceit. However, some professionals (e.g., members of the Secret Service, U.S. federal officers with experience and a special interest in deception and demeanor, sheriffs identified by their departments as outstanding interrogators) were better than others. Professionals are less susceptible to the truth bias than are lay observers, possibly because their jobs make them more suspicious and, hence, more likely to conclude that a person is lying.

## 9. COMMON PITFALLS IN DETECTING DECEIT

There are numerous reasons why people are generally poor at detecting deceit. Above all, lie detection is a notoriously difficult task because there are no typical behaviors associated with deception. In other words, a cue such as Pinocchio's growing nose does not exist. Despite this, people, laypersons, and professionals have strong views about how liars behave. Interestingly, their views do not differ substantially. In particular, people believe that liars display clear signs of nervousness such as averting one's gaze and making grooming gestures. In reality, however, gaze behavior and grooming gestures are unrelated to deception. If there is a relationship between movements and deception, it is more likely to be the converse, that is, that liars tend to decrease their movements rather than increase them. There are several explanations for this. First, it might be the result of cognitive load. Sometimes, liars find it difficult to lie because they must think of plausible answers, avoid contradicting themselves, and tell lies that are consistent with everything the observers know or might find out. Moreover, liars must remember what they have said so that they can say the same things again when asked to repeat their stories. This might be more difficult than truth telling, especially when liars are unprepared and must make up their stories instantly. Increased cognitive load results in a neglect of body language, reducing overall animation. Second, liars might worry that several cues will give their lies away and, therefore, try to suppress such signs and to make convincing impressions. However, it is not easy to suppress nervousness effectively, to mask evidence of concentration, to know how they normally respond so as to make honest and convincing impressions, and to display the intended responses. It may well be the case that, when trying to appear truthful, liars "overcontrol" their behavior and exhibit a pattern of behaving that appears planned, rehearsed, and lacking in spontaneity. For example, liars may believe that making lots of movements will give their lies away and, therefore, will move very deliberately and tend to avoid any movements that are not strictly essential. This will result in an unusual degree of rigidity and inhibition because people do normally make movements that are not essential. Another possible cue that may result from inadequate control of behavior is that performances might look flat due to a lack of involvement. For example, someone who applies for a job because the money appeals to him or her might not be enthusiastic when the

person talks about the job during the selection interview, and this in turn might result in a lack of animation.

Police officers' incorrect beliefs about cues to deception, such as liars' gaze aversion and grooming gestures, are particularly harmful for suspects who exhibit these patterns of behavior naturally. There is evidence that people from Western cultures exhibit this pattern of behavior less frequently than do people from other ethnic origins such as Black/African Americans and people from Mediterranean countries (e.g., Turkey, Morocco). Some behaviors are culturally defined. For example, looking an authority figure straight in the eyes is considered to be impolite in many cultures, but the opposite is true for Caucasian people living in, for example, North America or Western Europe. For them, averting gaze from a conversation partner is suspicious. Therefore, when a White police officer interviews a suspect from a different ethnic background, cross-cultural nonverbal communication errors may arise. Specifically, natural truthful behavior displayed by a member of a different ethnic group (e.g., gaze aversion) may be incorrectly interpreted as a sign of deceit.

Another pitfall in lie detection is that experiencing and displaying negative emotions is not the exclusive domain of liars. Truth tellers might experience similar negative emotions, albeit for different reasons. For example, lying suspects might be nervous during police investigations because they fear that they will be caught in their lies. Innocent suspects might also be nervous because they fear that they will not be believed by the police detectives when they maintain their innocence. Therefore, lie detectors should be careful in interpreting signs of fear as signs of deceit, and they should avoid making the mistake committed by Othello in Shakespeare's play. Othello falsely accused Desdemona (his wife) of infidelity. He told her to confess because he was going to kill her for her treachery. Desdemona asked for Cassio (her alleged lover) to be called so that he could testify to her innocence. Othello told her that he had already murdered Cassio. Realizing that she could not prove her innocence, Desdemona reacted with an emotional outburst. Othello misinterpreted this outburst as a sign of her infidelity.

## **10. TECHNIQUES TO IMPROVE PEOPLE'S ABILITY TO DETECT DECEIT**

To date, a truly accurate lie detection technique has yet not been developed. All of the proposed techniques

produce frequent errors. Police training manuals often emphasize that lie detectors should pay attention to someone's behavior to detect deceit because people are less able to control their behavior than to control their speech. Researchers expressed a similar view for many years. However, more detailed analyses of people's speech, starting in the late 1980s, revealed that there are many more speech-related cues to deceit than was thought previously. For example, compared with stories told by liars, truth tellers' stories sound more plausible and less ambivalent. Liars also tend to tell their stories in chronological order (this happened first, and then this, and then that, etc.), whereas truth tellers sometimes tend to give their accounts in unstructured ways, particularly when talking about emotional events. Sometimes, when people are clearly upset, they report what happened in chaotic and incomprehensible ways. In fact, their stories can be so incomprehensible that listeners must ask them to sit down for a while, to calm down, and to describe again exactly what happened, starting from the beginning. This unstructured production effect disappears when people have already told their stories a couple of times, or when they have frequently thought about the events, because this results in the telling of stories in a more chronological order. Moreover, truth tellers include more visual details (i.e., describe things they saw) and more auditory details (i.e., describe things they heard) in their accounts. Often, they also tend to repeat what was said literally (e.g., "Then he said, 'How are your parents?'" rather than "Then he asked me about my parents"). Moreover, truth tellers mention more spatial details (e.g., "He stood behind me," "The book was on the table under the window," "We were in the living room") and more temporal details (e.g., "He switched on the television first and then the video recorder," "The phone rang again approximately 15 minutes later," "He stayed with me for about 3 hours"). Research has demonstrated that observers who concentrate on these speech-related cues typically make better lie detectors than do observers who do not attend to these cues.

Another way in which to improve people's ability to detect deceit is to use a so-called baseline method. There are large individual differences in people's speech and behavior. Some people typically make many movements, whereas others do not. Some people are very talkative, whereas others are not. Because of these individual differences, a given response is often difficult to interpret given that it leaves the observer with a dilemma: Is a particular response a sign of deceit, or is it someone's natural truthful response? Therefore, it might facilitate

lie detection if someone's natural truthful speech style and behavior are known to the observer so that the behavior and speech style under investigation can be compared with their natural truthful speech style and behavior. A deviation from the natural truthful response might indicate that a person is lying. Research has shown that this baseline method works and that people are better lie detectors when they know somebody's natural truthful behavior. However, employing this method is more difficult than people might think initially. Not only do different people behave differently in the same situation (i.e., interindividual differences), but the same person behaves differently in different situations as well (i.e., intraindividual differences). Therefore, in applying the baseline method, it is crucial that only parts of the interview that are truly comparable be compared. Comparing the wrong parts of an interview is a common practice in police interviews, and unfortunately, police manuals sometimes advise police detectives to make comparisons that are inappropriate. Typically, comparisons are made between responses during "small talk" at the beginning of the interview and responses during the actual interrogation. This is an incorrect way in which to employ the technique because small talk and the investigative core of the police interview are fundamentally different. The formal interview matters to suspects because severe negative consequences may follow if they are not believed. In other words, this is a high-stakes situation. The small talk has no such consequences and, therefore, is a low-stakes situation. People tend to behave differently in low-stakes situations than they do in high-stakes situations. Both guilty and innocent suspects tend to change their behavior the moment their formal interviews start. For comparisons to be truly comparable, they must be made within formal interviews. For example, a suspect could be asked to describe in detail his or her activities during a particular day (i.e., the day the crime was committed), and the interviewer could observe the suspect's reactions while he or she gives the description. However, even when correct comparisons are made, the lie detector should be cautious about drawing conclusions. A difference in response between two comparable parts of the interview does not indicate conclusively that someone is lying; it only indicates that this is a possibility. Other explanations for the change in response should be taken into consideration as well. For example, there is some evidence that some behaviors are topic related. ("Topic" is not synonymous with "stakes." During a high-stakes interview, different

topics can be discussed.) Analyses of the behavior by Saddam Hussein when he was interviewed by Peter Arnett during the first Gulf War revealed that the supportive arm and hand movements Hussein made were, to some extent, related to the topics he discussed. When discussing Israel, he made specific movements with his left forearm. This behavioral pattern emerged only when he was discussing Israel and Zionism. Therefore, the lie detector should consider whether different response patterns shown during a high-stakes interview might be topic related.

## 11. CONCLUSION

People, including professionals, have difficulty in discerning lies. It is unlikely that detection of unintentional truth distortions is easier (although this is rarely investigated). Police investigators should question interviewees—witnesses, alleged victims, and suspects—in ways that decrease the likelihood of obtaining inaccurate information, as outlined in this article.

### See Also the Following Articles

Ethics and Social Responsibility

### Further Reading

- Baldwin, J. (1993). Police interview techniques. *British Journal of Criminology*, 33, 325–352.
- DePaulo, B. M., & Friedman, H. S. (1998). Nonverbal communication. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 3–40). New York: McGraw-Hill.
- DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74–118.
- Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: Norton. (Originally published 1985)
- Fisher, R. P., Brennan, K. H., & McCauley, M. R. (2002). The cognitive interview method to enhance eyewitness recall. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 265–286). Mahwah, NJ: Lawrence Erlbaum.
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. Chichester, UK: Wiley.
- Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, MD: Aspen.

- Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist*, 52, 221–233.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and its implications for professional practice*. Chichester, UK: Wiley.
- Vrij, A. (2003). “We will protect your wife and child, but only if you confess”: Police interrogations in England and The Netherlands. In P. J. van Koppen, & S. D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 55–79). New York: Plenum.
- Westcott, H., & Brace, N. (2002). Psychological factors in witness evidence and identification. In H. Westcott, & N. Brace (Eds.), *Applying psychology* (pp. 117–178). Milton Keynes, UK: Open University Press.
- Williamson, T. (1993). From interrogation to investigative interviewing: Strategic trends in police questioning. *Journal of Community and Applied Social Psychology*, 3, 89–99.



# Intrinsic and Extrinsic Motivation in Sport

Robert J. Vallerand

Université du Québec à Montréal, Montreal, Canada

1. Introduction
  2. Definitions and Assessment
  3. On the Determinants of Intrinsic and Extrinsic Motivation
  4. On Motivational Outcomes
  5. Interventions
  6. Conclusion
- Further Reading

*self-determination* A sense of personal initiative, freedom, and choice in behavior.

*self-determined motivation* Motivation that is characterized by high levels of intrinsic motivation and identified regulation but low levels of external and introjected regulation and amotivation.

*top-down effect* The impact of more generalized motivational orientations on more specific motivations.

## GLOSSARY

*amotivation* A lack of purpose and intentionality in one's action, that is, the relative absence of motivation.

*extrinsic motivation* Engaging in an activity as a means to an end and not for the activity's own sake.

*intrinsic motivation* Engaging in an activity for itself and for the pleasure and satisfaction derived from participation.

*motivation* The hypothetical construct that is used to describe the internal and/or external forces that lead to the initiation, direction, intensity, and persistence of behavior.

*need for autonomy* The desire to be the origin of one's behavior.

*need for competence* The desire to interact effectively with the environment.

*need for relatedness* The desire to be connected to significant others in one's interaction with the environment.

*recursive bottom-up effect* The impact that specific motivations have in a repeated matter over time on more generalized forms of motivation.

The purpose of this article is to provide an overview of what is known about intrinsic and extrinsic motivation in sport. After some definitional issues, the article focuses on the assessment of intrinsic and extrinsic motivation. It then proceeds to a discussion on the determinants and outcomes of intrinsic and extrinsic motivation. The article concludes with a discussion on intervention issues.

## 1. INTRODUCTION

Motivation represents one of the most important variables in sport. In fact, coaches and athletes agree that motivation is one of the key elements that will facilitate not only performance but also a positive experience in the sport area. But what is motivation? More specifically, what are intrinsic motivation and extrinsic motivation? Why do certain athletes seem more motivated

than others? Does motivation matter with respect to outcomes? If so, how can coaches and other practitioners best facilitate athletes' motivation? The purpose of this article is to deal with these questions and, thus, to provide a better understanding of the nature and functions of intrinsic and extrinsic motivation in sport.

## 2. DEFINITIONS AND ASSESSMENT

### 2.1. On Motivation

Motivation has been the subject of attention from a host of individuals, including playwrights, philosophers, and psychologists. Psychologists, in particular, have looked at the concept for the past two centuries and have provided some general definitions of the construct. A widely accepted definition of motivation is that it represents the hypothetical construct used to describe the internal and/or external forces that lead to the initiation, direction, intensity, and persistence of behavior. Thus, motivation leads to action.

One of the difficulties in defining motivation is that it is not directly observable. Thus, one strategy used to define motivation, implied by the preceding definition, has been to look at behavior to determine whether someone is motivated or not. For instance, a coach would infer that a player is motivated to perfect his or her dribble if the player stays after practices and works on dribble drills.

### 2.2. On Intrinsic and Extrinsic Motivation and Amotivation

#### 2.2.1. Intrinsic Motivation

Over the years, psychologists have come to realize the existence of different types of motivation. Two broad types that have been studied extensively are intrinsic

motivation and extrinsic motivation. Intrinsic motivation refers to engaging in an activity for itself and for the pleasure and satisfaction derived from participation. An example of intrinsic motivation is an athlete who plays basketball because he or she finds it interesting and satisfying to learn new moves with the ball. In 1992, Vallerand and colleagues proposed and showed that there are at least three types of intrinsic motivation: intrinsic motivation to know (engaging in the activity for the pleasure of learning), intrinsic motivation toward accomplishments (engaging in the activity for the pleasure of trying to surpass oneself), and intrinsic motivation to experience stimulation (engaging in the activity out of sensory and aesthetic pleasure).

#### 2.2.2. Extrinsic Motivation

When extrinsically motivated, individuals do not engage in the activity out of pleasure but rather do so to derive some kind of rewards that are external to the activity itself. An athlete who participates in the Olympics to obtain a gold medal and the associated fame and fortune that go with it represents an instance of an extrinsically motivated athlete. Avoiding punishment also pertains to extrinsic motivation.

In 1985, Deci and Ryan published a very influential book titled *Intrinsic Motivation and Self-Determination in Human Behavior*. In their book, they proposed the existence of a number of types of extrinsic motivation that vary in terms of their inherent levels of self-determination. From the lowest to the highest level of self-determination, these types of extrinsic motivation are external regulation, introjected regulation, identified regulation, and integrated regulation. Figure 1 presents the various types of motivation on the self-determination continuum.



FIGURE 1 The various types of motivation on the self-determination continuum. Adapted from Vallerand and Perreault (1999).

External regulation refers to behavior that is regulated through external means such as obtaining rewards (e.g., medals, trophies) and avoiding constraints (e.g., social pressure). For instance, an athlete might say, "I'm going to today's practice because I don't want the coach to make me sit on the bench during the next game." A second type of extrinsic motivation is introjected regulation. With this type of regulation, the individual has started to internalize the reasons for his or her actions. However, such internalization merely replaces the external source of control with an internal one, that is, self-imposed guilt and anxiety. Thus, an athlete might say, "I go to practices because I would feel guilty if I did not." To the extent that the behavior becomes valued and judged to be important for the individual, and especially to the extent that it is perceived as chosen by oneself, the internalization of extrinsic motives becomes regulated through identification. For instance, an individual might say, "I choose to go to practices because it will help me to perform better during the next game." Finally, an integrated regulation also involves engaging in an activity from an extrinsic perspective in a "choiceful" manner. However, in this case, one's choices are made as a function of coherence with the various aspects of the self. For instance, the integrated athlete might decide to stay home on a Saturday night, rather than go out with friends, so that the athlete will be ready for the next day's soccer game.

### 2.2.3. Amotivation

Finally, psychologists have also proposed the existence of the construct of amotivation, which refers to a lack of purpose and intentionality in one's action. When amotivated, athletes experience feelings of incompetence and expectancies of uncontrollability and an important loss of motivation. For instance, an amotivated athlete might feel incompetent and start to wonder whether participating in sport still represents a worthwhile option. Amotivation amounts to the relative absence of motivation.

## 2.3. Assessment of Intrinsic and Extrinsic Motivation and Amotivation

In his 1997 article on the hierarchical model of intrinsic and extrinsic motivation, Vallerand proposed that intrinsic and extrinsic motivation and amotivation,

exist at three levels of generality. From the highest to the lowest level in the hierarchy, these are the global, contextual, and situational levels.

Motivation at the situational level refers to a motivational state. It is the motivation that people experience toward a given activity at a specific point in time. For instance, situational motivation toward swimming would be the motivation experienced by a swimmer while training in the pool at 6 o'clock on a Friday morning.

Intrinsic motivation at the situational level has been typically measured in two ways. First, psychologists have used what is called the free choice period. Participants, after having been subjected to experimental conditions, are told that the experiment is over and are provided with the opportunity to engage again in the activity if they so desire. Participants are then left alone and observed unobtrusively. The number of seconds spent on the activity represents the measure of participants' intrinsic motivation toward the experimental task. The more time spent on the task, the higher participants' intrinsic motivation.

This behavioral index of intrinsic motivation is particularly useful in laboratory conditions. However, it is not as useful in real-life conditions, such as sport, where the free choice measure cannot be readily used for practical reasons. Researchers increasingly are turning toward self-report questionnaires to assess participants' motivation. In line with the definition of intrinsic and extrinsic motivation, psychologists have used participants' reasons to participate in the activity as a measure of intrinsic and extrinsic motivation. One self-report instrument that has been used at the situational level with much success in sport is the Situational Motivation Scale (SIMS). Developed by Guay and colleagues in 2000, the SIMS measures four types of motivation: intrinsic motivation without distinguishing the three types, identified regulation, external regulation, and amotivation. Only four subscales of four items each are used to keep the scale brief so as to capture situational motivation in field and laboratory situations. The scale has shown high levels of validity and reliability.

Motivation at the contextual level refers to a relatively stable and general motivational orientation that pertains only to activities that are subsumed under a specific broad life domain such as education, interpersonal relationships, or sport. Thus, individuals come to develop motivational orientations with respect to intrinsic and extrinsic motivation and amotivation in different contexts. A given contextual motivation always refers to one specific life domain. For instance, an

athlete with a high externally regulated motivational orientation toward basketball would generally tend to engage in various basketball situations to receive rewards and/or avoid punishment. Interestingly, a person's contextual motivational orientation may vary from context to context. Indeed, an athlete may have a high level of contextual intrinsic motivation toward his or her sport but a low level of contextual intrinsic motivation toward education.

One instrument that has been used to measure motivation in sport at the contextual level is the Sport Motivation Scale (SMS). In this questionnaire, validated by Pelletier and colleagues in 1995, athletes are asked the following question: "Why do you practice your sport?" Each item represents a perceived reason for engaging in the activity, thereby reflecting the various types of motivation seen previously. Items are assessed using a 7-point Likert-type scale. Sample items appear in Table I. As can be seen, the various types of intrinsic and extrinsic motivation and amotivation are assessed through the reasons for people's participation in sport. There are four items per subscale and seven subscales. The total score of each

subscale is a sum of the four items. It should be noted that the concept of integrated regulation is not assessed in the SMS because research has shown that this construct has not yet crystallized in adolescents and young adults. Research using the French and English versions of the SMS reveals that the scale has high levels of validity and reliability. The scale has also been validated in several other languages, including Chinese, Italian, Japanese, Russian, and Spanish.

Finally, motivation at the global level refers to a general motivational orientation to interact with the environment in an intrinsic, extrinsic, or amotivated way. Contrary to contextual motivation, global motivation does not pertain specifically to one given life domain or set of activities but rather is akin to a broad personality trait. Thus, certain individuals may have a global external regulation orientation. Such a personality orientation would predispose individuals to have an external regulation toward several life contexts, including sport. Guay, Pelletier, Vallerand, and others have developed the Global Motivation Scale to assess the seven types of motivational constructs that appear in Fig. 1 (except for integrated regulation) at the global level. It has shown high levels of reliability and validity.

TABLE I  
Sample Items from Each of the Seven Subscales  
of the Sport Motivation Scale

"In general, why do you practice your sport?"	
• Intrinsic motivation—knowledge	"For the pleasure of discovering new training techniques"
• Intrinsic motivation—accomplishments	"For the pleasure that I feel while executing certain difficult movements"
• Intrinsic motivation—stimulation	"For the pleasure I feel in living exciting experiences"
• Identified regulation	"Because it is one of the best ways to maintain good relationships with my friends"
• Introjected regulation	"Because I must do sports to feel good about myself"
• External regulation	"For the prestige of being an athlete"
• Amotivation	"I don't know anymore; I have the impression of being incapable of succeeding in this sport"

### 3. ON THE DETERMINANTS OF INTRINSIC AND EXTRINSIC MOTIVATION

Because motivation is present in people at three levels of generality, it is important to look at the determinants of motivation at each of the three levels: global, contextual, and situational.

#### 3.1. Where Does Situational Motivation Come From?

Research reveals that at least three types of variables produce an important effect on situational motivation: the task, the person, and the environment.

##### 3.1.1. The Task

Clearly, tasks differ in terms of their intrinsic properties in that certain tasks are more enjoyable than others. For instance, playing basketball would be more interesting to most basketball players than would running up the stairs of a stadium. Thus, it is

not surprising that certain tasks generate higher levels of situational intrinsic motivation than do others.

Yet individual differences and environmental variables may affect perceptions of tasks in such a way that even dull tasks may, at times, be experienced as enjoyable. Thus, much research has focused on the impact of environmental and individual difference variables on intrinsic and extrinsic motivation.

### **3.1.2. The Environment**

Historically, research on intrinsic and extrinsic motivation started during the early 1970s by looking at the effects of rewards on situational intrinsic motivation. In what is recognized as the first contemporary publication on intrinsic motivation, Deci showed in 1971 that receiving rewards to engage in an enjoyable activity produces a decrease in situational intrinsic motivation. Specifically, individuals who received monetary rewards to engage in a fun activity spent less time on the activity subsequently than did individuals who did not receive rewards to engage in the same activity. Subsequent research showed that the negative effects of rewards are especially potent when the rewards are expected and provided to entice participants to engage in the activity. Although there have been some controversies over the years, a meta-analysis of more than 100 studies conducted by Deci and colleagues in 1999 provided overwhelming support for the negative effects of rewards on situational intrinsic motivation. Research in sport, by researchers such as Orlick and Mosher in 1978, has replicated these findings and shown that trophies and awards do undermine intrinsic motivation.

Over the years, research in sport has expanded and looked at the impact of other situational factors on intrinsic motivation. Situational factors are variables that are present at a given point in time. Research has shown that factors such as competition, winning or losing, and performing well or poorly affect situational intrinsic and extrinsic motivation. Specifically, competing to win at all costs, as well as losing or not playing well, produces a decrease in intrinsic motivation and identified regulation and an increase in introjected and external regulation and amotivation. However, noncompetitive situations, as well as winning and doing well, produce the opposite effect.

Why do factors as diverse as rewards, competition, and losing influence athletes' intrinsic and extrinsic motivation? According to Deci and Ryan's self-determination theory, social factors influence motivation through their

impact on one's perceptions of competence, autonomy, and relatedness. Because individuals experience the needs to feel competent, autonomous, and related to significant others in their interaction with the environment, activities that allow people to satisfy these needs will be engaged in choicefully and on a regular basis. In other words, athletes experience important and powerful needs to feel connected to others in the sport milieu, to behave in an effective manner in that milieu, and to experience a sense of personal initiative in the process. Variables that nurture those needs help intrinsic motivation and identified regulation to flourish. Variables that thwart those needs produce negative effects on the latter types of motivation and facilitate the development of external and introjected regulation and amotivation. For instance, performing badly in a sport leads one to feel incompetent, and this in turn fuels a loss of intrinsic motivation and identified regulation and a concomitant increase in external and introjected regulation and amotivation. Research using sophisticated statistical analyses, such as structural equation modeling and path analysis, has shown that the impact of the environment on intrinsic and extrinsic motivation is indeed mediated by perceptions of competence, autonomy, and relatedness.

### **3.1.3. Individual Differences**

We have seen how social factors can have a very important impact on athletes' situational motivation. Yet athletes who play on the same team, and who are subjected to the same environmental conditions, often display different levels of situational intrinsic and extrinsic motivation. Why is this the case? Psychologists propose that these differences take place because athletes may differ with respect to their motivational orientations. A motivational orientation typically refers to a more or less stable predisposition to be intrinsically or extrinsically motivated or even amotivated. Such an orientation can be present at the contextual or global level. For instance, an athlete with an intrinsic contextual motivation toward basketball is more predisposed to display higher levels of intrinsic motivation in various basketball situations than is an athlete with a contextual motivation characterized by low intrinsic motivation. This refers to what Vallerand called the top-down effect. It is important to note that such effects appear to hold only for contextual motivational orientations related to the activity in which the individual is engaging. Thus, having a contextual intrinsic motivational orientation toward hockey would help an athlete to maintain a high level of situational

intrinsic motivation toward hockey. However, having a contextual intrinsic motivational orientation toward school would not produce such an effect. Finally, global motivation will influence situational motivation only when the activity being engaged in is rather ambiguous and unrelated to a given contextual motivation. In such instances, global motivation has room to exert its influence on situational motivation.

### **3.2. On the Determinants of Contextual Motivation**

One's contextual motivation toward a specific set of activities can also be affected by the same three factors just reviewed. Thus, task factors are also important influences on contextual motivation. For example, a baseball player's contextual motivation will likely be more intrinsic for actual baseball games than for weight training.

Social factors at the contextual level have been also found to affect contextual motivation. Such factors are those that are present on a general or recurrent basis in one life context (e.g., the sport in which one participates) but not necessarily in other life contexts (e.g., education, interpersonal relationships). Contextual factors studied in sport can be loosely classified into three categories: structural variables (e.g., the types of leagues in which athletes play), motivational climates that prevail in team settings, and interpersonal variables (e.g., the coach's managerial style). Highly structured and competitive league environments, performance-oriented team climates, and playing for a highly controlling coach who does not leave room for autonomy in the athletes all produce decreases in intrinsic motivation and identified regulation while producing increases in introjected and external regulation and amotivation. However, more relaxed league structures, mastery-oriented team climates, and playing for an autonomy-supportive coach who provides understanding, care, and room for decision making on the athletes' part all produce the opposite effect on contextual motivation.

A third source of influence on contextual motivation originates from individual differences. Just like situational motivation, contextual motivation is also subjected to top-down effects originating from individual differences. However, in this case, the influence comes from global motivation. Thus, athletes who entertain a high level of global intrinsic motivation generally possess high levels of intrinsic motivational orientations in various life contexts, including sport.

A final influence on contextual motivation comes from the cumulative effects of experiencing various motivational states on a regular basis. This refers to the recursive bottom-up effect. For example, a tennis player who interacts daily with an autonomy-supportive coach will likely experience daily doses of situational intrinsic motivation that, over time, should have positive recursive effects on the athlete's contextual motivation toward tennis. In the long run, the athlete's contextual motivation will likely become increasingly intrinsic in nature.

### **3.3. Motivational Determinants at the Global Level**

Very little is known about the determinants of global motivation, especially as it relates to the realm of sport. One source of influence that has been documented deals with the recursive bottom-up effect discussed previously. Thus, experiencing high levels of intrinsic motivation at the contextual level in sport leads to the development of a more intrinsic global motivation. Of course, this effect takes place to the extent that this life domain is important for the individual, as would be the case for athletes.

An additional source of influence comes from the impact of global factors. These refer to social factors that are present in most contexts of a person's life on a relatively permanent basis. Global factors help to shape a person's global motivation. In this regard, the motivational impact on global motivation of enrolling at a young age in soccer boarding schools, such as those found in several countries around the world, deserves attention. Such boarding schools represent a global environment in that athletes not only learn how to play soccer, in the hope of becoming professional soccer players in the future, but also go to school and actually live at the boarding schools, away from their families, for years. Therefore, boarding schools should profoundly affect the development of young individuals' personalities, including their global motivation. The type of effect induced, however, will likely depend on the climate that prevails in such schools. If the climate is one of autonomy support, care, and understanding, athletes will likely develop a self-determined global motivation. However, tightly controlling boarding schools will likely lead to the development of non-self-determined global motivation. Thus, being subjected to such global factors shapes athletes' global motivation and personalities. Such a global motivation,

in turn, greatly affects athletes' motivation toward sport through the top-down effect.

## 4. ON MOTIVATIONAL OUTCOMES

### 4.1. Motivation Causes Outcomes

It is intuitively thought that high levels of motivation will likely lead an athlete to play hard, persist more at training, and display more intense behavior compared with a less motivated athlete. But is it the case? Does motivation actually cause the outcome? The answer to this question is an unequivocal yes. Indeed, several studies have shown that experimentally inducing intrinsic motivation, as opposed to extrinsic motivation, leads to several positive outcomes, including enhanced creativity, performance, and persistence. In other words, motivation matters.

#### 4.1.1. Outcomes as a Function of the Various Types of Motivation

The type of motivation that one entertains toward an activity also matters in that outcomes are decreasingly positive as one moves from intrinsic motivation to amotivation (see Fig. 1). Because the various types of motivation are located on a continuum ranging from high to low self-determination, and because self-determination is associated with enhanced psychological functioning, a corresponding pattern of consequences ensues. Thus, the most positive outcomes result from the self-determined forms of motivation (intrinsic motivation and identified regulation), whereas negative outcomes follow from the least self-determined forms of motivation (external and introjected regulation and especially amotivation). This pattern of results has been obtained in sport and physical activity with the following outcomes: concentration, sportsmanship, and positive affective experiences such as flow, interest, enjoyment, satisfaction, and less anxiety as well as behavioral outcomes that include expanded effort, behavioral intentions, and actual levels of persistence, adherence in sport, and performance.

It should also be noted that these findings have been obtained with different populations, including elite and recreational athletes, athletes of various ages, physically challenged athletes, intellectually challenged children, and coaches. Thus, it would appear that the role of motivation in producing outcomes is well established

and that the more self-determined the motivation, the more positive the outcomes.

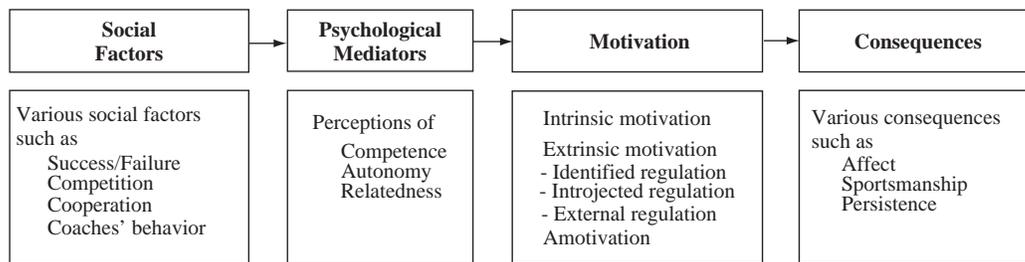
#### 4.1.2. Outcomes as a Function of Levels of Generality

Just like motivation, outcomes are experienced at various levels of generality. For instance, having the sudden intention to drop out of basketball is not the same as making such a decision after several weeks or months of reflection. Similarly, feeling momentarily dejected after a loss in a tennis tournament is different from experiencing a generalized negative mood each time one plays tennis. Why is this the case?

Psychologists have shown that outcomes not only differ in terms of generality but also are produced by motivations that differ in terms of their levels of generality. Thus, situational types of outcomes, such as feeling momentarily disappointed and not wanting to play any more at that point in time, originate from situational motivation. On the other hand, more contextually generalized thoughts, feelings, and behaviors that go beyond the moment and pertain to a specific area, such as tennis or football, are produced by sport contextual motivation specific to tennis or football. Finally, broad-level types of outcomes, such as full-fledged depression and apathy experienced across life domains, are typically the fruit of global motivation.

## 5. INTERVENTIONS

It is well known that participation in sport and exercise produces several physical and psychological benefits. Thus, it is unfortunate to see that a large number of children drop out of sport at a young age. For instance, in Canada, there is a 30 to 35% annual dropout rate in young competitive swimmers, some of whom have great potential. In 1999, Vallerand and Losier proposed the following causal sequence: Social Factors → Psychological Mediators → Motivation → Consequences. Such a sequence integrates the determinants and outcomes of motivation and, as such, charts the nature of the psychological processes through which athletes come to experience various outcomes, including dropping out of sport (Fig. 2). Indeed, research has shown that coaches who display controlling behavior (a social factor) undermine young athletes' sense of competence, autonomy, and relatedness (the psychological mediators). This, in turn, undermines their self-determined motivation,



**FIGURE 2** The motivational causal sequence: From social factors to outcomes. Adapted from Vallerand and Losier (1999).

leading to their dropping out of sport (a consequence). These findings have been obtained with both male and female athletes who engage in various sports such as handball and swimming.

If motivation has a causal influence on persistence, it might be possible to increase athletes' motivation and, in turn, their persistence toward sport by using the causal sequence depicted previously. This could be done by modifying social factors known to negatively affect motivation such as coaches' controlling interacting style. Although very few intervention studies have been conducted in the realm of intrinsic and extrinsic motivation, one sport study has followed this lead. In 2003, Pelletier and colleagues reported the development of an intervention program dealing with swimmers' motivation and persistence. This 18-month program mainly focused on the following:

- Helping coaches to become more autonomy supportive, thereby enhancing athletes' feelings of competence and autonomy
- Teaching athletes how to deal with the increased levels of competence and autonomy and to become more proactive in their sport environment

It was anticipated that changing coaches' interacting style to a more autonomy-supportive approach would trigger the previously described causal sequence and would lead athletes to experience higher levels of competence and autonomy, higher levels of intrinsic motivation, and higher levels of persistence in swimming.

Results supported this hypothesis. Approximately 1½ years into the intervention program, athletes perceived their coaches as significantly less controlling and as more autonomy supportive. In turn, athletes' levels of perceived competence and intrinsic motivation toward swimming increased, leading to several adaptive consequences. The levels of attendance in the practices increased dramatically, and the annual dropout rate

decreased from 35% to less than 5%. A control group of swimmers who were not part of the intervention still displayed a 35% dropout rate. In addition, 85% of athletes who were part of the intervention reached the national performance standards, four were selected to the Canadian Olympic swimming team, and one won a silver medal in the Seoul Olympics. Although preliminary, results from this intervention study underscore the potential for intervention programs that use the proposed motivational sequence as a guiding light.

## 6. CONCLUSION

Sport motivation is a complex and multidimensional phenomenon. It is influenced by numerous factors and can lead to a host of consequences. Furthermore, not only are intrinsic motivation and identified regulation important for allowing athletes to experience satisfying participation in sport, but these self-determined forms of motivation also lead to higher levels of achievement. In this vein, interventions that promote these forms of motivation have been found to be effective. The road to enjoyable and successful participation in sport has been mapped out.

### See Also the Following Articles

- Goal Setting and Achievement Motivation in Sport
- Motivation and Culture

### Further Reading

- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105–115.
- Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum.

- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.
- Goudas, M., Biddle, S., Fox, K., & Underwood, M. (1995). It ain't what you do, it's the way that you do it! Teaching style affects children's motivation in track and field lessons. *The Sport Psychologist*, 9, 254–264.
- Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of the situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24, 175–213.
- Orlick, T. D., & Mosher, R. (1978). Extrinsic awards and participant motivation in a sport related task. *International Journal of Sport Psychology*, 9, 27–39.
- Pelletier, G. L., Fortier, M. S., Vallerand, R. J., & Brière, N. M. (2001). Associations between perceived autonomy support, forms of self-regulation, and persistence: A prospective study. *Motivation and Emotion*, 25, 279–306.
- Pelletier, L. G., Fortier, M. S., Vallerand, R. J., Tuson, K. M., Brière, N. M., & Blais, M. R. (1995). Toward a new measure of intrinsic motivation, extrinsic motivation, and amotivation in sports: The Sport Motivation Scale (SMS). *Journal of Sport & Exercise Psychology*, 17, 35–53.
- Sarrazin, P., Vallerand, R., Guillet, E., Pelletier, L., & Cury, F. (2001). Motivation and dropout in female handballers: A 21-month prospective study. *European Journal of Social Psychology*, 31, 1–24.
- Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 271–360). San Diego: Academic Press.
- Vallerand, R. J. (2001). A hierarchical model of intrinsic and extrinsic motivation in sport and exercise. In G. Roberts (Ed.), *Advances in motivation in sport and exercise* (2nd ed., pp. 263–319). Champaign, IL: Human Kinetics.
- Vallerand, R. J., & Fortier, M. S. (1998). Measures of intrinsic and extrinsic motivation in sport and physical activity: A review and critique. In J. Duda (Ed.), *Advancements in sport and exercise psychology measurement* (pp. 83–100). Morgantown, WV: Fitness Information Technology.
- Vallerand, R. J., & Losier, G. F. (1999). An integrative analysis of intrinsic and extrinsic motivation in sport. *Journal of Applied Sport Psychology*, 11, 142–169.
- Vallerand, R. J., & Reid, G. (1984). On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport Psychology*, 6, 94–102.
- Vallerand, R. J., & Rousseau, F. L. (2001). Intrinsic and extrinsic motivation in sport and exercise: A review using the hierarchical model of intrinsic and extrinsic motivation. In R. Singer, H. Hausenblas, & C. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 389–416). New York: John Wiley.



# Intrinsic Motivation and Self-Determination

Edward L. Deci

University of Rochester, Rochester, New York, USA

1. Introduction
  2. Extrinsic Motivation Versus Intrinsic Motivation
  3. The Interaction of Extrinsic Motivation and Intrinsic Motivation
  4. Extrinsic Motivation
  5. Application of Research on Self-Determination
  6. Summary and Conclusions
- Further Reading

*intrinsic motivation* Doing an optimally challenging activity that is interesting, has an internal perceived locus of causality, and is energized by the basic psychological needs for competence, autonomy, and relatedness; such an activity does not require a contingent outcome that is separate from the activity itself.

*introjection* A form of internalization in which people take in an external control without ever really accepting it as their own and then use it to pressure and control themselves.

## GLOSSARY

*autonomy support* A quality of a person's interpersonal environment in which others acknowledge the person's perspective, provide choice, encourage initiative, and minimize the use of pressure and control.

*controlling environments* Environments in which people in positions of authority, expertise, or power pressure others to think, feel, or behave in specific ways; it is the antithesis of autonomy support.

*extrinsic motivation* Doing a behavior because it will lead to some separate outcome such as a reward, approval from others, or the avoidance of punishment.

*integration* The fullest type of internalization in which people take in an extrinsic motivation, identify with it, and integrate it into their sense of self.

*interpersonal context* The ambience or "climate" of a particular setting that can vary from supporting people's autonomy to controlling their behavior.

Self-determination in human behavior is based in autonomous motivation, which encompasses both intrinsic motivation and integrated extrinsic motivation. Intrinsic motivation involves doing an activity without the necessity of external prompts or rewards because it is interesting and satisfies the basic psychological needs for competence, autonomy, and relatedness. Extrinsic motivation involves doing an activity because it leads to some separate outcome such as a reward, approval from others, or the avoidance of punishment. Intrinsic motivation is invariably autonomous, whereas extrinsic motivation varies in its degree of autonomy. Extrinsic motivation becomes autonomous through the processes of internalization and integration. Both intrinsic motivation and integrated extrinsic motivation are facilitated in home, school, work, and clinical settings that are interpersonally supportive, that is, where people's perspectives are acknowledged and where people are

encouraged to experiment, allowed to try their own solutions to problems, provided with choice, and responded to when they initiate. Under such supportive conditions, individuals are more engaged and persistent, perform more effectively, and display higher levels of psychological health and well-being.

## 1. INTRODUCTION

The delight of discovery and the joy of leisure pursuits are experiences known to most everyone. They are stimulating, satisfying, and sustaining of the behaviors that promote them. Within the field of motivation, these experiences and the psychological processes that underlie them are encompassed by the concept of intrinsic motivation, which was introduced to the field by Harlow in 1950 and began receiving empirical attention two decades later. The concept has become pivotal in our understanding of human motivation and has provided important insights even about people's motivation for life's mundane activities such as studying uninteresting topics and doing boring work.

Intrinsic motivation is readily observable in young children; indeed, play is the "business of their lives." They happily turn random objects into toys, they role-play others such as their parents, and they invent all kinds of games. Intrinsic motivation is also apparent in older individuals, although perhaps not so ubiquitously. Leisure activities such as playing golf, looking at paintings, and pursuing a bargain on a new dress are often intrinsically motivated (i.e., done for the spontaneous satisfaction provided by the activities themselves), and some people are fortunate enough to be intrinsically motivated for their jobs.

Psychologists began to treat the experience of delight and joy (i.e., of interest and enjoyment) as an important motivator following from the realization that a fuller understanding of human motivation required distinguishing among types of motivation rather than just focusing on the amount of motivation. Stated differently, most theories of motivation or behavioral control, beginning with behavioral theories such as that of Hull and continuing through the cognitive theories of Bandura and Seligman, treated motivation as if it differed only in amount and not in type. People who were more motivated for a particular activity (or class of activities) were expected to do the activity more often and to do it more effectively than were people with less motivation. It was the amount of motivation that was considered critical.

The idea that there are different types of motivation implies that if two people have the same amount of motivation for a particular activity but the types of motivation are different, the quality of their engagement in and persistence at the activity will also be different. For example, there are times when employees may be highly motivated to do an activity at work because they are fearful that they will be punished if they do not do it, and there are other times when employees are equally highly motivated for such an activity because they find it interesting and personally important. Any individuals who can relate to this probably recognizes that these different types of motivation, even if the amounts are the same, can have a major effect on the quality of their engagement with the activities and perhaps even on their feelings about themselves.

## 2. EXTRINSIC MOTIVATION VERSUS INTRINSIC MOTIVATION

The first important differentiation of motivation into types distinguished between extrinsic motivation and intrinsic motivation. Extrinsic motivation refers to doing a behavior because it will lead to some separate outcome such as a reward, approval from others, or the avoidance of punishment. The concept of extrinsic motivators is linked to behaviorist theories that were prominent in psychology during the 1940s and 1950s, although those theories did not use the term "extrinsic motivation." Behavior modification, for example, was based solely on extrinsic motivators—do a desired behavior and get a reward (which behaviorists termed a reinforcement). Behavior modification was in widespread use in education, therapy, and job training during the 1970s, and then its use gradually tapered off.

### 2.1. Extrinsic Motivation and Behaviorism

The two primary strands of behaviorism, Hullian drive theory and Skinnerian operant theory, both were built around the concept of reinforcement. In both theories, people learn behaviors when they are reinforced. However, the two theoretical approaches define the concept of reinforcement very differently.

In drive theory, reinforcers are events that reduce basic physiological drives such as hunger and thirst. For example, food reinforces behaviors that lead to it while the hunger drive is operative. In contrast, within

operant theory, reinforcers are defined as any separable consequence (e.g., a food pellet, a dollar bill) that increases the likelihood of a particular response being emitted. In operant theory, reinforcers are not tied to underlying drives.

In operant theory, anything is a reinforcer if it increases the rate of a particular behavior, whereas in drive theory, only those things that reduce drives are called primary reinforcers. However, drive theory includes a concept of secondary reinforcers, which are other things (e.g., monetary rewards, smiles) that take on reinforcing value by being paired with primary reinforcers. For example, money can reinforce because it becomes associated with eating and/or drinking.

## 2.2. Intrinsic Motivation

Intrinsic motivation is based in people's inherent tendency to be proactive, to interact with the world in an attempt to have an effect, and to feel a sense of accomplishment. Indeed, when people are at their healthiest, they are curious, eager to take on challenges, partial to novelty, engaged with interesting tasks or stimuli, and ready to learn. All of these are manifestations of intrinsic motivation, and they are critical for optimal development across the life span.

The concept of intrinsic motivation arose, in part, in response to problems that became apparent within each strand of behaviorism. Accordingly, there were initially two different characterizations of the concept. The first suggested that not all behaviors are derived from the basic drives such as thirst, hunger, and sex. For example, children often forgo eating, even if they have not eaten for a long time, so as to continue playing a game they find to be fun. Accordingly, White asserted that, just as people have biological needs—the so-called drives—that must be satisfied for them to remain healthy, people also have psychological needs that must also be satisfied for them to thrive. White specified the need for competence as being important. Subsequently, Deci and Ryan maintained that there are three fundamental psychological needs—the needs for competence, autonomy, and relatedness—that are operative for all humans and must be satisfied for them to develop and function optimally.

The other approach to defining intrinsic motivation suggested that intrinsically motivated behaviors are those for which the behaviors themselves are rewarding and, thus, do not require separable consequences. This, of course, stands in contrast to the Skinnerian

perspective that suggests that all behaviors are controlled by reinforcing consequences that are separate from the behaviors themselves.

When the two characterizations of intrinsic motivation are considered together, the conclusion is that not all behaviors are controlled by reinforcements and that inherent psychological needs for competence, autonomy, and relatedness underlie those behaviors. Other discussions of intrinsic motivation have helped to flesh out the definition. DeCharms used attribution theory to suggest that extrinsic motivation is operative when people perceive the locus of causality for their behavior to be external to themselves, whereas intrinsic motivation is operative when people perceive the locus of causality to be internal or within themselves. For example, when people are doing an activity to get a reward, they would likely see the reward as the cause of their behavior; thus, the perceived locus of causality (PLOC) would be external. When people are doing an activity because they find it interesting and fun, they would likely see the enjoyment as the cause of the behavior, and so the PLOC would be internal. The idea of PLOC relates to the need for autonomy in that people will be satisfying their need for autonomy to the extent that the PLOC for a behavior is internal.

Another characterization that has been widely used in discussions of intrinsic motivation is that being intrinsically motivated for an activity involves doing an activity because people find the activity interesting. The idea is that people do not need a reinforcement or reward if the activity itself interests them. Thus, the concept of "interest" has been considered central to intrinsic motivation. Using the concept of interest highlights that intrinsic motivation exists in the relation between individuals and activities. That is, a person is intrinsically motivated for some activities and not for others, and not everyone is intrinsically motivated for any particular activity. Rather, an individual is intrinsically motivated for a task to the degree that he or she finds the task interesting, and people differ in the extent to which they find any particular task interesting.

A final important piece of the overall characterization of intrinsic motivation states that optimal challenge is central to intrinsic motivation. That is, people are intrinsically motivated to take on optimal challenges because those are the activities that are most interesting, facilitate competence, and promote development.

In sum, intrinsic motivation involves doing an optimally challenging activity that is interesting, has an internal PLOC, and is energized by the basic psychological needs for competence, autonomy, and

relatedness. Such an activity does not require a contingent outcome that is separate from the activity itself.

### **3. THE INTERACTION OF EXTRINSIC MOTIVATION AND INTRINSIC MOTIVATION**

When extrinsic motivation and intrinsic motivation were identified and defined, researchers began to wonder whether the two might interact. Are these two types of motivation additive, such that the amount of one's intrinsic motivation and the amount of one's extrinsic motivation for a particular activity can be summed to yield one's total motivation for the activity? Stated differently, would anything happen to people's intrinsic motivation for an activity if people began to receive extrinsic inducements for doing it? Would intrinsic motivation itself be enhanced, unchanged, or diminished? If the answer is anything other than "unchanged," this implies that the different types of motivation are not simply additive and, thus, that environmental conditions, such as reward contingencies and deadlines, could affect people's intrinsic motivation when they are operating within those environmental conditions.

#### **3.1. Reward Effects**

The first studies of intrinsic motivation in humans, published in 1971 by Deci, examined the effects of monetary rewards on the intrinsic motivation of college students. In subsequent studies that included participants from preschool and high school, the effects of other tangible rewards, such as prizes and good player certificates, were also examined. The studies consistently revealed that tangible extrinsic rewards, whether payments or prizes, undermined intrinsic motivation. Specifically, the undermining occurred when the activity was initially interesting, the rewards were salient, and the rewards were contingent on and expected while doing the activity. Deci interpreted these results as indicating that the extrinsic rewards tend to leave people feeling controlled by the rewards (rather than feeling autonomous). Accordingly, their PLOC tends to become more external.

The general finding that tangible rewards tend to undermine intrinsic motivation has been replicated numerous times, but the finding has remained controversial for three decades. Because rewards are the

easiest motivators to use, some people resist recognizing their potentially detrimental effects. Nonetheless, in 1999, Deci and colleagues published a meta-analysis involving 128 experiments on reward effects and showed definitively that tangible extrinsic rewards do reliably and clearly undermine intrinsic motivation for the rewarded activity.

#### **3.2. Other External Events**

Additional research has shown that other external events, such as threats of punishment, deadlines, directives, and competition, also undermine intrinsic motivation, presumably because people experience them as controls. On the other hand, the opportunity to make choices about what to do and how to do it was found to enhance intrinsic motivation by facilitating a greater experience of autonomy and a more internal PLOC.

The effects of all these studies were interpreted in accord with the postulate that people have a basic need for autonomy. Accordingly, factors that support people's autonomy and experience of an internal PLOC (e.g., choice, acknowledgment) tend to enhance intrinsic motivation, whereas those that leave people feeling controlled with an external PLOC (e.g., tangible rewards, deadlines, pressures) tend to diminish intrinsic motivation.

#### **3.3. Performance Feedback**

Other research relates to the basic psychological need for competence. Studies have shown that positive performance feedback (which is sometimes referred to as verbal rewards) enhances intrinsic motivation, that negative feedback diminishes it, and that perceived competence mediates these effects. There is also evidence, however, that the enhancement of intrinsic motivation following praise is less likely to occur for children than for college students. Furthermore, studies have found that increases in perceived competence must be accompanied by feelings of autonomy for the perceived competence to enhance intrinsic motivation. If people feel forced to perform well, their good performance and the positive feedback that might follow it are unlikely to enhance their intrinsic motivation.

#### **3.4. Interpersonal Contexts**

A final set of studies has shown that the ambience or "climate" of a particular setting (i.e., the general feeling of the setting) has an effect on people's intrinsic motivation. This factor is frequently referred to as the

interpersonal context, and the critical issue is whether the interpersonal context tends to support people’s autonomy or to control their behavior. For example, the coaching climate of some sporting teams tends to encourage athletes to take initiative and experiment, whereas the climate surrounding other teams conveys that athletes must do what the coach tells them to do. Studies show that if the interpersonal context tends to be autonomy supportive, people within it will be more intrinsically motivated, whereas if it tends to be controlling, people will be less intrinsically motivated. For example, if teachers use a lot of controlling language—demands and imperatives as well as words such as “should,” “must,” and “have to”—their students tend to be less intrinsically motivated. In fact, this issue has been studied in the laboratory, and it seems that if environmental factors such as rewards are administered in relatively autonomy-supportive ways, they are less likely to undermine intrinsic motivation. Furthermore, if factors such as positive feedback are administered in relatively controlling ways, they are less likely to enhance intrinsic motivation.

**3.5. Summary**

Research has shown that intrinsic motivation and extrinsic motivation are not additive; instead, they are interactive. That is, the addition of extrinsic motivators may either diminish or enhance intrinsic motivation. As such, the social environments within which people operate will significantly affect the degree to which they are intrinsically motivated for the target activity. Contextual factors that allowed satisfaction of the basic psychological needs for competence and autonomy were found to enhance intrinsic motivation, whereas those that thwarted satisfaction of these basic needs were found to undermine intrinsic motivation (Table I).

**3.6. Intrinsic Motivation and Outcomes**

Numerous studies have shown that intrinsic motivation was associated with better conceptual learning, greater creativity, more cognitive flexibility, greater behavioral persistence, and more enhanced well-being than was extrinsic motivation. For example, experiments showed that autonomy-supportive environments led both elementary school and college students to learn better than did controlling environments, especially when

**TABLE I**  
**Summary of the Effects of External Motivators on People’s Intrinsic Motivation**

Monetary rewards	Tend to decrease intrinsic motivation
Other tangible rewards	Tend to decrease intrinsic motivation
Deadlines	Tend to decrease intrinsic motivation
Threats of punishment	Tend to decrease intrinsic motivation
Competition against others	Tends to decrease intrinsic motivation
Directives and evaluations	Tend to decrease intrinsic motivation
Negative feedback	Tends to decrease intrinsic motivation
Controlling interpersonal climates	Tend to decrease intrinsic motivation
Choice	Tends to increase intrinsic motivation
Acknowledgment	Tends to increase intrinsic motivation
Encouragement for self-initiation	Tends to increase intrinsic motivation
Positive feedback	Tends to increase intrinsic motivation
Autonomy-supportive interpersonal climates	Tend to increase intrinsic motivation

the material was complex and required conceptual understanding.

Other research conducted in real-world settings such as homes, sport teams, and work groups has revealed that an autonomy-supportive environment catalyzes not only intrinsic motivation but also performance, level of satisfaction, and general well-being. Indeed, many studies have shown strong positive relations between being intrinsically motivated and having a higher level of psychological health and well-being. Such research makes clear why the differentiation of types of motivation is so important and confirms that the distinction between intrinsic motivation and extrinsic motivation is an empirically useful one.

It is critical to remember that intrinsic motivation is operative only for activities that hold intrinsic interest for individuals, for example, those that are novel and challenging. For activities that are not interesting, intrinsic motivation does not represent a potent motivator. Still, there are many activities that parents, teachers,

managers, coaches, and physicians believe are important for people, even if the activities are not interesting. To understand the motivation for these activities, one must examine the dynamics of extrinsic motivation.

## 4. EXTRINSIC MOTIVATION

As it turns out, many of the activities people do are not intrinsically motivated, especially from the time they move out of early childhood and face increasing demands to assume social roles and accept responsibilities. Children must begin to interact congenially with playmates, then to do schoolwork that they do not find interesting, and eventually (as adults) to hold gainful employment and function within the laws of society. Evidence shows, for example, that even during children's years in elementary school, their intrinsic motivation tends to become weaker with each passing year.

Recall that extrinsic motivation is operative when people do an activity to attain some separable outcome such as a pat on the back or a monetary bonus. When children are not intrinsically motivated for an important activity, parents or teachers must rely on the inducement of extrinsic motivators. Although extrinsic motivators have typically been found to be controlling rather than supportive of autonomy, there is still the important question of whether it is possible for people to be autonomous or self-determined when doing a task that is uninteresting and, therefore, was extrinsically prompted. The reason this question is so critical is that if it were possible for people to be autonomous while doing an activity for extrinsic or instrumental reasons, perhaps some of the performance and adjustment advantages that have been found for intrinsic motivation might also accrue when people are autonomously extrinsically motivated.

Deci and Ryan's self-determination theory (SDT) proposed that intrinsic motivation is invariantly self-determination but that extrinsic motivation can vary greatly in the degree to which it is self-determination. For example, a boy who takes out the garbage only because he knows that his mother will praise him for doing so is extrinsically motivated because the behavior is instrumental to the separate consequence. A girl who takes out the garbage because she personally values family harmony and believes that her participation in doing this task will help to maintain that harmony is also extrinsically motivated. The second instance, although still extrinsic, is characterized by a greater sense of personal endorsement and choice than is the

first instance, which involves compliance with the control that has been explicitly or implicitly created by the boy's mother. These two cases of extrinsically motivated behavior vary in the degree to which they are relatively autonomous.

### 4.1. Internalization of Extrinsic Motivation

Within SDT, extrinsically motivated behaviors can become more autonomous through a process in which people internalize the values and regulations associated with the behaviors. Internalization is the process of taking in a value or regulation and making it one's own. When the process functions optimally, individuals will have transformed an externally regulated extrinsic motivation into an internally regulated extrinsic motivation by integrating the regulation and its value into their sense of self.

Internalization has long been an important concept in many psychological theories. The idea is that the challenge of becoming self-regulating of activities that are important for people but are not themselves personally compelling can be met through the process of internalization. To be self-regulating, people must make internal what was initially external. The thing that makes SDT's conceptualization of internalization different from most others is that it includes different types or degrees of internalization. In other words, SDT proposes that regulations can be internalized more or less fully, such that people will, to differing degrees, accept the regulations as personally important for themselves and, thus, be more or less autonomous in enacting them. As such, there will be differences in the extent to which people do a behavior because they want to rather than because they believe that they have to, and this will reflect the degree to which the regulation has been internalized.

Thus, SDT views internalization in terms of a continuum that describes how fully the person has transformed an external prompt into an internal regulation that will allow volitional or "choiceful" behavior. This continuum, therefore, illustrates how instrumental (i.e., extrinsic) motivation for a behavior can range from passive compliance to active personal commitment.

One reason why this relative autonomy continuum is so important is that it addresses an important problem about internal regulation of behavior. Specifically, there are behaviors that people force themselves to do because they think they should do them and know they

will feel guilty if they do not. The regulation of these behaviors is certainly internal to these people, but it does not exhibit the qualities of volition or autonomy that, for example, are so evident in intrinsic motivation. SDT uses the concept of introjection to refer to the type of internalization that leads to this internally pressured regulation. In contrast, there are behaviors that people do not find interesting but that have become meaningful and important for their own self-selected goals and personal life plans. As such, people do these behaviors quite autonomously, even though the behaviors themselves are not inherently satisfying. In SDT, this is said to occur as people identify with the importance of the activities for themselves and then integrate that identification with other aspects of themselves. When this has occurred, people will feel fully autonomous as they enact an extrinsically motivated behavior because it would then emanate from their integrated sense of self.

The former example, of introjection, is about controlling oneself, and the process bears considerable similarity to being controlled by other people. The latter example, in contrast, is more about making a choice to do the activity because, all things considered, doing so feels desirable and right. Thus, in the latter case, people experience a truly internal PLOC, even though the behavior is still instrumental, that is, done for reasons other than the enjoyment of the activity itself. This important contrast, between what might be thought of as self-control and self-regulation, is highlighted in the SDT model of extrinsic motivation.

More specifically, the theory states that there are various forms of extrinsic motivation: one where the regulation is external to the person, one where the regulation has been merely introjected and so must be buttressed by internal reward or punishment (e.g., guilt) contingencies, one where the person has identified with the importance of the behavior for himself or herself, and one where the regulation has been fully integrated with other aspects of the person's self. Although the latter three of these all represent internal motivation, they vary in the degree to which the behaviors they motivate are autonomous. Integrated regulation represents a relatively full sense of volition and personal commitment, with introjection representing a relative lack of these qualities (Fig. 1).

Intrinsic motivation and fully integrated extrinsic motivation are the two bases for autonomous or self-determined behaviors. More than three decades of research have now shown that the quality of people's experience and performance vary as a function of the

<i>Extrinsic Motivation</i>			
External regulation	Introjected regulation	Identified regulation	Integrated regulation

**FIGURE 1** Types of extrinsic motivation based on the degree to which a regulation and its underlying value have been internalized. They range from external regulation (least autonomous) to integrated regulation (most autonomous).

degree to which a behavior is autonomous or self-determined. Thus, when people have identified with and integrated the regulation of an extrinsically motivated behavior, the behavior shares many of the qualities of behaviors that are intrinsically motivated. Not only do people feel a sense of choice and experience an internal PLOC, but more importantly, autonomous extrinsic motivation, like intrinsic motivation, is positively related to psychological well-being as well as to learning outcomes and effective performance, especially on activities that require a deeper or fuller engagement with the activity.

#### **4.2. Internalization in Social Contexts**

Various studies have examined how the social conditions within which a regulation is internalized affect how fully it is internalized and, thus, how autonomous the subsequent behavior will be. One factor that affects how self-determined people will become for behaviors that were initially externally prompted is whether the behaviors were clearly valued by others to whom they feel connected. This suggests, for example, that the starting point for facilitating internalization in children, students, or employees is providing them with a sense of personal relatedness, that is, the feeling of being cared about. In other words, it is the basic psychological need for relatedness that leads people to be willing to take on the regulation of behaviors that are valued by significant others or relevant reference groups.

In addition, people must feel sufficiently competent at the focal behavior to do it satisfactorily; that is, people will be more likely to internalize a goal if they have the relevant skills and understanding to become successful at it. Thus, supporting competence by offering optimal challenges and providing effectance-relevant feedback will also facilitate internalization. Finally, for people to become autonomous with respect to a behavior, they must grasp its personal meaning for themselves; that is, they must have their own perspective acknowledged and

must feel a sense of choice about doing the behavior. This suggests that whereas people may at least partially internalize a behavioral regulation if they feel a sense of (or desire for) relatedness to a relevant other or group and if they feel effective in doing the behavior, they are unlikely to fully integrate the regulation unless the relevant others in their social world provide autonomy support with respect to that behavior. Notably, then, whereas supports for autonomy and competence are particularly important for maintaining intrinsic motivation for an activity, supports for relatedness are also important for internalization. In short, when people feel supports for relatedness, competence, and autonomy with respect to an extrinsically motivated behavior, they are likely to internalize its value and regulation quite fully.

In sum, self-determination, which is based in intrinsic motivation and integrated extrinsic motivation, has been associated with a variety of positive performance and adjustment outcomes and has been found to depend on interpersonal supports for relatedness, competence, and autonomy. In these regards, then, one can now see how the study of joy and delight provided a starting point for an examination of how to facilitate optimal motivation for even the mundane tasks that are encountered each day in people's lives. By providing the kinds of supports that help to maintain intrinsic motivation, one can also facilitate internalization and integration that will enhance the quality of motivation and engagement with the more mundane activities as well as the more interesting ones.

## **5. APPLICATION OF RESEARCH ON SELF-DETERMINATION**

In various domains of life, there are people who have responsibility for motivating others—their children, their students, subordinates, patients, athletes, and the like. This can be a daunting challenge. Because research has shown that autonomous motivation, based in intrinsic motivation and integrated extrinsic motivation, is unequivocally linked to superior conceptual learning, greater creativity, more flexible problem solving, and better psychological health and well-being, it seems clear that working to facilitate autonomous motivation, rather than taking the easier route of relying on controls, is indeed important.

Numerous studies have examined the effects of supporting autonomy versus controlling behavior in

a variety of real-world domains. We now turn to a brief consideration of a few of these studies in some of the more important domains of life.

### **5.1. Autonomy Support in Parenting**

Supporting autonomous motivation and promoting healthy development begin in the home. Parents have the task of providing a context that will allow their children to experience a sense of autonomy, competence, and relatedness, and it is those contexts that foster autonomous motivation, internalization of values, and psychological well-being.

Parents, of course, are genuinely concerned about their children acquiring the skills and values that are necessary for functioning effectively in school, on the playground, and in their subsequent careers and personal development. Parents are also hopeful that their children will behave appropriately in a variety of everyday settings so as not to bring rejection or embarrassment on themselves and their families. At times, parents' concerns and desires can prompt conflict because their expectations can conflict with the interests and desires of the children. Children are not always eager to do their homework, take out the garbage, or attend the religious services in which their parents are invested. This resistance may prompt an urge in parents to apply pressure in an attempt to ensure that their children do the behaviors that the parents believe are in the children's best interests. Research makes clear, however, that the children pay a cost in terms of their motivation and self-reliance to the extent that parents' motivational strategies involve pressure and control. Whether the controls are blatant or subtle, and whether they are coercive or seductive, they seem to have some negative psychological consequences for children.

One common approach to motivation and discipline used by many parents is conditional regard, that is, providing more attention and affection when children do what the parents value and withdrawing affection and regard when the children do not. This parenting strategy turns out to be quite controlling, however, and research has shown that it leads children to merely introject behavioral regulations rather than integrate them. In other words, children will tend to do the behaviors on which affection and regard were made contingent, but they will experience a sense of inner compulsion and anxiety about the behaviors rather than a sense of choice and desire. Furthermore, this pressured form of regulation is associated with poorer well-being—with lower and more fragile self-esteem,

greater feelings of guilt and shame, and the like. In addition, to the extent that parents use conditional regard as a socializing strategy, their children still harbor resentment toward the parents many years later. Other more overtly coercive forms of control have been found to have even more negative effects.

The alternative to pressure and control, of course, is autonomy support, which has been found to enhance motivation, performance, and well-being. In one interview study, for example, Grolnick and Ryan went into the homes of late elementary school-aged children. The researchers conducted separate interviews with mothers and fathers of the children regarding how the parents motivate the children to do schoolwork and to carry out chores around the house. Results of the study made clear that when parents were autonomy supportive—when they took their children’s perspective into consideration, encouraged initiative, and provided choice—the children were more autonomously motivated, felt better about themselves, and were rated by teachers as being more engaged in their schoolwork and more competent.

Autonomy support represents a challenge for parents. To try to understand children’s behavior from their perspectives can be a formidable undertaking when parents have their own perspectives and desires. To allow and support children’s choices when it is easier to impose parents’ own choices is asking a lot of parents, especially if they are stressed or tired from dealing with demands in other domains of their lives. Yet the evidence is clear that it is only through autonomy support that children will come to the point of accepting responsibility for the behaviors that are in their best interests. In other words, pressure and control can prompt compliance in the immediate situation (although it sometimes prompts rebellion), but even when it prompts compliance, it is a relatively ineffective socializing strategy because the behavior tends not to persist over the long term. Autonomy support, in contrast, promotes internalization and integration of the behaviors so that they become part of the children’s own values and sense of self and, thus, are more likely to be maintained over time.

It is very important to recognize that autonomy support is not the same thing as permissiveness. To be autonomy supportive does not mean to allow children to do anything they want. Rather, autonomy support involves the use of limits and discussions about the reasons why certain behaviors are important. The important thing, however, is the style in which that is done. When parents demand that their children accept

limits and punish them for not accepting these limits, the parents are creating a power struggle that is doomed to fail. For limits to work effectively, it is important to acknowledge the children’s point of view, for example, recognizing that they might not want to behave in the way that is being discussed. It is also important to provide a meaningful rationale for doing the behaviors and to refrain from using interpersonal pressure. It is the pressure, rather than the limits, that creates the problem. As such, autonomy support means being neither pressuring nor permissive; instead, it involves working with children to support their acceptance of the values and behaviors that are important for their own development and well-being.

## 5.2. Autonomy Support in Education

Several studies, using questionnaires, observational methods, and experimental designs, have examined the facilitation of motivation and performance in students. Studies have shown that there is substantial variability in public school classrooms regarding the degree to which teachers tend to support students’ autonomy versus control their behavior. Furthermore, evidence from research in elementary school classrooms indicates that when teachers are more supportive of students’ autonomy—when teachers provide meaningful explanations for why certain behaviors are important, allow students to experiment with alternative ways of doing things, and encourage students to take initiative and try solving their own problems—students display higher levels of intrinsic motivation, curiosity, and desire for challenge and also are more likely to internalize extrinsic motivation in a relatively full way. One study showed that within 2 months in classrooms, students had made an accommodation to the teachers’ styles such that students of relatively controlling teachers had become less intrinsically motivated. Because controlling methods have been related to poorer conceptual learning, poorer performance on problem-solving tasks, less creativity, and lower feelings of self-worth, the evidence is compelling that more autonomy-supportive approaches in classrooms have substantial advantages for learning, performance, and development.

During recent years, there have been some efforts to include the concept of supporting students’ and teachers’ autonomy within school reform approaches, and initial evidence suggests that this is an important element of effective reform. For example, the Institute

for Research and Reform in Education has carried out districtwide reform programs that have the concept of autonomy support as a central element, and these efforts have resulted in a variety of positive educational outcomes, including decreased dropout rates, enhanced engagement in classroom activities and homework, and improved performance on standardized tests.

At the college and medical school educational levels, comparable results have been reported. For example, in a study of undergraduates learning organic chemistry, results indicated that when instructors were perceived as more autonomy supportive, students became more autonomously motivated over the semester, resulting in better performance on a standardized course exam. Studies of medical students similarly showed that instructors who were more autonomy supportive had students who were more likely to maintain interest, internalize important values associated with the course, feel more competent, and perform better in the practical application of course material.

### **5.3. Autonomy Support in the Workplace**

Work groups, like classrooms, often vary in their climate or ambience. Some managers tend to encourage subordinates to participate in goal setting, performance evaluation, and other important decisions that are relevant to the subordinates' work lives, whereas other managers prefer to make the decisions by themselves and then tell the subordinates what needs to be done. The former approach tends to integrate the planning and doing of the work, with subordinates being involved in both, whereas the second tends to separate them, with the managers planning and the subordinates doing.

Research has shown that when managers have a more autonomy-supportive orientation toward their employees, the employees are more motivated, satisfied, and trusting and also perform more effectively than when managers have a more controlling orientation. In one study, researchers developed an intervention in which a change agent worked with managers in a major corporation to train the managers how to be more autonomy supportive in their interactions with subordinates. Results of the study indicated that over the period of the study, managers did become more autonomy supportive. More importantly, however, the subordinates of those managers who had been involved

in the training became more motivated and more satisfied with their work than did comparable employees who worked for managers who had not received the training. It appears that when managers become more autonomy supportive, the change in their orientation and behavior radiates to their subordinates, resulting in improved engagement and effectiveness of the organizational unit.

In other research, Deci and colleagues studied managers and their subordinates in Bulgaria and the United States to confirm whether the positive effects of autonomy support would appear in a collectivist totalitarian society (Bulgaria) as well as in an individualistic democratic society (United States). Results showed that when managers were autonomy supportive, their subordinates experienced greater satisfaction of the basic needs for competence, autonomy, and relatedness, leading to greater engagement in work and better psychological health. Thus, autonomy support seems to be important in the workplace regardless of cultural values or the economic and political systems that are operative in the culture.

### **5.4. Autonomy Support in Health Care**

The importance of supporting autonomy versus controlling behavior for the promotion of maintained health behavior change has also been demonstrated in several studies. Specifically, research has shown that health care practitioners who are more autonomy supportive are more successful in promoting healthy behavior among their patients. Because unhealthy behaviors—tobacco use, poor diets, lack of exercise, and the like—are responsible for well over half a million American deaths each year, facilitating change toward healthier behaving is clearly a critical issue.

Williams and colleagues have examined practitioner autonomy support with respect to a variety of health behaviors. For example, provider autonomy support has consistently been found to promote more autonomous motivation for healthier living and greater feelings of competence about making successful maintained change. In one study, these enhanced experiences of autonomy and competence among patients with diabetes resulted in physiologically indexed improvement in glucose control over a year-long treatment period. In another study, the enhancement of patients' autonomous motivation by autonomy-supportive physicians led patients to be more adherent to long-term

medication prescriptions. Because only approximately half of all prescriptions written by physicians in the United States are followed to a degree necessary for the medications to be effective, promotion of adherence is an important issue for health care. It seems that when providers engage patients as partners in the health care process by encouraging them to be more active in their own care, the patients become more autonomously motivated and are more successful in carrying out these difficult health-relevant behaviors.

Another study examined the autonomy support of providers in a weight loss program where morbidly obese individuals were on a medically supervised, very low-calorie diet. Results indicated that the degree of provider autonomy support made a significant difference. Specifically, patients who experienced the staff as more autonomy supportive reported more autonomous motivation for participating in the program, attended weekly program meetings more regularly, and lost more weight during the program. Furthermore, at a 2-year follow-up, they were exercising more regularly and were more effective in maintaining the weight loss.

In one study, autonomy-supportive principles were used as the central element in developing a medically based counseling intervention for smokers. Using more than 1000 smokers whose educations and incomes were significantly below those of the urban county where they lived, the study evaluated the new counseling intervention relative to the usual care for treating tobacco dependence that is available in the community. Results indicated that the new intervention was substantially more effective than typical community care and that it also linked provider autonomy support directly to patients' autonomous motivation for smoking cessation. This, in turn, was a significant predictor of the patients' 6- and 18-month biochemically validated smoking cessation.

### 5.5. Autonomy Support in Athletics

Field studies have also shown the importance of supporting autonomy and competence for enhancing autonomous motivation in the domain of sport and exercise. There is a popular image of coaches as tyrants who care only about winning and who pressure their players to work hard and win. Although that might not be the norm, there does appear to be a tendency toward pressure and control in sports. Many parents put strong pressure on their young children to win at Little League baseball, for example, and college athletics has become something of a professional endeavor, at least in

some circles. Professional athletics, of course, is increasingly tied into huge salaries, endorsement contracts, and intense pressure to win. Studies of competition and motivation in athletics indicate that this may be counterproductive, however, both for performance and for the health and well-being of athletes. For example, there is evidence that pressuring athletes to win decreases their intrinsic motivation for the sport, and athletic scholarships to college football players have been found to have a negative motivational impact.

An interesting study by Pelletier and colleagues showed that when coaches of competitive swimmers were autonomy supportive, the athletes tended to be autonomously motivated and, in turn, persisted at the sport over a 22-month period. On the other hand, when coaches were controlling, the athletes tended to be either external or introjected in their motivation. Externally motivated swimmers had a high rate of dropout from the sport by the time of a 10-month follow-up, and those who were introjected in their motivation tended to persist through the 10-month follow-up but evidenced a high rate of dropout by the 22-month assessment point. It seems clear, then, that although controlling approaches to motivation can bring about some behavior change, that change is unlikely to persist over the long term, whereas autonomy-supportive approaches are much more likely to promote maintained long-term change.

## 6. SUMMARY AND CONCLUSIONS

Autonomous motivation, which is the basis for self-determination in human behavior, consists of intrinsic motivation (i.e., doing an activity because it is interesting and does not require a separable reward) and integrated extrinsic motivation (i.e., doing an activity because it has become personally important for the person's self-selected values and goals). Both intrinsic motivation and integrated extrinsic motivation are facilitated in situations that are interpersonally supportive, that is, where parents, teachers, managers, health care providers, and coaches understand and acknowledge the perspectives of their children, students, employees, patients, and athletes, respectively, and encourage them to experiment, allow them to try their own solutions to problems, provide them with choices, and respond to their initiations.

Under conditions that are supportive of individuals' autonomy, those individuals are more engaged with the task and persist longer, perform more effectively

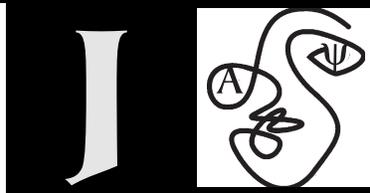
(especially on tasks that require greater creativity, flexibility, and conceptual understanding), and display higher levels of psychological health and well-being.

### See Also the Following Articles

Motivation and Culture ■ Work Motivation

### Further Reading

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- DeCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York: Academic Press.
- Deci, E. L., Betley, G., Kahle, J., Abrams, L., & Porac, J. (1981). When trying to win: Competition and intrinsic motivation. *Personality and Social Psychology Bulletin*, 7, 79–83.
- Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, 74, 580–590.
- Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, 62, 119–142.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627–668.
- Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, 52, 890–898.
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York: Appleton–Century–Crofts.
- Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic rewards: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, 28, 129–137.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66, 297–333.
- Zuckerman, M., Porac, J., Lathin, D., Smith, R., & Deci, E. L. (1978). On the importance of self-determination for intrinsically motivated behavior. *Personality and Social Psychology Bulletin*, 4, 443–446.



# Job Analysis, Design, and Evaluation

Robert Smither

Rollins College, Winter Park, Florida, USA

1. Job Analysis
2. Job Design
3. Job Evaluation
4. Other Issues in Job Analysis, Design, and Evaluation  
Further Reading

## GLOSSARY

**autonomy** The level of independence the worker has in deciding how to accomplish the tasks associated with the job.

**commensurable factors** Behaviors typical of different levels of skill and assigned a number that reflects those differences.

**critical incidents** Method of job analysis that studies jobs by having knowledgeable individuals provide real workplace examples of successful and unsuccessful job performance.

**feedback** The degree to which a worker receives information about his or her performance.

**growth need strength** A worker's need for personal growth through job performance.

**hygiene factors** Herzberg's term for conditions of work, including salary, management, work environment, and other factors.

**job** An ongoing set of cognitive or physical activities performed by one person that furthers organizational goals.

**job analysis** Identification of tasks or behaviors necessary to perform a job successfully.

**job design** Structuring of tasks and behaviors that creates a job.

**job elements** A method of analyzing the duties associated with a job.

**job enlargement** The redesign of a job to make it more interesting by adding duties.

**job enrichment** The redesigning of a job so that workers have more responsibility for their accomplishment, thereby motivating workers to accomplish their jobs at a higher level of performance.

**job evaluation** A means of determining the economic value of a job and the setting of salary levels.

**job grading** A job evaluation method that places jobs into levels or grades.

**job ranking** A method of job evaluation in which an expert ranks all jobs in an organization or a department in terms of the importance of their contribution to the overall organizational goals.

**job simplification** The belief that jobs should be designed so that any worker with the requisite skills could perform the job successfully by knowing the procedures to follow.

**motivators** Tasks and opportunities for advancement and professional growth associated with a job.

**physical abilities** An analytical approach that focuses on the strength, cognitive abilities, stamina, and other factors required by a specific job.

**point method** A method of job evaluation in which dimensions of a job are described by subject matter experts and considered the basis for assigning salary levels.

**skill variety** Abilities that a worker uses to accomplish a job.

**subject matter experts (SMEs)** People who are familiar with the requirements of a job, such as incumbents, former job performers, supervisors, and job analysts.

*task identity* The degree to which the worker sees the outcome of his or her efforts.

*task significance* The impact of a job on others, both inside and outside the organization.

In its most basic sense, a job is an ongoing set of cognitive or physical activities performed by one person that furthers organizational goals. Information about jobs comes from job analysis, which is the identification of tasks or behaviors, sometimes referred to as the knowledge, skills, and abilities, necessary to perform a job successfully. Job analysis is used for a variety of organizational purposes, including setting selection standards, designing departments and divisions, training, performance appraisal, job design, and job evaluation. Job design refers to the structuring of tasks and behaviors that creates a job and is often undertaken after a job analysis. Job evaluation is the procedure for determining the economic value of a job and the setting of salary levels.

## 1. JOB ANALYSIS

Researchers often trace the beginnings of job analysis back to the work of Frederick W. Taylor, an industrial engineer at Bethlehem Steel, at the beginning of the 20th century. Taylor's goal was to find what he considered the one best way to perform a specific job and then develop step-by-step procedures for jobholders to follow. Taylor's approach focused solely on the tasks associated with a job—he did not consider social, organizational, or motivational factors in studying jobs. Nonetheless, his studies resulted in significant improvement in efficiency and productivity at Bethlehem Steel.

Taylor's research was guided by the concept of job simplification, or the belief that jobs should be designed to minimize individual differences in workers so that any worker with the requisite skills could perform the job successfully by knowing the procedures to follow.

A few decades after Taylor's original studies, however, the context of most jobs had changed dramatically and job analysis became a more complex procedure than simply identifying tasks and the best way to perform them. For example, Taylor had assumed that worker motivation was primarily economically based. Later research, such as the Hawthorne bank wiring studies or studies regarding the longwall method of coal getting, suggested that factors other than economics affect worker performance.

This increased knowledge about the complexities of worker motivation affected the field of job analysis by broadening the area of study to motivational, social, and organizational considerations in addition to the tasks associated with performing the job. Recently, changing demographics of the workforce, and concerns about making jobs equally available to all persons who can perform them, further broadened the field of job analysis. Although the most basic model of job analysis still consists of identifying tasks, worker characteristics, and working conditions, this identification is now done in the context of a much broader range than the issues Taylor originally considered.

Information about the tasks or behaviors associated with a particular job can be collected in a variety of ways. Typical methods of collecting data include observation of the person performing the job; interviews with the job incumbent, the incumbent's peers, immediate supervisor, or other supervisory personnel with the organization; diaries and other forms of record keeping; and questionnaires. Sources for job analysis information can include, in addition to job incumbents and their supervisors, vendors, customers, or other organizational members who interact with the incumbent.

Throughout the years, researchers have developed a number of approaches for analyzing the duties associated with a job. These procedures, which focus on different aspects of the job, can be quite precise and detailed, and they can focus mainly on either tasks to be performed or characteristics needed to perform the tasks. The job elements approach, for example, focuses on the knowledge, skills, and abilities (KSAs) necessary to perform a job successfully. Information about the KSAs comes from job incumbents and supervisors, who are considered subject matter experts (SMEs). Other approaches take a more generic approach to analyzing jobs. The Position Analysis Questionnaire method, for example, relies on the assumption that there is an underlying taxonomy to all jobs, and that all tasks associated with a job can be classified into six categories.

Another approach to job analysis is to focus on the physical abilities associated with successful job performance. This approach considers strength, cognitive abilities, stamina, and other factors required by a specific job. Physical abilities job analysis has been particularly important in addressing gender issues in personnel recruitment and selection.

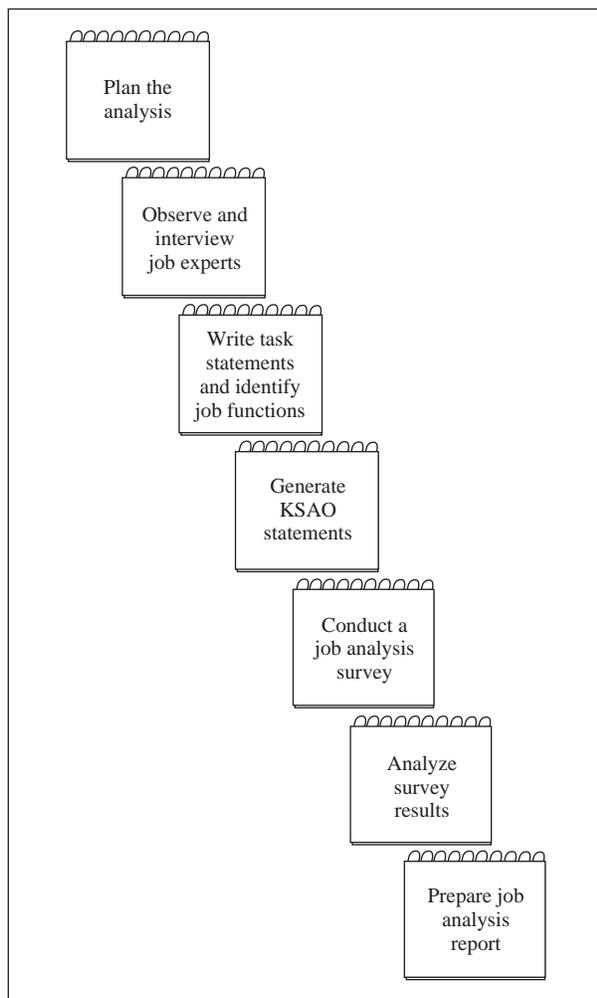
Finally, the critical incidents method of job analysis studies jobs by having incumbents and other knowledgeable individuals provide real workplace examples of successful and unsuccessful job performance. From

these examples, job analysts identify the abilities and behaviors necessary to perform a job successfully.

Although each of these methods takes a different approach to job analysis, the basic steps associated with analyzing a job are illustrated in Fig. 1. From the perspective of job design, the writing of task statements and identifying functions of a job is one of the most important parts of a job analysis. Job functions are groups of activities that relate to a particular purpose or outcome. Figure 2 lists some typical job

functions. Each job function consists of several tasks necessary to complete the job function. Tasks generally refer to an action that changes an input to an output.

From the identification of functions and tasks, the job analyst should be able to draw a fairly accurate picture of what a job entails. Nonetheless, a thorough job analysis also considers other important information that might affect performance, such as nonroutine tasks (tasks performed only occasionally), interactions with other individuals both inside and outside the organization, tasks performed as part of a team, and the position of the job within the overall organizational context.



**FIGURE 1** Steps in a job analysis. Reproduced by special permission of the publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, Florida 33549, from the *Job Analysis Kit Manual*, by Sandra McIntire, Ph.D., Mary Ann Bucklan, and Deonda R. Scott, Copyright 1995 by PAR, Inc. Further reproduction is prohibited without permission of PAR, Inc.

## 2. JOB DESIGN

Using the information gathered from a job analysis, jobs can be designed, redesigned, created, or eliminated. Most job design starts by determining where a job fits in the overall organizational process. One way to do this is to identify the precise steps a worker is expected to perform to change an input into an output—for example, to change a consumer's application for a mortgage loan into either a commitment or denial from a mortgage lender. From this information, the job designer can identify aspects of a job aside from the specific tasks and functions identified in the job analysis. These aspects may affect performance of a job's tasks and functions. In addition, knowing each step in the process of changing an input to an output allows the job designer to evaluate the expectations associated with the job.

However, as suggested previously, focusing solely on the major functions and tasks associated with a job may not provide a complete picture of the requirements for successful performance. Other factors that may affect job performance include (i) the technology the worker would be expected to use in performing tasks and functions (e.g., machinery vs computers), (ii) the environment in which the work is performed (e.g., indoors vs outdoors or in front of customers vs producing goods for inventory), and (iii) the structure of the work organization (e.g., assembly line vs team environment).

The final part of the job design considers matching the job functions and tasks to worker abilities. That is, the job design may be affected by the abilities of workers expected to perform the job. For example, an ideal job design may call for workers to handle both production and customer service duties, but it may be unrealistic to expect workers to have high levels of ability in both these areas. In this case, the job designer would need to

Administering operations	Filing, sorting, and routing
Analyzing competition	Financial planning and review
Analyzing financial data	Gathering information
Buying, ordering, and stocking	Handling outside contacts
Checking and testing	Improving work procedures and practices
Cleaning and maintaining	Maintaining and operating
Communicating orally	Machines/computers
Composing or editing	Managing the floor
Computing	Merchandising and displaying products
Controlling expenses	Operating the plant
Coordinating interdepartmental communications	Operating the store
Coping with difficulties and emergencies	Processing records
Developing and implementing technology	Promoting community relations
Developing and improving oneself	Promoting safe attitudes and practices
Developing employee potential	Relating to customers
Developing group cooperation and teamwork	Scheduling
Developing sales	Setting organizational objectives
	Supervising employees

**FIGURE 2** Some job functions. Reproduced by special permission of the publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, Florida 33549, from the *Job Analysis Kit Manual*, by Sandra McIntire, Ph.D., Mary Ann Bucklan, and Deonda R. Scott, Copyright 1995 by PAR, Inc. Further reproduction is prohibited without permission of PAR, Inc.

modify the design to reflect the practical reality of the abilities of the workers expected to perform the job.

Another consideration in job design mentioned earlier is the issue of fairness. Although it may be expedient to design jobs best suited to the mental or physical characteristics of a particular group, this kind of job design may not meet either the legal requirements or goals of the organization with regard to workforce composition. The job designer may need to modify the ideal design to reflect considerations other than performing the tasks associated with the job functions. Another important consideration in job design is worker motivation, which is discussed next.

## 2.1. Job Design and Worker Motivation

Three influential models of job design that consider the issue of worker motivation are the two-factor model, sociotechnical systems, and the job characteristics model.

### 2.1.1. The Two-Factor Model

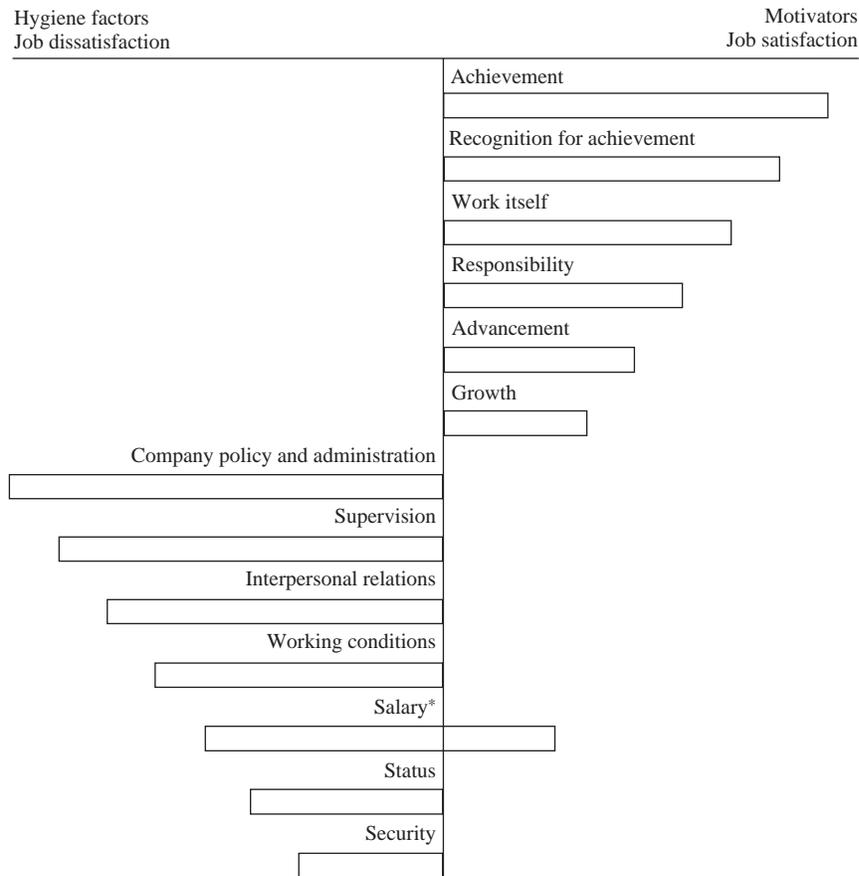
The job simplification model developed by Taylor was later challenged by Frederick Herzberg and associates, who argued that motivational factors are a critical aspect of job design. In Herzberg's formulation,

known as two-factor theory and illustrated in Fig. 3, different aspects of jobs could be categorized into what he called hygiene factors and motivators. Hygiene factors refer to the conditions of work and include salary, management, work environment, and other factors.

Motivators, on the other hand, include the tasks associated with the job as well as responsibility, autonomy, and opportunities for advancement and professional growth. According to two-factor theory, motivators have a more important impact on job performance than hygiene factors. Whereas employees may complain if they are dissatisfied with hygiene factors, their performance is more influenced by motivators. For this reason, job designers need to pay particular attention to motivators.

Herzberg's theory contradicted Taylor's belief that simplification should be the basis for designing jobs. By recognizing that performance is affected by factors other than tasks, Herzberg advocated an approach known as job enrichment. Job enrichment refers to designing jobs so that workers have more responsibility for their accomplishment. In this way, workers are more motivated to accomplish their jobs at a higher level of performance. Job enlargement is a kind of enrichment that attempts to make jobs more interesting by adding duties rather than adding responsibility.

Although studies of the two-factor model indicate that motivation and performance are more complex



**FIGURE 3** Motivators and hygiene factors. \*Because of its ubiquitous nature, salary commonly shows up as a motivator as well as a hygiene factor. Although primarily a hygiene factor, it also often takes on some of the properties of a motivator, with dynamics similar to those of recognition for achievement. From Herzberg, F. (1982). *The managerial choice: To be efficient and to be human* (2nd rev. ed.). Salt Lake City, UT: Olympus.

than the model suggests, Herzberg’s theory had significant intuitive appeal when applied to understanding worker motivation. From the perspective of job design, a shortcoming of two-factor theory is its lack of specificity in identifying how job tasks become motivators that affect performance.

**2.1.2. Sociotechnical Systems**

A second important influence on job design was developed at the Tavistock Institute of Human Relations in England. This model, developed during the period after World War II as Europeans introduced technology into their factories, argued that organizations consist of two systems—the social and the technical—that affect worker performance. According to sociotechni-

cal systems theory, jobs should be designed with both their social and their technical aspects in mind.

Sociotechnical theorists believe that introducing technological changes into a workplace almost always disrupts the social system. These disruptions, in turn, are likely to lower worker satisfaction and productivity. Therefore, a job design that does not consider the social aspects of a job is unlikely to result in optimal performance.

An important aspect of the sociotechnical approach to job design—and an important difference from job simplification—is its emphasis on work groups rather than individual jobs. Sociotechnical theorists believe that designing jobs in terms of work groups that have autonomy to plan how to accomplish their tasks results in both greater productivity and greater worker satisfaction. They also argue that increases in productivity

occur more often if workers develop multiple skills and if jobs have a variety of duties, which is a kind of job enrichment.

**2.1.3. The Job Characteristics Model**

Probably the most influential approach used by modern job designers is the job characteristics model developed by J. R. Hackman and Greg Oldham. Although Hackman and Oldham’s research focused on worker motivation, their identification of five domains associated with any job has been useful in the study of job design. Figure 4 illustrates the job characteristics model.

Skill variety refers to the abilities a worker uses to accomplish a job. According to the job characteristics model, jobs that require a wider range of skills are more motivating. Task identity refers to the degree to which the worker sees the outcome of his or her efforts, and task significance relates to the impact of the job on others, both inside and outside the organization. According to the theorists, the three aspects of skill variety, task identity, and task significance affect the worker’s assessment of the importance of his or her job.

The other two factors in the model address the responsibility associated with the job and the worker’s perception of outcomes. Autonomy refers to the level of independence the worker has in deciding how to accomplish the tasks associated with the job, and

feedback describes the degree to which the worker receives information about his or her performance.

A final aspect of the job characteristics model addresses an important individual difference variable that affects job performance. Growth need strength refers to the worker’s need for personal growth through job performance. Growth need strength can affect job design in that workers with high growth need strength seek jobs that offer challenge and a variety of duties; workers with low growth need strength prefer more routine types of jobs. If a job designer is dealing with jobs requiring a highly educated workforce, for example, growth need strength may require that jobs be made more complex and that levels of autonomy be higher.

**3. JOB EVALUATION**

As suggested previously, job evaluation is the procedure used to determine the relative worth of jobs within an organization. Most jobs are evaluated in terms of education required, skills involved in job performance, level of responsibility, working conditions, and factors in the local economy. Information from the evaluation is used to determine salary levels for jobs.

There are several methods for job evaluation. The point method is the most widely used. Under this procedure, subject matter experts identify job dimensions

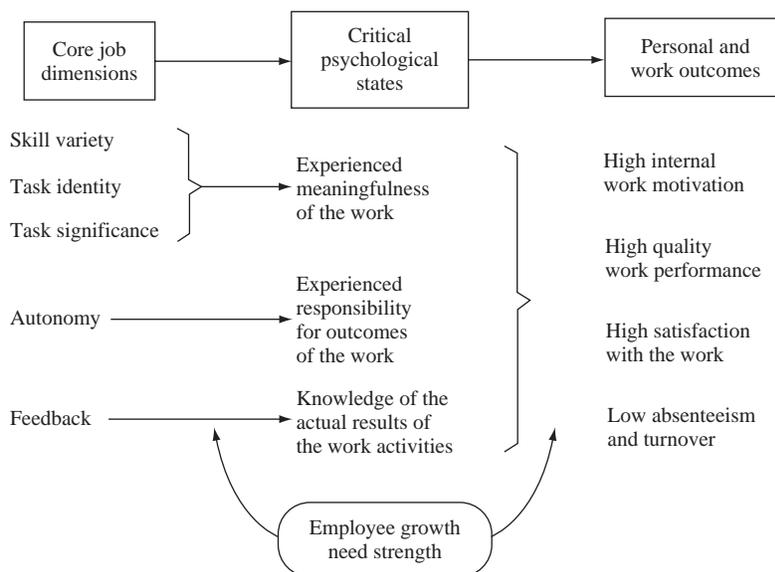


FIGURE 4 Job characteristics model. Reprinted from *Organizational Behavior and Human Performance*, 16, Hackman, J. R., & Oldham, G. R., Motivation through the design of work: Test of a theory, 250–279, Copyright 1976, with permission from Elsevier.

that will be the basis for assigning salary level. These dimensions—known as commensurable factors—are then described in terms of behaviors typical of different levels of skill and assigned a number that reflects differences in skill levels. Jobs can then be scored in terms of the number of points and compared to each other to determine salary level.

Job ranking is the most straightforward method of job evaluation and is most often used in small organizations. In this procedure, a SME ranks all jobs in an organization or a department in terms of the importance of their contribution to the overall organizational goals. In contrast to the point method, job ranking is a subjective procedure that does not consider the magnitude of the difference between jobs.

Finally, the job grading method places jobs into levels or grades. Jobs requiring less skill, for example, would be placed together—irrespective of the work performed—in a lower grade, whereas jobs requiring more skill would be grouped in a higher grade. The U.S. government's GS system is an example of job grading.

Regardless of the procedure used, job evaluation results in a hierarchical ranking of the worth of each job within the organization. The final step in determining salaries is external validation of pay scales. This is usually done through surveys of salaries paid for similar jobs in other organizations.

#### **4. OTHER ISSUES IN JOB ANALYSIS, DESIGN, AND EVALUATION**

Because of the complexities of modern organizations, considerations other than functions and tasks affect the way jobs are structured and salary levels assigned. Some issues that may influence the way an organization analyzes, designs, and evaluates jobs include individual differences, organization development, and teams.

##### **4.1. Individual Differences**

When Taylor first starting designing jobs at Bethlehem Steel, his goal was to standardize procedures for job performance so that any person could do the work. In general, standardization remains the basic principle of job design, which is that anyone with the requisite KSAs should be able to do the work.

However, as mentioned previously, job performance is greatly influenced by level of motivation, which is an individual difference variable. When

motivation to perform is high, jobs can be designed that encompass a variety of tasks and challenge the worker. When motivation is lower, jobs may need to be simpler, encompassing fewer tasks. Similarly, ability may also impact a worker's motivation and performance.

Another relevant individual difference variable is the tendency of individual workers to redesign their jobs so that they spend more time performing duties they enjoy and less time performing duties they find unappealing. In these cases, the job design provides a general outline for performance rather than a strict guideline. An employee may redesign a job so that he or she spends more or less time on certain tasks and still maintain a high level of performance overall.

##### **4.2. Organization Development**

During the 1990s, organizational theorists increasingly argued that the organization of the future would emphasize flexibility over structure. This argument was based on a variety of factors, but among the most important were changes in workforce demographics, technological innovations, and globalization.

As the U.S. workforce widened to include individuals from different races and cultural backgrounds, some theorists argued that traditional models of workplace organization were not appropriate for everyone. In particular, U.S. managerial practices were criticized for not taking differences such as race, culture, and gender into consideration. These criticisms, in turn, gave rise to the diversity movement. Diversity considerations may not affect the design of every job, but, at a minimum, they may be a factor in determining the physical requirements for performing the tasks associated with a job. For example, women, as a group, are not as physically strong as men, and Asians, as a group, are not usually as large as Westerners. Changes in the composition of the workforce may require job designers to consider the type of person most likely to do the job being designed as well as the tasks and functions associated with the job.

Technology is also likely to change job design. Although technology minimizes certain differences between workers (e.g., physical strength), technology often requires specific skills, such as facility with computers and computer programs, vigilance, and an ability to grasp the underlying model under which technology operates. A related challenge to traditional job design and evaluation is the changing nature of technology. Because technology changes rapidly in

some occupations, constant revision of the design and requirements of specific jobs may be necessary.

Finally, globalization affects job design in a variety of ways. Aside from confronting cultural differences, workers with foreign colleagues or who produce products for foreign markets may encounter factors such as language differences, time differences, incompatibilities in technology, or even political considerations that affect the way jobs are designed and evaluated.

### 4.3. Teams

During the 1980s, organizations began to pay more attention to the role of teams in the work environment. Although sociotechnical theorists had advocated teamwork as a means of both making work more meaningful to the worker and enhancing productivity since the 1950s, only a few U.S. organizations had adopted a teamwork approach. Building on an interest in Japanese quality circles that emerged in the 1970s, U.S. managers began to adapt team concepts to their workforces. In the modern team approach, teams often have responsibility for determining work schedules, methods, and hours, as well as team leadership and communication procedures, all of which can affect job design.

Teamwork calls for a radically different approach to job design. Rather than focus on requirements associated with a specific job, job design for teams requires the analysis of several jobs concurrently as well as the establishment of requirements based on the relationship of each job to other jobs represented by team members. This can be a complicated procedure, and most approaches to job design have not addressed the question of interrelationships among

jobs held by members of the same team. Similarly, organizing workers into teams responsible for accomplishing a particular task may complicate the job evaluation process.

### See Also the Following Articles

Motives and Goals ■ Organization Development ■ Work Motivation ■ Work Teams

### Further Reading

- Davis, L. E., & Wacker, G. J. (1988). Job design. In S. Gael (Ed.), *The job analysis handbook for business, industry, and government* (Vol. 1). New York: Wiley.
- Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, 16, 250–279.
- Herzberg, F., Mausner, B., & Snyderman, B. (1959). *The motivation to work*. New York: Wiley.
- McIntire, S. A., Bucklan, M. A., & Scott, D. R. (1995). *Job analysis kit*. Odessa, FL: Psychological Assessment Resources.
- Parker, S. K., & Wall, T. D. (2001). In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 1). London: Sage.
- Smither, R. D. (1988). *The psychology of work and human performance*. New York: Harper & Row.
- Trist, E. L. (1981). The evolution of sociotechnical systems as a conceptual framework and as an action research program. In A. H. Van de Ven, & W. F. Joyce (Eds.), *Perspectives on organization and behavior*. New York: Wiley.



# Job Search

Alan M. Saks

University of Toronto, Toronto, Ontario, Canada

1. The Job Search Process
  2. Job Search Research
  3. Job Search Behaviors
  4. Antecedents of Job Search Behaviors
  5. Consequences of Job Search Behaviors
  6. Conclusion
- Further Reading

## GLOSSARY

**active job search behavior** Applying for positions by sending out resumes and interviewing with prospective employers.

**assertive job seeking** The ability to identify one's rights and choices during job search and to act on them.

**coping resources** Individual and environmental factors and conditions that an individual can draw on in coping with involuntary job loss and that have direct and indirect effects on coping strategies.

**emotion-focused strategies** Attempts to change the way that the stressor is construed through techniques such as distancing and avoidance.

**emotional response model** Job seekers experience high levels of stress and frustration that can lead to avoidance, helplessness, and withdrawal.

**employment commitment** The importance and centrality of employment to the job seeker.

**employment outcome** Whether or not a job seeker obtains employment and the nature of that employment.

**employment quality** Job search outcomes that occur once the job seeker assumes a position and begins working, such as job satisfaction and intentions to quit.

**employment status** Whether or not a job seeker has obtained employment.

**formal sources** Public intermediaries such as advertisements, employment agencies, and campus placement offices.

**informal sources** Private intermediaries such as friends, relatives, or persons who are already employed in an organization.

**job-finding club** An intensive, highly structured, group-based behavioral counseling program that emphasizes motivation, maintenance of behavior, feedback and reinforcement, imitation, role playing, and practice activities.

**job information sources** The means by which job seekers learn about job opportunities.

**job search** A dynamic process through which individuals pursue employment goals.

**job search effort** The amount of energy, time, and persistence that a job seeker devotes to his or her job search.

**job search intensity** The frequency with which job seekers, during a set period of time, engage in specific job search behaviors or activities, such as preparing a resume or contacting an employment agency.

**job search outcomes** The immediate outcomes that occur during one's job search and lead to subsequent employment outcomes.

**job search self-efficacy** Confidence in one's ability to perform a variety of specific job search tasks and activities.

**job-seeking social support** The network of friends and family who provide counseling, assistance, and encouragement to job seekers.

**JOBS intervention** A job search intervention that includes problem-solving and decision-making processes, inoculation against setbacks, social support, and job-seeking skills.

**learning model** Job seekers learn more efficient and effective search techniques during their job search.

**networking** Contacting friends, acquaintances, and referrals to obtain information and leads about job opportunities.

**preparatory job search behavior** Gathering job search information and identifying potential leads during the planning phase of job search.

**problem-focused coping strategies** Activities directly aimed at eliminating the source of stress (i.e., job loss) such as proactive job search.

**sequential model** Job search follows a logical and systematic progression of stages.

**symptom-focused coping attempts** Attempts to eliminate the symptoms of stress through involvement in nonwork activities or seeking social support from friends and family.

Job search is a dynamic process through which individuals pursue employment goals. Whether one is entering the workforce for the first time, seeking reemployment following job loss, or changing jobs to pursue new career opportunities, job search is a necessary and inevitable part of the process. Numerous individual difference variables and situational factors combine to shape a person's job search. In turn, the nature and scope of an individual's job search can influence the probability and speed that one will obtain employment. For unemployed job seekers, job search is a form of problem-focused coping in response to job loss. An effective job search can increase an individual's chances of reemployment following job loss and improve long-term psychological well-being. Among employed job seekers, job search is a major predictor of voluntary turnover. The extent to which one engages in an effective job search and finds satisfactory employment can be enhanced through job search training interventions.

## 1. THE JOB SEARCH PROCESS

Job search is a process that consists of gathering information about potential job opportunities, generating and evaluating job alternatives, and choosing a job from the alternatives. These activities determine the type and amount of information that job seekers obtain about job openings as well as the number of job opportunities from which a job seeker has to choose.

The job search process has often been conceived of as a logical sequence of activities. Soelberg's model of job search consists of two phases: planning job search

and then job search and choice. Thus, job search begins with an extensive search to gather information and identify job opportunities followed by a more intensive search that involves the acquisition of specific information about jobs and organizations. Job search has also been described as consisting of a preparatory and active phase. Preparatory job search behavior involves information gathering about job opportunities, and active job search behavior involves applying for positions.

Kanfer *et al.* conceptualized job search as a self-regulatory process that involves "a purposive, volitional pattern of action that begins with the identification and commitment to pursuing an employment goal." Employment goals activate job search behaviors that are intended to obtain employment. Thus, "individuals identify, initiate, and pursue actions for the purpose of obtaining new employment or reemployment." Job search and the self-regulatory process ends when the employment goal has been achieved or is abandoned.

Finally, it is worth noting that during the search process job seekers change their behaviors. Three models describe how the behavior of job seekers can change during their job search. As indicated previously, the sequential model proposes that job search follows a logical and systematic progression of stages. According to the learning model, job seekers learn more efficient and effective search techniques during their job search. The emotional response model asserts that job seekers experience high levels of stress and frustration that can lead to avoidance, helplessness, and withdrawal.

Two studies have examined changes in job seeker's search behavior. Barber and colleagues measured the job search activities of college and vocational-technical school graduates early in the search process, at graduation, and 3 months following graduation for those who remained unemployed. They found a significant decrease from initial search to late search in the use of formal sources, job search intensity, and in information related to obtaining a job. They also reported a significant increase in the use of formal sources and intensity following late search for unemployed job seekers.

Saks and Ashforth studied recent university graduates who had not found a job just prior to graduation. Follow-up 4 months later indicated that job seekers increased their active job search behavior, formal job source usage, and search effort and decreased their job search anxiety.

## 2. JOB SEARCH RESEARCH

Research on job search can be distinguished in terms of the major area or domain of interest as well as the sample. Job search research has tended to fall into one of the following areas: (i) coping with job loss and reemployment, (ii) job search interventions, (iii) job search and voluntary turnover, and (iv) antecedents and consequences of job search.

The samples used in job search research have varied across and within the main research areas. For example, research on coping with job loss and reemployment necessarily includes unemployed and recently laid-off workers. Research on job search and turnover includes samples of employed job seekers. Research on the antecedents and consequences of job search has included new entrants to the workforce (e.g., college graduates), employed job seekers, as well as unemployed job seekers. However, it is worth noting that college students have been used most extensively in job search research, followed by unemployed job seekers.

### 2.1. Coping with Job Loss and Reemployment

Studies on job search and involuntary job loss focus on the coping strategies of recently unemployed individuals. Job loss coping strategies can be problem-focused, emotion-focused, and symptom-focused. Problem-focused coping strategies involve activities directly aimed at eliminating the source of stress (i.e., job loss) such as proactive job search. Emotion-focused strategies involve attempts to change the way that the stressor is construed through techniques such as distancing and avoidance. Symptom-focused coping attempts to eliminate the symptoms of stress through involvement in nonwork activities or seeking social support from friends and family.

Research on coping with job loss measures the predictors and consequences of coping strategies. Predictors of coping strategies consist of a number of coping resources, including individual and environmental factors and conditions that an individual can draw on in coping with involuntary job loss and that have direct and indirect effects on coping strategies. Coping resources such as education, financial resources, and social support predict cognitive appraisal of job loss and coping strategies, including job search activities. Reemployment coping goals and gender also predict proactive job search. Men engage in more problem-

focused coping strategies such as job search, whereas women engage in more symptom-focused strategies such as social support.

The consequences of coping behaviors include mental health and reemployment. Problem-focused coping (i.e., proactive job search) relates to obtaining reemployment but can also reduce mental health in the short term.

### 2.2. Job Search Interventions

Job search interventions can lessen the negative effects of job loss and improve individuals' job search and probability of reemployment. Research on employment counseling and job loss has found that the job-finding club is one of the most effective methods of job search training.

The job-finding club is an intensive, highly structured, group-based behavioral counseling program that emphasizes motivation, maintenance of behavior, feedback and reinforcement, imitation, role playing, and practice activities. Participants receive assistance in all areas of job search, including coping with discouragement, preparing resumes, obtaining and pursuing job leads, learning interviewing skills, scheduling time and record keeping, dress and grooming, and using the telephone for making inquiries and contacts.

Job seekers who have attended a job-finding club have been found to be more likely to obtain employment compared to individuals in a control group or who participated in other types of programs. Some studies have also found the job-finding club to result in higher starting salaries, more work hours, lower depression, and greater advancement and job satisfaction.

A number of other job search interventions have also been shown to be effective for coping with job loss and reemployment. For example, Caplan and colleagues developed a program called the JOBS intervention to improve the motivation and skills necessary for unemployed job seekers to engage in job-seeking behavior. The intervention includes problem-solving and decision-making processes, inoculation against setbacks, social support, and job-seeking skills.

Compared to a control group, the JOBS intervention resulted in higher rates of reemployment as well as higher quality reemployment in terms of earnings and job satisfaction for its participants. For those who remained unemployed, participants in the JOBS program have higher job search motivation and self-efficacy compared to control group participants. The benefits of the intervention for reemployment and employment quality have been shown to persist. Individuals who have participated in the JOBS program have higher levels

of employment, higher paying jobs, work more hours, and had fewer changes in jobs and employers up to 2½ years following the program. The JOBS intervention also has a significant effect on the mental health of individuals at high risk for depression.

Eden and Aviram studied the effects of a behavioral modeling job search workshop designed to boost the general self-efficacy (GSE) of unemployed vocational workers. Participants who attended the workshop had higher self-efficacy at the end the workshop and 2 months later compared to those in a control group. The workshop also increased the job search activity and reemployment of individuals with low initial GSE. Furthermore, the effect of the workshop on reemployment was mediated by job search activity.

Finally, a disclosive writing intervention in which laid-off professionals wrote about their thoughts and feelings about the layoff and how it affected their personal and professional lives resulted in higher levels of full-time employment compared to those for a control group.

### 2.3. Job Search and Voluntary Turnover

Models of turnover and the turnover process often include job search as a key predictor of voluntary turnover. Several studies have found that job search is highly predictive of turnover and even more strongly related to turnover than more traditional predictors such as perceptual, affective, attitudinal, and intention measures.

In terms of specific job search variables, Gary Blau has shown that active job search behavior is a better predictor of voluntary turnover than preparatory job search behavior and general search effort, and that preparatory job search behavior is indirectly related to turnover through active job search behavior. In addition, active job search behavior predicts voluntary turnover beyond job attitudes and withdrawal cognitions. Job search has also been found to predict the voluntary turnover of managers and to mediate the relationship between motivation variables and turnover.

Thus, it appears that job search, rather than turnover intentions, is the best predictor of voluntary turnover. Although this stream of research does not measure the employment status of those who quit, it is safe to assume that if an employee has quit he or she has found another job. Therefore, the finding that job search is related to turnover is akin to stating that job search is related to finding a new job.

### 2.4. Antecedents and Consequences of Job Search

In 1987, Schwab and colleagues presented a job search model in which self-esteem and financial need predicted job search intensity, and job search intensity predicted employment. Since then, additional antecedents, job search variables, and outcomes have been investigated. Given that this research represents the most extensive area of job search and overlaps with the other areas, the following sections focus on job search behaviors, antecedents, and consequences.

As indicated previously, research on the antecedents and consequences of job search has varied as a function of sample type (i.e., new entrants, employed, and unemployed job seekers). Since the characteristics of these samples differ in terms of experience, motivation, and the use of job search strategies, the antecedents and consequences of job search might differ across studies and samples. In fact, in a meta-analysis on the antecedents and consequences of job search conducted by Kanfer and colleagues, sample type was a significant moderator variable.

## 3. JOB SEARCH BEHAVIORS

In their review of the job search literature, Schwab and colleagues indicated that job search success and employment outcomes are a function of the sources used to acquire information about job openings and the intensity with which one pursues such information. Although sources and intensity are among the most often studied job search dimensions, job search research has included a number of other dimensions of job search, the four most prominent being job information sources, job search intensity, job search effort, and job search information.

### 3.1. Job Information Sources

Job information sources is considered to be one of the most important dimensions of job search and is among the most often studied. Sources refer to the means by which job seekers learn about job opportunities. Both the recruitment and the job search literature have tended to distinguish between formal and informal job sources. Formal sources involve the use of public intermediaries, such as advertisements, employment agencies, and campus placement offices. Informal sources are private intermediaries, such as friends,

relatives, or persons who are already employed in an organization.

### 3.2. Job Search Intensity

Job search intensity refers to the frequency with which job seekers, during a set period of time, engage in specific job search behaviors or activities, such as preparing a resume or contacting an employment agency. Various indicators have been used to measure search intensity, such as the number of employers contacted, the number of hours per week spent searching, and the number of sources used.

In response to the failure to distinguish between job search activities at different stages of the job search process as well as the frequent use of single-item measures and dichotomous response scales, Gary Blau developed two distinct measures of job search intensity that he called preparatory job search behavior and active job search behavior. Preparatory job search behavior involves gathering job search information and identifying potential leads during the planning phase of job search. Active job search behavior involves the actual job search and choice process, such as sending out resumes and interviewing with prospective employers. It reflects an individual's behavioral commitment to job search.

Several studies have measured specific job search behaviors that have been recommended in the popular job search literature, particularly assertiveness and networking. Assertive job-seeking applies the concept of assertiveness to job search and refers to the ability to identify one's rights and choices during job search and to act on them. Specific behaviors include making follow-up calls regarding the status of a job application and making calls to arrange meetings with organizational representatives to discuss employment opportunities. Networking is another specific job search behavior and involves contacting friends, acquaintances, and referrals to obtain information and leads about job opportunities. Networking involves the use of informal job information sources.

### 3.3. Job Search Effort

Job search effort refers to the amount of energy, time, and persistence that a job seeker devotes to his or her job search. Unlike measures of job search intensity, they do not focus on particular job search behaviors. Blau developed a scale to measure general effort in job search that asks respondents about the time and effort they have devoted to their search. Based on confirmatory factor

analyses across these diverse samples, he found empirical support that general effort is related to, but distinguishable from, preparatory and active job search behavior.

### 3.4. Job Search Information

Barber and colleagues argued that the type of information sought and acquired is also an important dimension of job search because individuals use this information to make job choice decisions. In addition to information about employment opportunities, job seekers also require accurate, detailed, and specific information about job and organizational attributes. Few studies, however, have actually examined this dimension of job search.

## 4. ANTECEDENTS OF JOB SEARCH BEHAVIORS

A number of variables have been found to influence job search and employment outcomes. In general, the antecedents of job search can be categorized into three groups: biographical/demographic variables, individual difference variables, and situational factors.

However, the size of the relationships between the antecedents and job search behavior has been found to differ as a function of the job search measure and the sample type. In a meta-analysis, Kanfer and colleagues concluded that a number of antecedents–job search relationships were stronger for measures of job search intensity compared to measures of job search effort. Also, for some antecedents (i.e., extroversion, neuroticism, conscientiousness, and education) the relationships were stronger for new entrants, whereas for other antecedents (i.e., self-esteem, employment commitment, and social support) the relationships were stronger for unemployed job seekers. In addition, tenure related positively to search behavior for new entrants but related negatively to search behavior for employed and unemployed job seekers. It is also worth noting that the relationships between the antecedents and search behavior tend to be stronger than the relationships between the antecedents and employment outcomes.

### 4.1. Demographic/Biographical Variables

Demographic/biographical variables in job search research include age, gender, education, race, and tenure.

In general, biographical variables tend to be only weakly related to job search behavior and to a lesser extent than the other antecedents. In their meta-analysis, Kanfer and colleagues concluded that age, race, gender, education, and tenure were related to job search behavior; that is, older, nonwhite, female, less educated, and longer tenured individuals engaged in less job search behavior than younger, white, male, more educated, and less tenured individuals. Furthermore, men have been found to be more likely to use informal sources than women.

Demographic variables are also related to job search outcomes. Younger job seekers as well as more educated ones and males (new entrants only) are more likely to find employment, and more educated and white job seekers experience a shorter period of unemployment.

Finally, there is evidence that cognitive ability may also play a role in job search. For example, the cognitive ability of employed managers has been shown to relate positively to job search. Also, grade average, a surrogate measure of cognitive ability, has been found to correlate positively to active and preparatory job search behavior as well as employment status among a sample of recent graduates.

#### 4.2. Individual Difference Variables

Individual difference variables tend to be personality variables (e.g., self-esteem), motivational factors (e.g., self-efficacy), and attitudes toward employment and work (e.g., employment commitment). Among the personality variables, self-esteem has been found to relate negatively to the use of formal job sources and relate positively to job search intensity and assertive job-seeking behavior, as well as a shorter period of unemployment, the number of job offers received, and the likelihood of obtaining employment.

There is evidence that the dimensions of the five-factor model of personality are related to job search behaviors. In their meta-analysis, Kanfer and colleagues organized personality constructs used in job search research on the basis of the five-factor model of personality. They found that extroversion and conscientiousness were the strongest positive predictors of job search behavior, followed by openness to experience and agreeableness. Neuroticism related negatively to job search behavior. In addition, higher levels of extroversion, conscientiousness, openness to experience, and agreeableness related to a shorter period of unemployment, and conscientiousness was also related to employment status.

A motivational variable that has received a considerable amount of attention is job search self-efficacy—that is, confidence in one's ability to perform a variety of specific job search tasks and activities. The results of several studies indicate that job search self-efficacy predicts preparatory and active job search behaviors, assertive job search behavior, job source usage, and search intensity, as well as search duration, number of job offers, and employment status.

Another motivational variable that has been shown to be important is job seekers' perceived control over the outcomes of their job search. Although locus of control has been found to relate only weakly to job search behavior, perceived situational control over the outcomes of job search has been found to relate positively to job search behaviors and employment status.

In terms of employment attitudes, employment commitment or the importance and centrality of employment to the job seeker has been found to relate positively to job search intensity. Among employed job seekers, job attitudes such as job satisfaction and organizational commitment have been found to relate negatively to job search behavior.

#### 4.3. Situational Variables

Among the situational variables, financial need and social support have received a considerable amount of attention. Job seekers who have a greater financial need tend to be more intense in their search for employment. Financial need has been shown to relate positively to both preparatory and active job search behaviors. Job seekers with higher benefit levels or a longer duration of benefits tend to remain unemployed for longer periods. Thus, financial hardship tends to result in a shorter period of unemployment.

One of the most important situational variables in job search and coping with job loss is social or job-seeking support. Social support refers to the network of friends and family who provide counseling, assistance, and encouragement to job seekers. Social support has been shown to be particularly important in assisting individuals following job loss and to predict job search behavior and employment status.

Finally, several other situational factors have been found to predict the job search behavior of employed individuals. In particular, compensation, perceived organizational success, and organizational work–family balance policies have been found to relate negatively to job search. Job insecurity and desire for more work–family balance have been found to relate positively to

job search. These results are particularly noteworthy because they suggest that organizations have some influence over a number of job and organizational variables that affect employees' job search behavior.

## 5. CONSEQUENCES OF JOB SEARCH BEHAVIORS

In their model of job search, Schwab and colleagues proposed two outcomes of job search: employment and the quality of employment (e.g., job satisfaction and tenure). Job search research, however, has focused primarily on preentry outcomes, especially employment status (i.e., whether or not the job seeker has obtained employment). In general, the consequences of job search can be categorized in terms of (i) job search outcomes (outcomes that occur during the search process), (ii) employment outcomes (outcomes that are a result of one's job search), (iii) psychological well-being (the job seeker's mental health), and (iv) employment quality (outcomes that occur on-the-job).

### 5.1. Job Search Outcomes

Job search outcomes refer to the immediate outcomes that occur during one's job search and lead to subsequent employment outcomes. This includes the number of job interviews and job offers that a job seeker receives as well as the speed with which one obtains employment. Research supports a positive relationship between job search intensity and the number of job interviews and offers received and a negative relationship with search duration.

### 5.2. Employment Outcomes

Employment outcome measures assess the result of one's job search and refer to whether or not a job seeker obtains employment and the nature of that employment. The most common outcome measure of job search is employment status. There is strong support for a positive relationship between job search behavior and employment status. Furthermore, as indicated previously, research showing a relationship between job search intensity and turnover provides additional support that job search results in subsequent employment. Job search behavior has also been found to relate positively to preentry person–job (P–J)

fit and person–organization (P–O) fit and the initial compensation of recent graduates.

Interestingly, Kanfer and colleagues found that measures of job search effort related more strongly to employment duration and employment status compared to measures of job search intensity. In addition, the relationship between job search behavior and employment status was strongest for employed job seekers, followed by new entrants and unemployed job seekers.

Research on job sources has consistently found that job seekers are more likely to find employment through the use of informal sources of information. In particular, friends and personal acquaintances tend to be the main source through which job seekers obtain employment.

However, although informal sources may be more likely to result in employment, there is evidence that formal sources are more likely to result in a better job. For example, graduates who used formal sources have been found to be more likely to find work in higher level occupations and jobs that are more closely related to their college training than those who used informal sources. Also, the number of formal sources used by recent graduates has been found to be positively related to preentry perceptions of P–J and P–O fit. Formal sources have also been found to result in higher paying jobs.

### 5.3. Psychological Well-Being

A major concern in the job search literature is the psychological well-being of job seekers. This is especially the case in research on job loss, unemployment, and job search interventions. It is well known that the experience of job loss and unemployment is associated with negative physical and psychological effects, including symptoms of psychiatric disorder, distress, and depression.

Unlike the other outcomes discussed in this section, indicators of psychological well-being tend not to be directly associated with job search behavior but, rather, the experience of (un)employment. In this regard, job seekers who obtain satisfactory employment tend to show an improvement in their mental health and psychological well-being. However, given that job search is related to (re)employment, it is fair to say that job search behavior is indirectly related to psychological well-being through its influence on employment status.

Finally, job search interventions have also been shown to influence job seekers' psychological well-being. A number of such interventions have improved unemployed job seekers' self-esteem, self-efficacy, depression, and mental health.

### 5.4. Employment Quality

Employment quality refers to job search outcomes that occur once the job seeker assumes a position and begins working. With the exception of research on recruitment and job sources, job search research has seldom measured the quality of employment. Research on unemployment has also tended to neglect the quality of reemployment.

The relationship between recruitment sources and postentry outcomes has been extensively studied in the recruitment literature. Accordingly, job seekers who find employment through informal sources tend to have more positive job attitudes and lower turnover.

Job search research, however, has only recently begun to examine the links between job search behavior and employment quality, and the results have not been very supportive. For example, Wanberg and colleagues did not find a significant relationship between job search intensity and job satisfaction, job improvement, and intentions to turnover. In another study, they found that neither job search nor networking intensity related to job satisfaction or intention to turnover. Werbel also failed to detect a significant relationship between job search intensity and job satisfaction. Furthermore, research on involuntary job loss has also failed to detect a significant relationship between problem-focused coping and the quality of reemployment.

However, there is evidence that job search behavior is indirectly related to employment quality. Saks and Ashforth found that job search behavior was indirectly related to job and organizational attitudes through a direct relationship with preentry perceptions of P-J and P-O fit. In a related study, they found that preentry fit perceptions mediated the relationship between job information sources and job satisfaction, intentions to quit, and turnover.

## 6. CONCLUSION

Job search is a major factor in the experience of job loss, turnover, psychological well-being, and (re)employment. Both individual differences and situational factors influence job search behavior, and the nature and scope of one's job search are related to the probability, speed, and quality of employment that one obtains. Given that workers today are expected to experience more than a dozen job transitions during the course of their careers, the need for individuals to have job search skills is

critical for job seekers' continued employment and psychological well-being. Fortunately, the knowledge exists for career counselors to assist job seekers in coping with job loss and finding satisfactory (re)employment.

### See Also the Following Articles

Coping ■ Employment Interviewing ■ Job Analysis, Design, and Evaluation ■ Occupational Choice ■ Occupational Psychology, Overview ■ Social Networks

### Further Reading

- Barber, A. E., Daly, C. L., Giannantonio, C. M., & Philips, J. M. (1994). Job search activities: An examination of changes over time. *Personnel Psychology, 47*, 739-765.
- Blau, G. (1993). Further exploring the relationship between job search and voluntary individual turnover. *Personnel Psychology, 46*, 213-330.
- Blau, G. (1994). Testing a two-dimensional measure of job search behavior. *Organizational Behavior and Human Decision Processes, 59*, 288-312.
- Caplan, R. D., Vinokur, A. D., Price, R. H., & van Ryn, M. (1989). Job seeking, reemployment, and mental health: A randomized field experiment in coping with job loss. *Journal of Applied Psychology, 74*, 759-769.
- Eden, D., & Aviram, A. (1993). Self-efficacy training to speed reemployment: Helping people to help themselves. *Journal of Applied Psychology, 78*, 352-360.
- Granovetter, M. S. (1995). *Getting a job* (2nd ed.). Chicago: University of Chicago Press.
- Hanisch, K. A. (1999). Job loss and unemployment research from 1994 to 1998: A review and recommendations for research and intervention. *Journal of Vocational Behavior, 55*, 188-220.
- Kanfer, R., Wanberg, C. R., & Kantrowitz, T. M. (2001). Job search and employment: A personality-motivational analysis and meta-analytic review. *Journal of Applied Psychology, 86*, 837-855.
- Latack, J. C., Kinicki, A. J., & Prussia, G. E. (1995). An integrative process model of coping with job loss. *Academy of Management Review, 20*, 311-342.
- Leana, C. R., & Feldman, D. D. (1992). *Coping with job loss*. New York: Lexington.
- Saks, A. M., & Ashforth, B. E. (1997). A longitudinal investigation of the relationships between job information sources, applicant perceptions of fit, and work outcomes. *Personnel Psychology, 50*, 395-426.
- Saks, A. M., & Ashforth, B. E. (2000). Change in job search behaviors and employment outcomes. *Journal of Vocational Behavior, 56*, 277-287.
- Saks, A. M., & Ashforth, B. E. (2002). Is job search related to employment quality? It all depends on the fit. *Journal of Applied Psychology, 87*, 646-654.

- Schwab, D. P., Rynes, S. L., & Aldag, R. J. (1987). Theories and research on job search and choice. In K. M. Rowland, & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 5, pp. 129–166). Greenwich, CT: JAI Press.
- Soelberg, P. O. (1967). Unprogrammed decision making. *Industrial Management Review*, 8, 19–29.
- Wanberg, C. R. (1997). Antecedents and outcomes of coping behaviors among unemployed and reemployed individuals. *Journal of Applied Psychology*, 82, 731–744.
- Wanberg, C. R., Kanfer, R., & Rotundo, M. (1999). Unemployed individuals: Motives, job search competencies, and job search constraints as predictors of job seeking and reemployment. *Journal of Applied Psychology*, 84, 897–910.



# Job Stress

James Campbell Quick and  
Marilyn Macik-Frey

University of Texas, Arlington, Texas, USA

Debra L. Nelson

Oklahoma State University, Stillwater, Oklahoma, USA

1. Introduction
  2. Causes of Job Stress
  3. Consequences of Job Stress
  4. Modifiers of the Stress Response
  5. Measures of Job Stress
  6. Early Warning Signs of Strain
  7. Preventive Stress Management
  8. Implications for Executives and Leaders
- Further Reading

*tertiary prevention* The treatment or caregiving intervention to heal a person's strain.

*trauma* A highly stressful, deeply disturbing, or dramatically unexpected stressful event experienced as traumatic by the individual.

Job stress is the psychophysiological arousal resulting from workplace demands. Optimally, it results in enhanced job performance. When mismanaged, job stress leads to job strain, which includes psychological, medical, and behavioral costs.

## GLOSSARY

*eustress* The healthy, positive, constructive outcome of stressful events and the stress response.

*job stress* The mind-body arousal resulting from physical and/or psychological stressors associated with a job and work.

*preventive stress management* An approach and set of principles for promoting eustress and health while preventing strain and dysfunction.

*primary prevention* The modification, reduction, or management of sources of stress.

*secondary prevention* The modification, reduction, or management of the individual's stress response.

*strain* The physiological, psychological, and/or behavioral deviation from an individual's healthy functioning.

*stress response* The generalized, patterned, unconscious mobilization of the body's natural energy and resources.

*stressor* A physical or psychological stimulus that causes the individual stress response.

## 1. INTRODUCTION

Job stress is important for two reasons, one economic and one humanitarian. Direct and indirect organizational costs result from mismanaged job stress. In addition, mismanaged job stress causes physical and emotional suffering in the workplace. The National Institute for Occupational Safety and Health (NIOSH) identified job stress and psychological disorders among the top 10 occupational health hazards in the United States and the American Institute of Stress determined that job stress reached epidemic proportions in the by the end of the 1990s. When life expectancy at birth is used as the standard, Americans are passing the stress test. However, although Americans are living longer, they are suffering more. Anxiety disorders affect one in every six people in

the United States and one in every five employed people in the United Kingdom. In 2001, 35% of Americans reported that jobs were harming their physical and emotional health. Layoffs and downsizing during the 1990s had adverse effects on employee, managerial, and executive stress. Translated literally from the Greek, epidemic means "upon the people." Epidemiology is the basic science and fundamental practice of public health and preventive medicine. Treatment alone is rarely effective in the management of epidemics. Prevention is the preferred mode of epidemic management.

## 2. CAUSES OF JOB STRESS

Job stress is caused by demands within the work environment or by contextual or nonwork demands. The major categories of job stressors are physical demands, task demands, role demands, and interpersonal demands. In addition, extraorganizational stressors (e.g., marital discord) and transitional factors (e.g., preparing for retirement) can have spillover effects into the workplace, in part through role conflicts. Understanding an individual's stress and strain requires examining the entire life experience (i.e., work plus nonwork).

Physical stressors vary by occupation and working environment. Extreme work conditions, such as in tropical or arctic work environments, are a source of physical stress, as are strenuous working conditions, such as in industrial construction or steel making. Business travel is a major physical stressor for those who must engage in this activity.

Task stressors include a wide range of job content and job context factors. Work overload, time pressure, and changing tasks or activities are among the greatest sources of job stress for people in many occupations and organizations. Four other task-related sources of job stress are lack of control, career progress, future job uncertainty, and new technologies. Lack of control and lack of decision latitude are major contributors to job strain, especially when work demands are very high for a particular job. Blocked career progress leads to strain from frustration of natural achievement drive and ambition, whereas future job uncertainty creates strain from the anxiety of job loss. New technologies cause job stress due to the requirement for developing one's abilities and learning new skills.

Role stressors are among the best understood sources of job stress. The two major categories of roles stressors are role conflict and role ambiguity. Role conflict may emerge in a variety of ways. Interrole conflict occurs

when the expectations of different roles (e.g., the role of employee and the role of wife) are in conflict with each other. Intrarole conflict occurs when the expectations from different role senders (e.g., one's supervisor and one's colleague) are in conflict with each other. A variation of intrarole conflict occurs when one role sender, such as one's supervisor, sends conflicting expectations. Person-role conflict occurs when one's personal beliefs or values are in conflict with the expectations of one's job. Role ambiguity occurs in several forms. Ambiguity may arise when it is unclear what the consequences might be for failure to meet others' expectations, or it may arise as to exactly what others' expectations really are or what they really mean.

Interpersonal stressors are somewhat broader and more personal than role stressors. Leadership styles, abrasive personalities, organizational politics, poor diversity management, and sexual harassment give rise to interpersonal stress on the job. The two most stressful leadership styles for people at work are rigid, autocratic leadership behaviors and *laissez-faire*, or very passive, leadership behaviors. Abrasive personalities can be the source of intense interpersonal stress and emotional pain on the job. Organizational politics is a source of job stress, with some research showing that it has a greater adverse impact on women than men. Poor diversity management leads to job stress for the minority worker who feels there are unequal workplace barriers to success. Sexual harassment is a major job stressor, most often for women. In addition to being significant interpersonal stressors, diversity issues and sexual harassment can lead to tens of millions of dollars in legal costs and settlements.

Nonwork stressors, such as marital and family expectations, child-rearing responsibilities, and parental care, create role conflicts for people on the job and have spillover effects into the work environment. Traumatic events outside of the workplace, such as the loss of loved ones or the reality of terrorism and war, also may have spillover effects into the workplace, thus magnifying any job stress for the individual.

## 3. CONSEQUENCES OF JOB STRESS

The two broad categories of job stress consequences are functional (eustress, or healthy stress) and dysfunctional (strain). The former includes both health and performance benefits. The health benefits of well-managed job stress include enhanced focus during peak work activities and workplace emergencies, as well as

cardiovascular efficiency from aerobic fitness. The performance benefits include enhanced activity level and arousal with their accompanying bursts of physical strength and peak performance. The state of active engagement in meaningful work that produces eustress represents the healthy side of stress consequences. Workers who experience eustress can translate their experience into benefits for the organization, including productivity, organizational agility, and efficiency.

The dysfunctional consequences of job stress are individual and organizational strain. Because job stress triggers a systemic response within individuals and organizations, individual and organizational strain may manifest in a wide variety of ways. Job strain's manifestation is importantly influenced by the particular vulnerability of the individual or weak link in the organization. This is commonly called the Achilles' heel phenomenon or weak organ hypothesis. Some individuals, for example, tend to respond to stress via the respiratory system. They may find that their allergies worsen during periods of high stress. In some organizations, the weak link may be ethical behavior, with more unethical actions occurring during periods of high stress. For both individuals and organizations, it is at the point of the Achilles' heel that many strain symptoms emerge.

Individual strain can take the form of psychological disorders, medical illnesses, and/or behavior problems. Anxiety and depression are two of the more common psychological consequences of strain. Individuals may also experience family problems and sleep disturbances related to work strain. Within the services sector, individuals are particularly prone to burnout, another psychological consequence. Among the medical consequences of strain are heart disease and stroke, cancer, diabetes mellitus, and emerging diseases such as HIV and chronic fatigue syndrome. Behavioral consequences include tobacco and alcohol abuse, eating disorders, and violence.

Because of the exchange relationship between employees and organizations, the consequences of strain for employees translate into costs for the organization. Organizational strain can take the form of direct costs and/or indirect costs. Direct costs include strikes and work stoppages, declining quality and quantity of performance, compensation awards, and increased health care costs. Among the indirect costs of distress are employee dissatisfaction, communication breakdowns, faulty decision making, and diminished quality of work relationships.

#### 4. MODIFIERS OF THE STRESS RESPONSE

The stress response is the psychophysiological pattern that results when an individual is exposed to stressors. The stress response pattern is universal; the intensity, frequency, and consequences of the stress response are not. The ways in which the stress response is exhibited in any individual are moderated by individual and interpersonal factors, including interpersonal communication abilities, personality differences, emotional intelligence, and gender differences. The interaction of these variables with the stress response determines whether the stressor results in positive outcomes (eustress) or negative outcomes (strain). Although many of these moderators are inherent traits, most can also be consciously changed through prevention efforts.

Poor interpersonal communication has been cited as one of the precursors to strain. Therefore, an individual's ability to effectively communicate when faced with stressors can positively influence the outcomes of the stress. Interpersonal communication can be in the form of effective management of difficult work situations, people, or role conflicts. It can also involve effective use of communication forums to gain social support and to positively funnel stress-induced emotions. Effective interpersonal communication skills are in some cases inherent skills related to personality variables, but they can also be acquired through active learning of productive communication processes.

Personality differences also modify the individual's response to stressors and the consequences of the individual's response. Personality affects the person's perception or appraisal of a specific situation as stressful and also affects the person's response to the perceived stressor. Although multiple personality traits influence the stress response, the most often cited is Friedman and Rosenman's type A vs type B behavior patterns. Research has consistently shown that type A individuals exhibit more distress than their type B counterparts. The type A individual's high-intensity work level, incessant need for achievement, inability to relax, and general tendency to lack a balanced life approach result in both a higher number of situations being perceived as stressful and poor coping responses when stressors occur. Another perspective on personality and stress is illustrated by the Myers-Briggs Type Indicator. This popular measurement tool for personality traits depicts a combination of traits that describe how a

person perceives the world, processes information, and makes decisions. The individual's personality type explains why one person who is comfortable with less structure, independent work, and frequent changes may have a more positive response to an organization in flux than an individual who prefers structure, stability, and has a strong need for closure. Two other personality variables linked to the stress response are hardiness and self-reliance. Hardiness is defined as responding to stressful situations with commitment (vs alienation), control (vs powerlessness), and challenge (vs threat). Self-reliance is defined as the ability to form and maintain multiple healthy and mutually supportive interpersonal attachments that are used appropriately to deal with stressors. Some personality traits may be malleable, such as "learned optimism," and thus can be developed in order to help individuals manage stress more effectively.

Emotional intelligence (EI) includes a set of competencies that refer directly to the way an individual deals with emotions, a key determinant of how they also deal with stress. The components of EI include the abilities to be self-aware of emotions, to express them appropriately, to judge sincere vs insincere expressions of emotions, to distinguish and prioritize different emotions, and to assess problems from multiple emotional perspectives. EI components also include the ability to understand complex dual emotions and natural transitions from one emotion, such as fear, to another, such as anger or hostility, and to connect or disconnect from an emotion depending on its usefulness to the situation. EI may naturally occur in some individuals, but it can also be improved through targeted learning.

Gender differences also moderate the stress response. Research suggests that the natural response to stressors differs for women from the traditional fight-or-flight pattern. In addition to the fight-or-flight response, women "tend and befriend" in times of high stress. The moderating effects of tend and befriend may explain differences in psychophysiological consequences of stress for men and women. There are also issues related to the conflict between work and family roles that appear to affect women to a larger extent than men.

## 5. MEASURES OF JOB STRESS

Valid, reliable job stress measurement requires multiple psychological, environmental, and medical measurements. There are four construct categories for

stress instruments: measures of (i) environmental demands and sources of stress; (ii) healthy, normal stress response; (iii) modifiers of the stress response; and (iv) psychological, behavioral, and medical strain. For the purposes of workplace stress measurement, the focus is on four measures that offer a range of construct assessment. For a more clinical assessment of organizational stress, Levinson's classic organizational diagnostic approach may be used.

### 5.1. Job Stress Survey

Spielberger's Job Stress Survey (JSS) measures job pressure and lack of organizational support as the two primary sources of job stress. The JSS measures the severity and the frequency of 30 different job demands. Examples from the JSS scale items are "excessive paperwork" and "poorly motivated coworkers." Severity and frequency scores are also computed for 10-item job pressure and organizational support subscales that were factor analytically derived from the 30 JSS items. The subscales show adequate internal consistency and reliability.

### 5.2. NIOSH Generic Job Stress Questionnaire

The Generic Job Stress Questionnaire was developed by NIOSH and contains measures of 13 different job stressors as well as a host of measures of individual distress and modifiers of the stress response. The questionnaire assesses constructs within domains contained in the NIOSH job stress model. Specific stressor, strain, and modifier variable constructs were selected for inclusion in the instrument on the basis of a content analysis of the job stress literature, and the scales selected to measure these constructs were adapted from scales with known reliability and validity. The instrument was designed to be modular in form so that the diagnostician or stress researcher can select individual scales, or the entire instrument can be used.

### 5.3. Occupational Stress Inventory

The Occupational Stress Inventory (OSI) measures three domains of adjustment: occupational stress, personal strain, and coping resources. Six occupational stress scales measure role overload, role insufficiency, role ambiguity, role boundary, responsibility, and the physical environment. Four personal strain scales

measure vocational strain, psychological and/or emotional problems, interpersonal problems, and physical illnesses. Four coping scales measure recreational activities, self-care, social support, and rational/cognitive coping. Plotting standardized scores on each of the 14 scales produces a “stress profile” for an individual manager or executive, or the scales can be used to develop organizational profiles.

### 5.4. Pressure Management Indicator

The Pressure Management Indicator (PMI) is a measure developed from the earlier OSI and based on the Cooper and Marshall’s stress model. The PMI is used increasingly as a diagnostic instrument in Europe. It is more reliable, more comprehensive, and shorter than the OSI. The PMI measures sources of pressure, such as workload, relationships, work–home balance, and daily hassles; individual differences, such as type A behavior/drive, locus of control, and social support; and job stress effects, such as job satisfaction, security, mental health, and physical symptoms.

## 6. EARLY WARNING SIGNS OF STRAIN

From the prevention perspective, it is important to recognize the earliest warning signs of strain in order to intervene early in the process. Worker strain can result following prolonged exposure to stressors when paired with poor response patterns. Pains of unknown origin, fatigue, inability to concentrate, and irritability are early warning signs for individuals. In addition, symptoms of depression, increased anger and hostility, increased accidents, or signs of substance abuse can be clues that interventions are warranted. These early warning signs may be accompanied by increased aggressive behavior, disrespect, and withdrawal from relationships.

For organizations, early warning signs that normal work stressors are leading to strain include general patterns such as slight changes in productivity with decreased quantity and/or quality of work, increased absenteeism, decreased commitment to the organization, poor interpersonal work relationships, increased tardiness, and more conflict among workers. Other early signals of distress include general dissatisfaction, low motivation, and low morale. Increased accident rates and machine breakdowns may also constitute early warn-

ings. There may be a general loss of vitality within the organization and an atmosphere of distrust and animosity.

As with any preventive strategy, early detection of strain can help to reduce the long-term negative consequences. Key to early detection and sound prevention is vigilance on the part of the management team.

## 7. PREVENTIVE STRESS MANAGEMENT

Prevention is the best public health strategy for any disease epidemic. Because job stress is a health epidemic, prevention holds the best hope for addressing this epidemic. The theory of preventive stress management translates the public health notions of prevention into an organizational context and overlays them on a stress process model as shown in Fig. 1. Stress is one of several chronic organizational health problems—with others being workplace violence, sexual harassment, and suicide—for which prevention is appropriate.

Preventive stress management is an organizational philosophy and set of principles that employs specific

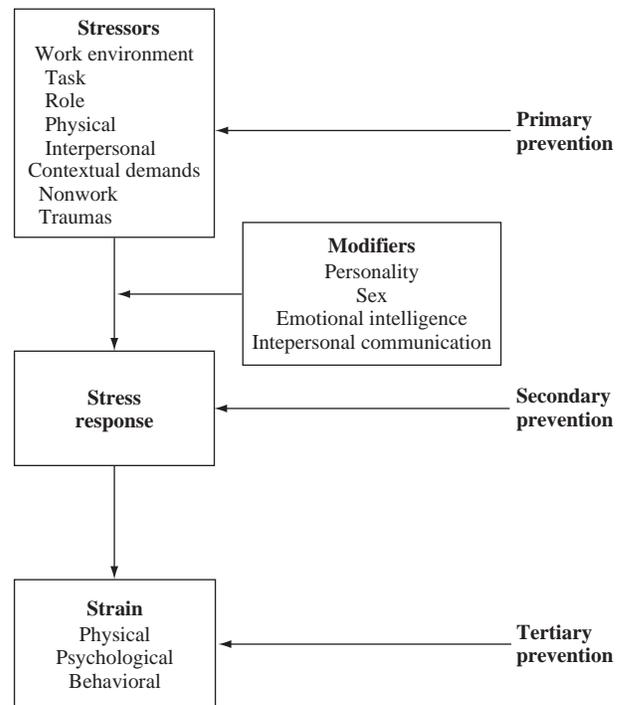


FIGURE 1 Preventive management model for job stress.

methods for promoting individual and organizational health while preventing individual and organizational distress. This philosophy is based on the following five guiding principles that motivate the practice of preventive stress management and provide a framework for healthy organizations and healthy leaders:

Principle 1: Individual and organizational health are interdependent.

Principle 2: Leaders have a responsibility for individual and organizational health.

Principle 3: Individual and organizational distress are not inevitable.

Principle 4: Each individual and organization reacts uniquely to stress.

Principle 5: Organizations are ever-changing, dynamic entities.

Chronic diseases do not arise suddenly; instead, they develop gradually through a progression of disease stages, a “natural life history.” This is true for chronic individual disorders, such as heart disease, as well as for chronic organizational problems, such as workplace violence. The natural history of most diseases is one of evolution from a stage of susceptibility to a stage of early disease and finally to a stage of advanced or disabling disease. At the stage of susceptibility, the individual is healthy but is exposed to certain risk factors or disease precursors. For example, individuals who choose a sedentary life or who choose to smoke cigarettes are at the stage of susceptibility for coronary artery disease as well as several other diseases. When these and other risk factors lead to the development of arteriosclerosis or hardening of the arteries to the heart, the individual is at the stage of early disease or preclinical disease. In other words, the person’s body has responded to the disease precursors, but there are few, if any, symptoms. As the disease advances, it becomes symptomatic or clinical disease. Angina pectoris (heart pains) and heart attacks are advanced manifestations of coronary artery disease.

Preventive stress management is rooted in the public health notions of prevention, which were first used in preventive medicine. The term public health encompasses a broad array of health protection activities inspired by the practice of viewing illnesses within a social context. The dominant diagnostic model in public health involves the interaction between a host (the individual), an agent (health-damaging organism or substance), and the environment. One of the fundamental concepts of preventive medicine is that there is an

opportunity for preventive intervention at each stage in the life history of a disease, as noted previously and depicted in Fig. 1. These interventions are aimed at slowing, stopping, or reversing the progression of disease.

As Fig. 1 shows, there are three stages of prevention strategies: primary prevention, secondary prevention, and tertiary prevention. Primary prevention aims to modify and manage the job stressors and other demands in the work environment. Secondary prevention aims to modify and manage the individual’s response to these job stressors and other demands. Tertiary prevention aims to help and provide aid to those who are experiencing behavioral, psychological, or medical strain symptoms. From a public health perspective, primary prevention is always the preferred point of intervention. For job stress and workplace health, this implies that job redesign efforts and other interventions that alter, modify, or eliminate stressful work conditions are the preferred category of preventive stress management interventions.

Primary prevention is the protection of health directed at the stage of susceptibility and aims to eliminate or reduce the impact of risk factors; it is intervention before the onset of problems or disorders. Primary prevention may be either organizational or individual. Organizational strategies include the following:

- Job redesign
- Participative management
- Flexible work schedules
- Career development
- Design of physical settings

Individual strategies include the following:

- Learned optimism
- Planning and time management
- Social supports and secure attachments
- Maintaining a work–home balance
- Changing type A behavior pattern

It is important to recognize that some primary prevention techniques directly alter stressors. Job redesign, for example, may incorporate more control into a job, or more decision latitude, thereby reducing the stress from lack of control. On the other hand, some primary prevention methods alter the individual’s perception of the stressor. Social support, for example, can help individuals reframe stressors as challenges because they believe they possess the resources to manage a stressor. In both cases, pri-

mary prevention is the preferred first defense against stressors.

Secondary prevention aims at the early detection of problems or disorders and prompt, early interventions to correct departures from health. Secondary prevention may also be either organizational or individual. Organizational strategies aim at improving relationships in the work environment. Organizational strategies include the following:

- Goal setting
- Role analysis
- Team building
- Social supports and secure attachments
- Diversity training and programs

Individual strategies include the following:

- Relaxation training
- Prayer, spirituality, and faith
- Emotional outlets
- Physical fitness and exercise
- Nutrition

Secondary prevention efforts focus on the development of strengths for both individuals and organizations. Team-building efforts, for example, can lead members to understand each other better, manage conflict more effectively, and be more productive. Individual strategies, such as relaxation training, meditation, and exercise, make the body and mind more resilient and more capable of dealing with stress.

Tertiary prevention is prevention of more serious problems, disorders, or death. It is directed at expediting or improving treatment for symptomatic or advanced disease and aims to alleviate discomfort and restore effective functioning. Tertiary prevention is primarily focused at the individual level rather than the organizational level, as is done in treating mental health referrals. Tertiary prevention also includes psychological counseling and medical or surgical treatment. Individual strategies include the following:

- Individual psychotherapy
- Behavior therapy
- Medical care and treatment
- Traumatic event debriefings
- Physical therapy

Tertiary prevention is thus focused on procuring care from qualified professionals who can help

individuals heal. The goal of the preventive stress management approach is to utilize sufficient primary and secondary prevention such that individuals seldom, if ever, need to turn to tertiary prevention for relief.

## **8. IMPLICATIONS FOR EXECUTIVES AND LEADERS**

Job and work stress, as a component of the contemporary organization, do not show signs of decreasing in frequency, intensity, complexity, or outcomes. Effective management of organization and worker stress by executives and leaders begins with the effective management of their own stress. Learning preventive stress management can benefit the individual leader and positively affect stress management throughout the organization. As the workforce continues to change with increased globalization, diversity, and age variation, effective leaders must learn better ways to deal with the inevitable stressors of the workplace. Future leaders may very well set themselves apart by developing the stress prevention skills needed to move the response to stress in a positive direction so that individuals and organizations can reap the benefits of eustress and avoid the pitfalls of strain.

The key tasks for managers and executives are clear. Within the preventive stress management framework, the goal is to increase eustress—the healthy, positive consequence of stress—and to prevent or resolve distress. The three-tiered model presented in Fig. 1 can be used as a road map for these prevention efforts. Primary prevention represents the most efficacious set of techniques because it consists of modifying the stressor or the individual's response to the stressor. Secondary prevention efforts, focused on increasing individual and organizational resilience, should supplement primary prevention methods. Finally, tertiary prevention methods are available for targeting symptoms and healing the wounded within organizations.

Increasing eustress within organizations presents a challenge to managers and to researchers, who are beginning to identify the aspects of work that many employees find pleasurable and motivating. The feeling of eustress has been described as being in the “flow,” similar to a “runner's high” in which time suspends and the person is fully absorbed in the joy of work. It is a state of active engagement, zest, and achievement. Eustress is characterized by feelings of hope,

commitment, and meaningfulness of work. Managers and executives must explore ways to design jobs that tap this positive experience for employees. Managers should also find opportunities to improve policies, procedures, and the physical work environment in an effort to enhance employees' exposure to work that they find engaging and to eliminate potential barriers to eustress.

Executives are vulnerable to the same medical, psychological, and behavioral consequences that affect any workers. They are at increased risk for isolation and loneliness, anxiety, burnout, workaholism, and depression. Their positions in organizations place them as role models whom others look to for guidance regarding effective stress management. Because of their positions, they can model effective prevention efforts that others can follow, thus benefiting the entire organization. By taking a holistic view of health and developing its multiple dimensions (physical, psychological, spiritual, and ethical), executives can lead the way in tapping into the benefits of eustress and possessing the strength and security to effectively manage distress.

Because job stress leads to a systemic response both within the organization and within the individual, executives and workers alike can benefit from strengthening themselves through physical fitness, psychological well-being, spiritual development, and strong character. Physical fitness includes cardiovascular well-being, muscular flexibility, and appropriate muscular strength. Psychological well-being centers on the ability to deal constructively with reality, the capacity to adapt to change, relative freedom from tension or anxiety-related symptoms, and the capacity to love. Spiritual development should not be confused with religious affiliation and is characterized by hope, a sense of purpose, and healthy relationships. Finally, strong character is achieved through ethical behavior and positive personal integrity.

In addition, executives can be proactive in creating healthy work environments in which job stress is optimized, rather than either maximized or minimized. High-performance, healthy work environments are ones characterized by a sense of challenge coupled with a sense of psychological safety and security. Healthy job stress leads to a work environment in which people may produce, serve, grow, and be valued.

Future areas of study in work and job stress include exploration of the individual differences in the stress response and how to incorporate the positive traits into effective stress prevention. Training in skills shown to increase the positive outcomes of stress and reduce strain needs to be incorporated into educational

programs at an early stage and throughout an individual's work life. Finally, shifting the perception of job stress to a neutral construct is both necessary and inevitable. The human response to stress, not the stress itself, determines whether negative or positive outcomes prevail.

### See Also the Following Articles

Environmental Stress ■ Homosexuality ■ Job Stress and Burnout ■ Stress

### Further Reading

- Bar-On, R., & Parker, J. D. A. (2000). *Handbook of emotional intelligence: Theory, research, assessment and application at home, school and in the workplace*. New York: Jossey-Bass.
- Cooper, C. L. (1996). *Handbook of stress, medicine and health*. Boca Raton, FL: CRC Press.
- Hurrell, J. J., Jr., Nelson, D. L., & Simmons, B. L. (1998). Measuring job stressors and strains: Where we have been, where we are, and where we need to go. *Journal of Occupational Health Psychology, 3*, 368–389.
- Levinson, H. (2002). *Organizational assessment: A manual*. Washington, DC: American Psychological Association.
- Macik-Frey, M., Quick, J. C., & Quick, J. D. (In press). Interpersonal communication: The key to unlocking social support in preventive stress management. In C. L. Cooper (Ed.), *Handbook of stress, medicine and health* (Rev. ed.). Boca Raton, FL: CRC Press.
- Perrewé, P. L., Ferris, G. R., Frink, D. D., & Anthony, W. P. (2000). Political skill: An antidote for workplace stressors. *Academy of Management Executive, 14*(3), 115–123.
- Quick, J. C., Cooper, C. L., Nelson, D. L., Quick, J. D., & Gavin, J. H. (2003). Stress, health, and well-being at work. In J. Greenberg (Ed.), *Organizational behavior: The state of the science* (pp. 53–89). Mahwah, NJ: Laurence Erlbaum Associates.
- Quick, J. C., Cooper, C. L., Quick, J. D., & Gavin, J. H. (2002). *The Financial Times guide to executive health: Building your strengths, managing your risks*. London: Financial Times/Prentice Hall.
- Quick, J. C., Quick, J. D., Nelson, D. L., & Hurrell, J. J., Jr. (1997). *Preventive stress management in organizations*. Washington, DC: American Psychological Association. (Original work published 1984 by J. C. Quick & J. D. Quick).
- Salovey, P., Rothman, A. J., Detweiler, J. B., & Steward, W. T. (2000). Emotional states and physical health. *American Psychologist, 55*, 110–121.
- Taylor, S., Klein, L., Lewis, B., Gurung, R., Grunewald, T., & Updegraff, J. (2000). Biobehavioral responses to stress in females: Tend and befriend, not fight or flight. *Psychological Review, 107*, 411–429.



# Job Stress and Burnout

Philip Dewe

*Birkbeck University of London, London, United Kingdom*

1. Introduction
  2. Defining Job Stress
  3. Models of Job Stress
  4. Causes of Job Stress
  5. Coping with Job Stress
  6. Burnout
  7. Other Job Stress Responses
  8. Job Stress Interventions
- Further Reading

(strain). Contemporary definitions of stress view it as relational, resulting from the transaction between the individual and the environment. This approach has provided an opportunity to consider those processes that link the individual to the environment. The shift in emphasis away from the different components of the stress process to the process itself has raised a number of issues in terms of how measurement practices should develop and how researchers fulfill their obligations to those whose working lives they study.

## GLOSSARY

**appraisal** The processes an individual goes through in determining the significance and meaning of a stressor (primary appraisal) and what resources are available to cope (secondary appraisal).

**strain** Responses that indicate that the individual is under stress.

**stress management programs** Interventions at the level of the individual and organization aimed at reducing stress and allowing the individual to fully contribute to the goals of the organization.

**stressors** Situations or properties of a situation that have the potential to cause stress.

Job stress has traditionally been defined in terms of a stimulus, response, or an interaction between the two. These approaches to defining job stress have allowed researchers to identify an array of different causes of job stress (stressors) and relate them to different responses

## 1. INTRODUCTION

The growing awareness of the impact of job stress is best illustrated by the sheer increase in the volume of job stress research in recent years. Job stress is a condition familiar to many. In some form or another, it influences all working lives, leaving those who experience job stress emotionally drained and frequently more vulnerable to other illnesses and disease. It is not just individuals who experience the impact of job stress. Organizations, families, communities, and society all bear the costs of job stress to such an extent that job stress is frequently viewed as an inevitable feature of 21st-century living. However, when sifting through all the information on job stress it is important to maintain a sense of perspective. It is necessary to guard against feeling that the problem of job stress has been overexaggerated, that it is not worth worrying about, or that it is someone else's responsibility. The

issue is not whether there is too much or too little stress in our working lives but what actually causes job stress, what can be done to cope with it, and what needs to be done to provide a working environment within which individuals can maximize their potential and enhance their well-being. This is the responsibility that job stress researchers have to those whose working lives they study.

## 2. DEFINING JOB STRESS

The difficulty when talking about job stress is that it means different things to different people, with the result that the term stress has taken on a variety of meanings. Traditionally, stress research has been influenced and guided by definitions that focus on stress as a stimulus, a response, or on the interaction between the two. These three approaches to defining stress frequently reflect the discipline or orientation of researchers. Stress as a stimulus is defined in terms of those events or properties of events that place demands on an individual over and above what is normal. Response definitions of stress are concerned with identifying particular responses or patterns of responses that indicate that the individual is confronted with a demanding event. Stimulus and response definitions of stress have provided researchers with an opportunity to identify both a range of events that have demanding properties and responses that indicate that the individual is under stress. However, such definitions have been criticized because each reflects just one aspect of the stress process; defining stress in this way fails to provide any understanding of the nature of the stress process. They have also been criticized because they fail to take into account individual differences and therefore cannot account for the fact that what may be stressful for one individual may not be stressful for another.

Interactional definitions of stress tried to overcome these difficulties by describing stress in terms of the interaction between the stimulus and the response. This definition is one in which a relationship, usually correlational, is hypothesized between the stimulus and the response. Cooper *et al.* argue that defining stress like this, where the focus is simply on the interaction between the stimulus and the response, fails to provide any understanding of the complexity of the interaction or of the nature of the processes that may be involved. As a consequence of the difficulties associated with stimulus, response, and interactional

definitions of stress, more contemporary definitions adopt a definitional approach described by Lazarus in which stress is viewed as relational in nature involving some sort of a transaction between the individual and the environment. Transactional definitions of stress imply that stress is neither solely in the individual nor in the environment but in the transaction between the two. It is the transaction that links the individual to the environment. Furthermore, it is by thinking of stress in transactional terms that draws attention to what the nature of the transaction may be. In this way, defining stress in transactional terms achieves what earlier definitions failed to do: It focuses attention on those processes that link the individual and the environment and that are at the heart of the transaction.

Lazarus describes these processes as primary and secondary appraisal. Primary appraisal is when the individual evaluates the importance of the event in terms of harm, threat, or challenge to well-being. In short, it is the realization by the individual that something important is at stake. Secondary appraisal is when the individual considers what can be done. It involves an evaluation of those resources that are available to cope with the stressful event. Identifying the processes that are at the heart of a stressful transaction in terms of how an event is appraised and how it can be coped with means that contemporary definitions of stress are more concerned with capturing its transactional qualities than focusing, as was once the tradition, on individual components to that transaction. As a result, stress is best defined in terms of those encounters in which the demands from the encounters are appraised as taxing the individual's resources, thus threatening well-being.

## 3. MODELS OF JOB STRESS

Models of job stress have, quite naturally, closely followed how stress has been defined. Early models of job stress followed an interactional framework. That is, they were interested in the interaction between the cause of job stress (stimulus) and its impact (response). As outlined, the interactional framework suggests a relationship between the perceived presence of demanding work conditions (stressors) and different stress responses (strain). The relationship, usually correlational, postulated that the more demanding the stressor, the greater the probability it would result in strain. As job stress researchers began to explore different relationships between stressors

and strains, they soon realized that the stressor–strain relationship may be influenced by other variables, such as age, gender, organizational level, and individual differences. Therefore, a second stage in the development of job stress models began. This stage is described as the moderator stage and was concerned with identifying organizational and individual variables that moderate (influence) the stressor–strain relationship.

The interactional and moderator models, although important in providing information on the causes and consequences of job stress and on variables that influenced the relationship between the two, were limited in their ability to explain the stress process. So began in the development of job stress models the transactional stage. Transactional models, like transactional definitions of job stress, were concerned with the sequence of events and the processes that linked the stressor to the strain. McGrath had for some time urged researchers to approach the investigation of job stress by developing models of job stress that mirrored the sequencing of events from stressor to strain. To capture this sequence, many models of job stress adopted the concept of fit. The idea of fit was coupled with the idea of equilibrium or balance. In general, where there was a misfit or imbalance between the person and the environment, such as in the case of a person failing to cope with the demands of a job, a state of disequilibrium would exist and this state would be associated with strain. Implicit in the concept of misfit is the notion of an individual's ability to manage or deal with a demanding job event.

As job stress researchers began to explore the nature of this misfit, models of job stress began to outline more explicitly what misfit may involve. Job stress models began to build into their frameworks different transactional qualities, such as how stressors are appraised, how individuals cope, and organizational and individual resources available to individuals to manage the demanding encounter. What distinguishes the transactional model from the earlier interactional and moderator models is its focus on process. It requires researchers to identify the processes that link the individual to the environment and to consider the sequencing of those processes and their role over time. However, although job stress researchers have long accepted at the theoretical level the importance of identifying process elements in job stress models, empirical work is predominately still being carried out using a interactional–moderator perspective. Nevertheless, job stress research has contributed much to our

understanding of its causes and consequences, not at times without controversy and intense debate.

#### 4. CAUSES OF JOB STRESS

The causes of job stress are many and varied. Perhaps the easiest way to develop a comprehensive picture of what causes job stress is to start with the early work of Kahn and colleagues and two causes of work stress: role conflict and role ambiguity. These two work role demands along with role overload have become the most frequently researched causes of job stress. In general, role conflict refers to job situations in which individuals are confronted with conflicting demands while performing their organizational role so that by complying with one of the demands it is very difficult to comply with the other(s). Role conflict frequently arose from having to work under policies that conflicted with one another, being responsible for tasks over which there was very little control, and having to do things that were likely to be accepted by one person but not accepted by another. Perceived role conflict led many to experience dissatisfaction and strain. Similar findings have been found for role ambiguity. In the case of role ambiguity, the significant issue was a lack of information. Individuals experiencing role ambiguity more often than not found that they lacked the relevant information to perform their work roles, determine how their performance was being judged, and consider whether the decisions they made were the right ones. Role ambiguity was consistently associated with levels of psychological strain.

A third form of difficulty associated with performing one's role in the organization was role overload. Role overload was first measured in terms of the demands that arose from the volume of work that had to be done. Frequently associated with time pressures, this type of role overload became known as quantitative overload and was measured in terms of the amount of work to be done in the available time, the idea that all tasks have a sense of urgency, not being able to do the work because of constant interruptions, and the belief that no matter how much work was done there was still as much to be done at the end of the day as there was at the beginning. As organizational roles became more complex, role overload took on a second dimension. This type of role overload was described in terms of qualitative role overload and referred to demands on individuals required to ensure that their skills matched the increasing requirements of the job. This form of

role overload was concerned with the need to keep up with the technical requirements of the job, having to spend increasingly more time learning new policies and procedures, keeping up with constant job changes, and sensing that at times there was an imbalance between current skills and required skills. Both these forms of role overload had negative consequences in terms of well-being and satisfaction on those who perceived them to be present in their jobs.

The importance of discussing these types of role demands in such detail is that even though they were first identified in the 1960s, they are—despite the significant social, economic, political, and organizational changes that have occurred since that time—still by far the most frequently measured causes of job stress. Other causes of job stress have been identified, and it is worth exploring these as well. In addition to organizational roles, in 1976 Cooper and Marshall identified a range of different causes of job stress, including extrinsic aspects of the job, such as working conditions and job design; structure and climate issues, such as levels of participation, supports, controls, and rules and regulations; career development issues, such as promotion, job security, and goal achievement; and relationship issues involving, for example, subordinates, colleagues, and the boss, and difficulties in delegating and dealing with sensitive interpersonal issues.

Cooper and Marshall also included “extraorganizational sources of stress” in their list of causes of job stress. These sources relate to the home–work interface and include balancing family and work responsibilities; family problems; financial and other life difficulties; and the interference of work with nonwork activities, such as leisure, hobbies, community life, and social activities. Other sources of job stress include the level of local and international travel, working abroad, life-threatening events and safety at work, redundancy, bullying, harassment, social responsibility, and ethics. Beehr also provided a similar list of the causes of job stress.

Although a wide range of events, usually defined as stressors, can be identified as having the potential for causing job stress, there has been considerable debate about how best to measure such stressors. The common approach has been to use self-report measures in which individuals generally indicate on a 5-point scale whether they agree or disagree as to the presence of different stressors in their job. The major criticism of this approach is that simply agreeing or disagreeing that a stressor is present in a job does not actually

indicate how demanding that stressor may be or its significance in terms of well-being or complexity. As a consequence, a number of refinements to self-report measures have been suggested. First, there has been discussion as to whether stressors should be measured globally, allowing comparisons to be made across different occupational groups, or tailored to meet the requirements of specific occupations so that the unique stressors for that occupation can be identified. This discussion, particularly in relation to global measures, often spills over into what events should be measured. That is, have job stress researchers, by focusing too much on organizational role stressors, overemphasized the presence of these work role events and ignored the significance of other events, particularly those that reflect the enormous change that has occurred in society and organizations?

Second, when measuring stressors should more attention be paid to the characteristics of the stressors? This would include measuring the frequency, duration, and intensity of the event; how demanding it is; and the way in which different events may result in the accumulation of stress. Then, thinking of stress in transactional terms, there is a need to consider the difference between the presence of an event and its significance and meaning. Here, the emphasis shifts to determining how an event is interpreted or appraised. This means developing measures that capture the meanings individuals give to the event. In this way, the focus would be on whether the event is appraised in terms of, for example, harm/loss and threat or challenge. Focusing on the meaning and salience of an event may be a more significant predictor of job stress than the more traditional measurement approach that simply captures whether the event is present in the job. Of course, focusing on the meanings that individuals give to events raises the issue of subjective versus objective measurement and the fact that objective measures are also an important component contributing to our understanding of the causes of job stress.

Debating how to refine self-report measures of stressors is one way of capturing the complexity and character of different stressors. However, as Cooper *et al.* suggest, there are other more qualitative ways of measurement that should not be overlooked, including the use of critical incident techniques, daily diaries, and structured interviews. These measurement strategies allow stressors to emerge from the narrative allowing individuals to identify the significant events in their working lives. Whatever measurement method is selected, it is clear that identifying stressors is a crucial

part of furthering our understanding of the stress process and in considering the development of different interventions strategies.

## 5. COPING WITH JOB STRESS

When stress is defined in transactional terms, coping is fundamental to our understanding of the stress process. Coping is part of what Lazarus describes as secondary appraisal. Secondary appraisal is concerned with “what can be done” to deal with the demanding situation and involves identifying and evaluating the available coping resources. Secondary appraisal therefore involves what a person actually thinks and does; thus, as Dewe *et al.* suggest, coping can be defined as “the cognitions and behaviors, adopted by the individual following the recognition of a stressful encounter, that are in some way designed to deal with that encounter or its consequences.” From their review of the coping literature, Dewe *et al.* suggest that it is possible to identify three themes associated with defining coping in this way: Coping is (i) relational in the sense that it reflects the relationship between the individual and the environment, (ii) process focused, and (iii) integrative in that it is part of the appraisal process that links the different dimensions to the stress process.

The most common approach to researching coping in relation to job stress is what Cox describes as taxonomic, where researchers describe and categorize different coping actions that appear broadly applicable to work situations. By far the most popular schema for classifying coping strategies distinguishes between what Lazarus and Folkman describe as problem-focused and emotion-focused strategies. Problem-focused strategies are primarily concerned with problem solving or altering the source of the problem, whereas emotion-focused strategies are aimed at reducing the emotional discomfort. This is not the only schema for classifying coping strategies. In general, however, most classification attempts describe variations on the problem- and emotion-focused themes. Nevertheless, O'Brien and DeLongis suggest that any classification of coping strategies should include what they describe as relationship-focused coping. They argue that it is important to recognize that coping requires, in addition to problem solving and emotional management, relationship maintenance, and that coping classifications should be expanded to include this type of coping. From their more detailed examination of coping functions, Ferguson and Cox suggest that

there is a degree of consensus among researchers that there may be four qualitatively distinct coping types: emotional regulation (dealing with the emotional consequences of the problem), approach behaviors (aimed at directly confronting the problem), reappraisal (strategies for readjusting and reinterpreting the meaning of the event), and avoidance behaviors (ignoring the existence of the event).

Classifying coping strategies into different categories such as problem- and emotion-focused should not be regarded as indicating that one form of coping is better than another. Placing different strategies into one or another category is not the same as understanding how they work or the manner in which they are used. Researchers have come to realize that, when considered in the context of a particular stressful encounter, what function a coping strategy may be performing is not as easy as to define as independently describing the focus of separately derived categories. Dewe noted that researchers have commented on the fact that the same coping strategy may play a number of different functions depending on the particular encounter. Similarly, the sequencing, patterning, and combined use of different coping strategies all contribute to the difficulties of determining whether a coping strategy is problem- or emotion-focused.

Despite these difficulties, coping research has expanded to explore the influence of personality and gender on the use of different coping strategies, the categories of coping strategies associated with different work stressors, the relationship between different categories of coping and different outcomes, the role of different coping types over the longer term, the nature of the organizational culture and climate and its impact on the use of different coping strategies, and the role of coping as a mediator or moderator in the stress process. Many studies have, somewhat implicitly, drawn attention to the question of coping effectiveness. Regarding the issue of how to determine coping effectiveness, the following question needs to be answered: Effectiveness for whom and at what cost? This has led researchers, when faced with the question of how to establish coping effectiveness, to identify a range of issues that need to be considered, including the need to distinguish between coping and ordinary everyday adapting; short-term versus long-term effectiveness; the difference between effort and effect; the individual's competency to use different coping strategies; the goals an individual may be trying to achieve; the nature of the encounter, particularly in terms of the levels of control an individual may have; and the goodness of fit between how an event is appraised and the coping strategy used.

In addition to the work on identifying specific coping strategies that individuals may use, classifying them, and evaluating their use, research has also focused on “capacity techniques” for coping with job stress. These strategies are not so much designed to deal specifically with a stressful encounter as to develop a person’s capacity to deal with a stressful encounter should it occur. This is often done in a number of ways—exercise, relaxation, meditation, biofeedback, and a philosophy of life—but the end result is to provide an inner sense of energy and well-being and a greater sensitivity to the signals our bodies may send us and thus a greater capacity for dealing with and building resistance to stressful encounters. Moreover, these techniques are frequently discussed in terms of lifestyle changes, how they can play a supportive role in conjunction with more specific coping strategies and so become part of a repertoire of coping activities, how they depend to a considerable degree on individual commitment and preference, and how each technique should be reviewed occasionally so that the intended results continue to be positive.

Many of the issues discussed previously, in terms of both classifying coping strategies and determining their effectiveness, have led researchers to consider how best coping strategies should be measured. Again, researchers have considered changes to self-report coping instruments, including considering the type of instructions given so as to differentiate between what individuals may generally use and what they specifically use in a particular situation, the way in which coping strategies are worded so as to avoid ambiguity in meaning, and whether scoring coping strategies in terms of the frequency with which they are used best captures the nature of the coping process. The general consensus from the debate on coping measurement is that identifying circumstances that perhaps best suit the use of coping checklists will provide researchers with the opportunity to bring other more qualitative techniques to the study of coping so that our understanding can continue to develop.

## 6. BURNOUT

Cooper *et al.* describe burnout as “a special form of strain.” These authors point to the concept of burnout as reflecting a state of psychological strain initially associated with those working in the human service professions. Since the early studies on burnout, researchers have extended their investigations of the

phenomenon to working life in general, considering in detail the consequences of burnout for individuals and their organizations. In the early 1980s, Maslach provided a description of burnout involving three major elements: emotional exhaustion described in terms of not having the emotional energy to sufficiently manage the encounter; depersonalization, in which individuals simply become seen as objects and are treated in a detached way; and a lack of personal accomplishment, in which the tendency is to devalue performance in negative ways. Since this three-dimensional view of burnout was first proposed by Maslach, much discussion has centered on whether emotional exhaustion is the essential feature of burnout, with the roles played by the other two dimensions being disputed.

Cordes *et al.* describe burnout as a developmental process. These authors go on to describe burnout as a gradual eroding process and note that by emphasizing the process of burnout, researchers and organizations are provided with a mechanism for understanding what to look for and the types of interventions that may be necessary. Their work supports a process in which the onset of burnout is marked by emotional exhaustion. Depersonalization follows, as Ashforth and Lee suggest, because it is a “means (albeit futile) of staunching the flow of emotional energy, of coping with growing exhaustion.” The issue of whether, as depersonalization occurs, the individual begins to sense a loss of accomplishment and hence a degrading of achievements is less clear. Cordes *et al.* note that one possible reason for this is that a lack of personal accomplishment may also be explained in terms of a range of constraining organizational factors and hence depersonalization may develop somewhat independently of the other two dimensions.

Correlates of burnout are many and varied, and researchers have explored these at a number of levels, including the individual level (e.g., gender, age, commitment, and individual differences), the job level (e.g., work role demands, client relationships, and autonomy), and the organizational level (e.g., organizational culture, management style, and communications). In general, it is clear that a range of individual, job, and organizational level factors influence the experience of burnout. However, it is also clear that more work is still needed to understand where in the process these different factors have their most significant effect, how far their effect can be generalized, and what this means in terms of the development of intervention strategies.

## 7. OTHER JOB STRESS RESPONSES

The fact that “being under stress” has become such a common phrase has meant that an array of responses (strains) have been measured as indicating that an individual is stressed. The general approach to outlining job-related stress or strain is to divide the different strains into three categories: physiological, psychological, and behavioral. Reviews of physiological indicators used to measure strain would include measures that capture cardiac, biochemical, and gastrointestinal activity. Under the heading of “poor physical health” associated with job stressors, Kahn and Byosiere list increased pulse rate, high blood pressure, high cholesterol levels, ulcers, and cardiovascular heart disease. Physiological indicators may also include, as Jex and Beehr suggest, behaviors that are closely associated with disease, such as smoking and drinking. The use of physiological measures presents researchers with a range of measurement challenges. However, evidence suggests that the demands of the job can produce physiological reactions and so this approach to measurement offers what Cooper *et al.* describe as important causal pathways with considerable explanatory potential.

Not surprisingly, as Jex and Beehr note, psychological strains are by far the most common type of strain measured. To give an idea of the range of psychological strains measured, these authors list strains from recent studies such as depression, alienation, boredom, frustration, anxiety, hostility, and dissatisfaction. All are correlates of job-related stressors. Similarly, Kahn and Byosiere, in their table of “psychological responses to work stress,” identify 43 different strains, including in addition to those noted previously, confusion, irritation, resentment, tension, and emotional arousal. Kahn and Byosiere conclude that the psychological effects of job stressors are “real, painful, and costly.” Moreover, they extend to health behavior, job performance, and the enactment of other life roles.

Jex and Beehr note that behavioral strains are the least studied. They discuss two broad categories of behavioral strain: organizationally relevant behaviors (e.g., job performance, absenteeism, and turnover) and individually relevant behaviors (e.g., smoking, drinking, and substance use). Similarly, Kahn and Byosiere identify degradation/disruption of the work role, aggressive behavior at work, flight from the job, degradation/disruption of other life roles, and self-damaging behaviors. Both sets of authors agree that behavioral

effects both on and off the job are extremely costly to both the individual and the organization and warrant more research. Such behaviors are the result of a complex process, and researchers need to try to develop an understanding of the process and the role that job stressors play in it.

Although different emotional responses are included among the list of psychological strains, there is growing acceptance that more attention should be paid to workplace emotions. If, as Lazarus suggests, stress is essentially about emotional reactions, then exploring the specific role of emotions in the stress process may help to provide a richer understanding of what happens to a person in a stressful encounter. Lazarus explains that it is the way an event is appraised that gives that event its emotional quality. Taking this one step further, Lazarus suggests that different emotions are associated with particular appraisals, thus providing a causal pathway for investigation. Lazarus identifies 15 different emotions and categorizes them as negative (e.g., anger, shame, envy, and guilt), positive (e.g., pride and relief), and mixed (e.g., hope and gratitude).

## 8. JOB STRESS INTERVENTIONS

The most common approach to describing stress management programs is to divide them into three categories depending on the level at which the intervention takes place. The first category describes primary interventions aimed at reducing job stressors, including, as Newman and Beehr suggest, strategies for changing organizational characteristics (e.g., organizational restructure and redesigning organizational processes), changing role characteristics (e.g., redefinition of roles, increasing participation in decision making, and reducing ambiguity), and changing task characteristics (e.g., better match of individual abilities with job requirements, provide skill training, and job redesign).

The second category describes secondary interventions. The focus here is on providing stress management training. Secondary interventions are probably the most common job stress interventions and involve a range of techniques, strategies, and skills aimed at equipping individuals to better manage the demands of the job. Newman and Beehr identify four categories of intervention: those aimed at psychological conditions (e.g., planning ahead and realistically assessing one's expectations), those aimed at physical or physiological

conditions (e.g., diet, exercise, and sleep), those aimed at changing one's behavior (e.g., learning relaxation strategies, taking time off, and developing supportive networks), and those aimed at changing one's work environment (e.g., changing to a less demanding job).

The third category describes tertiary interventions, and here the provision is in terms of employee assistance programs (EAPs). Defining EAPs is somewhat problematic because of the array of perceptions and the range of programs. The definition developed by Berridge and Cooper reflects this need to be comprehensive. They define EAPs as

A programmatic intervention associated with the work context, usually at the level of the individual employee, using behavioral science knowledge and methods for control of certain work-related problems that adversely affect job performance, with the objective of enabling the individual to return to making her or his full job contribution and to attaining full functioning in personal life.

The components of an EAP should reflect a preferred model of practice that reflects the needs of the employees, the goals of the organization, and the underlying principle that for both humanitarian and economic reasons rehabilitation is better than replacement. EAPs are generally operated in two modes: in-house provision of the program and the counseling service or externally located EAP and counseling service. Referrals can be through constructive discussion, in which typically the supervisor or manager makes the referral, or through self-referral, in which the employee initiates proceedings. Most EAPs are developed around a systematic set of procedures designed to identify and respond to personal issues that interfere with work performance. These procedures include a needs analysis to determine the nature of the problem; the development of a program of counseling, advice, and assistance depending on the nature of the problem; a procedure for referrals; a protocol outlining ethical issues, including informed consent, confidentiality, and anonymity; follow-up procedures; and a program of evaluation.

Although EAPs have become an established part of many organizational stress management procedures, less attention has been given to evaluating their impact. A number of reasons have been given, including methodological difficulties in relation to collecting and analyzing what is confidential data, establishing appropriate criteria against which success can be gauged, and establishing what is best practice with regard to carrying out evaluative procedures. Although both

qualitative and quantitative criteria have been suggested as criteria for evaluation, the reported findings are mixed. Because of the variety of different programs, it may be best to focus on whether the goals of specific programs have been successful rather than attempt to identify universal criteria that cover all programs.

One of the issues regarding stress management programs is whether many of the programs assume that work environments have the same impact on everyone and therefore many of these programs fall short because they offer only partial solutions, place all responsibility for change on the individual, or fail to take into account the wider organizational structure and context within which behavior takes place. This debate has widened to include discussion about the role that theory should play in the design and development of stress management programs. The thrust of these arguments has been that by viewing stress in transactional terms, stress management programs need to acknowledge that stress does not reside solely in the individual or solely in the environment but in the transaction between the two. In this way, the focus for the design of a stress management program would shift to the relational nature of stress and the need to consider both individual and environmental issues rather than simply placing the responsibility on one. The continuing debate reflects a wider concern held by those researching job stress; that is, what needs to be done in terms of measurement practices and theory development so that we can contribute significantly to those whose working lives we research.

### See Also the Following Articles

Environmental Stress ■ Homosexuality ■ Industrial/Organizational Psychology, Overview ■ Job Stress ■ Stress

### Further Reading

- Arthur, A. (2000). Employee assistance programmes: The emperor's new clothes of stress management. *British Journal of Guidance & Counselling*, 28, 550-559.
- Ashforth, B. E., & Lee, R. T. (1997). Burnout as a process: Commentary on Cordes, Dougherty and Blum. *Journal of Organizational Behavior*, 18, 703-708.
- Beehr, T. A. (1995). *Psychological stress in the workplace*. London: Routledge.
- Berridge, J., & Cooper, C. L. (1993). Stress and coping in U.S. organizations: The role of the employment assistance programs. *Work & Stress*, 7, 89-102.

- Cooper, C. L., & Marshall, J. (1976). Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology*, 49, 11–28.
- Cooper, C. L., Dewe, P. J., & O'Driscoll, M. P. (2001). *Organizational stress: A review and critique of theory, research, and applications*. Thousand Oaks, CA: Sage.
- Cordes, C. L., Dougherty, T. W., & Blum, B. E. (1997). Patterns of burnout among managers and professionals: A comparison of models. *Journal of Organizational Behavior*, 18, 685–702.
- Cox, T. (1978). *Stress*. London: The Macmillan Press.
- Dewe, P., Cox, T., & Ferguson, A. (1993). Individual strategies for coping with stress at work: A review. *Work & Stress*, 7, 5–16.
- Dewe, P. J. (2001). Work stress, coping and well being: Implementing strategies to better understand the relationship. *Research in Occupational Stress and Well Being*, 1, 63–96.
- Ferguson, A., & Cox, T. (1997). The functional dimensions of coping scale: Theory, reliability and validity. *British Journal of Health Psychology*, 2, 109–129.
- Jex, S. M., & Beehr, T. A. (1991). Emerging theoretical and methodological issues in the study of work-related stress. *Research in Personnel and Human Resources Management*, 9, 311–365.
- Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress: Studies in role conflict and ambiguity*. New York: John Wiley & Sons Inc.
- Kahn, R. L., & Byosiere, P. (1991). Stress in organizations. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 571–649). Palo Alto, CA: Consulting Psychologists Press.
- Lazarus, R. S. (1999). *Stress and emotion: A new synthesis*. London: Free Association Books.
- McGrath, J. E. (Ed.) (1970). *Social and psychological factors in stress*. New York: Holt, Rinehart and Winston, Inc.
- Maslach, C. (1982). *Burnout: The cost of caring*. Englewood Cliffs, NJ: Prentice-Hall.
- Newman, J. E., & Beehr, T. A. (1979). Personal and organizational strategies for handling job stress: A review of research and opinion. *Personnel Psychology*, 32, 1–43.
- O'Brien, T. B., & DeLongis, A. (1996). The interactional context of problem-emotion and relationship-focused coping: The role of the big five personality factors. *Journal of Personality*, 64, 775–813.



# Jury Decision Making

Stephanie Albertson, Erin J. Farley, and Valerie P. Hans

University of Delaware, Newark, Delaware, USA

1. Introduction
  2. Decision Making of Individual Jurors
  3. The Psychology of Evidence
  4. The Jury as a Group
  5. Comparative Research on Juries
  6. Conclusion
- Further Reading

This article discusses trial by jury, including empirical jury research and applications. The use of the jury allows laypersons, rather than legally trained judges, to decide court cases. Critics question whether juries are able to decide cases competently and fairly. Addressing these criticisms, social scientists have examined the influence of individual factors and group variables on the jury decision-making process. Scholars have also studied the effects of jury reforms aimed at improving jury decision making. After summarizing the empirical findings and applied research, the article closes by discussing new directions in research on jury decision making.

## GLOSSARY

**anonymous jury** A jury for which the jurors' names are not revealed to the defendant or to the media. Used most often in high-profile trials or trials with security risks.

**evidence-driven deliberation** When jurors engage in general discussion of evidence in the case prior to voting, in contrast to verdict-driven deliberation, in which voting typically precedes evidence discussion.

**hung jury** A case in which the jury cannot agree on one or more verdict outcomes.

**jury selection** The process by which individual citizens are chosen for the jury.

**scientific jury selection** The use of scientific techniques by jury consultants to determine the individual demographic and attitudinal characteristics that are linked to views of a case. Techniques include community surveys, mock trials, focus groups, and in-court selection.

**story model** Theory of juror decision making in which jurors arrange evidence in the form of a narrative or story and match the story to the closest verdict alternative.

**voir dire** Questioning process during jury selection.

## 1. INTRODUCTION

Trial by jury, in which citizens from the community hear evidence and decide on a verdict, is an important symbol of democracy. It allows peers, rather than legally trained judges or magistrates, to evaluate cases and reach binding decisions. As a vehicle for injecting lay participation into legal decision making, the jury has broad democratic appeal. Many contemporary legal systems incorporate juries, including those of the United States, Great Britain, Canada, Australia, New Zealand, and more than 40 other nations in Africa, Asia, South America, and the South Pacific. Some newly democratic countries, such as Spain, following the Franco regime, and Russia, after the breakup of the Soviet Union, have recently introduced or reintroduced the jury.

Public participation in the jury system is said to foster numerous important social and political purposes. First, it promotes accurate decision making that reflects community values. Close scrutiny of an individual case allows the jury to consider the specific facts as well as any extenuating circumstances, increasing the likelihood that the verdict will be fair and equitable. It permits citizens to engage in a key governmental function, educating them about the courts and law in the process. Juries also protect individuals against government oppression and provide a check on politically corrupt prosecutors and judges. Juries send signals through their verdicts about the community's standards and expectations. Thus, although juries today decide relatively few cases compared to earlier times, their verdicts remain important.

Although the jury system serves many positive functions, it has been subject to sharp criticism. The jury's competence has been attacked. Are juries able to comprehend the increasingly complex technical and scientific evidence in today's criminal and civil trials? Critics voice doubts concerning the jury's fairness, and express worry that verdicts are tainted with biases and prejudices of various kinds. For example, juries are viewed as overly sympathetic toward injured plaintiffs, yet at the same time as unduly harsh toward corporations and the wealthy. Jurors are described as "bleeding hearts" who acquit guilty defendants. Jurors are seen as too gullible in accepting illegitimate psychological defenses to criminal charges, such as the insanity defense, battered women's syndrome, battered child's syndrome, and post-traumatic stress disorder. In addition, there is also widespread sentiment in the United States that litigation has spiraled out of control, in part because of the promise of excessively generous jury awards to undeserving plaintiffs.

Debates surrounding the jury provide fertile ground for the application of psychological theory and research. Critically examining these debates and criticisms, psychologists have researched a wide range of jury topics, using mock jury simulation experiments, juror interviews, case analysis, and other methodologies. Psychological research on individual cognition, as well as the impact of juror demographic characteristics, attitudes, biases, prejudices, and personalities on verdict preferences and persuasion has uncovered important individual differences. This research has been applied to such areas as pretrial questioning, voir dire, and jury selection. Psychologists have also analyzed jurors' ability to integrate information, their perceptions and comprehension of evidence, and the dynamics of group decision making. Psychologists

have used this research to propose specific jury reforms to trial procedure. In sum, the jury trial is a complex and fascinating institution that has been the site for advances in the field of applied psychology.

## **2. DECISION MAKING OF INDIVIDUAL JURORS**

### **2.1. Individual Differences among Jurors**

Psychological research has found a juror's decision to be the product of a complex set of factors. Not surprisingly, the trial evidence has the greatest impact. The evidence presented at a trial, and its strength and quality, constitute the prime determinants of jury verdicts. That finding has been replicated in both actual and mock jury studies time and again.

Nonetheless, research has also found certain other aspects of the trial that can affect juror judgments. In deciding criminal and civil cases, jurors draw on their life experiences such as crime victimization or physical injury. Jurors' own knowledge is brought to bear as they evaluate evidence in the case. Other factors that potentially impact individual cognition include a juror's social background, character, ideology, attitudes, and values. The nature of the evidence presented at trial, the way that attorneys structure their arguments, and the order in which witnesses testify can also influence evidence assessments.

#### **2.1.1. Impact of Juror Demographic Characteristics**

Although the strength of the evidence presented at a trial is the strongest predictor of the jury's verdict, many legal actors continue to believe that the juror's demographic characteristics are critically important. Oftentimes attorneys will exercise their peremptory challenges on the basis of characteristics such as race, ethnicity, occupation, gender, and religious preference. However, psychological research indicates that utilizing juror demographic characteristics is not a particularly useful or reliable method for predicting a prospective juror's propensity to favor one side over the other. The factors of age, gender, or race generally bear no relationship to verdict preferences, or have only slight predictive power. In fact, most individual differences among jurors typically account for only a small proportion of the variance in verdict outcomes. As a result,

eliminating or selecting prospective jurors based upon individual demographic characteristics may have little impact or might even backfire on attorneys.

Jury research on capital punishment is one of the few areas in which demographic variables are consistently related to punishment preferences. Women and racial and ethnic minorities are more likely than men and whites to oppose capital punishment. Even though strong opponents of the death penalty punishment do not sit as capital jurors, race and gender continue to play a role in capital jury decision making. Juries with a greater number of white men, for example, are more likely to recommend a death sentence than juries composed of more women and minorities. A juror's gender is also a reliable predictor in cases of rape and sexual harassment, with women generally giving more credence to the victim's testimony and evidence.

Demographic variables have low predictive ability in most cases. A prospective juror's knowledge, ideologies, beliefs, and experiences appear to be more significant than demographic characteristics in helping to explain why a juror reaches a particular decision.

### **2.1.2. Role of Juror Attitudes**

Attitudes tend to be more reliable predictors of verdicts than demographic characteristics. Jurors' attitudes are influential in shaping their views and judgments of the parties, the attorneys, and the evidence presented at the trial. In turn, these assessments of evidence are related to the juror's verdict choices.

Psychological research on attitudes related to criminal jury trials has identified clusters of crime control and due process views. Those who endorse crime control beliefs and attitudes tend to have greater concern about crime, more distrust of defense attorneys, greater support for the police and prosecution, and more worry about defendants going free because of legal technicalities and illegitimate defenses. Crime control adherents tend to favor the prosecution in criminal trials. In contrast, those whose views reflect more due process concerns, such as worry about false convictions and support for the insanity defense and the exclusionary rule, are more likely to favor the defense. Research has also found a relationship between jurors' attitudes toward capital punishment and their verdicts about guilt or innocence. All else being equal, jurors opposing the death penalty are relatively more likely to favor acquittal than supporters of capital punishment.

In civil jury trials, found mostly in the United States, attitudes also play a role. Civil jurors frequently consider

the plaintiff's claim with skepticism. They often question the severity of the plaintiff's claimed injury, sometimes blaming the plaintiff for the injury. Two psychological phenomena, the belief in a just world and defensive attribution, are linked to victim blame. In order to believe in a just and stable world, those observing an injury may minimize the injury and derogate the victim. To believe otherwise, that an innocent person is suffering, would challenge the prevailing belief that the world is predictable and fair. Psychologists have also theorized that when a victim is severely injured, observers are motivated to distinguish themselves from the victim and the threatening possibility of their own injury by engaging in defensive attribution. Defensive attribution may result in derogating the victim, blaming the victim for the injury, or minimizing the injury.

As for attitudes toward defendants in civil cases, civil jurors hold corporate defendants to a high standard of responsibility, and judge civil liability against this elevated standard. Respondents in several polls were found to have negative views of corporate management, and said they would be unable to be fair and impartial in cases involving tobacco companies and health maintenance organizations. However, whether general attitudes toward such businesses and organizations actually affect jury verdicts has not been established.

Pretrial publicity is another important factor that can help to shape initial views of trial evidence in a case. Exposure to pretrial publicity, particularly certain kinds of publicity such as criminal record information or emotional material, can influence the jury's decision making. Research has shown that negative pretrial publicity works in multiple ways: creating negative initial judgments of the defendant's character, shaping the processing of evidence, and influencing jury deliberations.

## **2.2. Psychology of Persuasion**

The adversary trial is a two-way effort at persuasive communication. Psychologists as well as communication scholars have tested and applied theories of communication to the jury trial context. Expertise and trustworthiness are two dimensions of witnesses that influence how jurors view their credibility. Expertise is influential in creating attitude change, especially when the position advocated by the communicator is quite different from the recipient's initial attitude. Communicators arguing against their vested interests tend to be more persuasive compared to those arguing for their personal benefit; in addition, communicators with persuasive intent are viewed as less trustworthy.

An expert witness who is paid a sizeable amount of money to testify and does this for a living may be viewed by jurors as having low credibility compared to an expert witness who does not make a career from providing expert testimony.

Other factors found to increase persuasiveness include the communicator's attractiveness, similarity, and message comprehensibility. The more attractive the communicator, the more favorably he or she is viewed. Greater similarity to the observers (jurors) also leads to more positive appraisals. When it takes less effort to comprehend a message, individuals are more persuaded and are able to recall the information and arguments in the message to a greater extent. The degree of comprehension can become especially crucial during expert testimony, such as when scientific evidence is being presented. A juror's difficulty in comprehending complex evidence may reduce the juror's recall and hence the testimony's impact.

## 2.3. Applications

### 2.3.1. *Voir Dire*

During a typical jury selection, the judge and the attorneys may pose questions to prospective jurors during a period called the *voir dire*, to assess whether jurors can be fair and impartial in the upcoming case. The judge has unlimited ability to remove jurors for cause, if the judge determines that a prospective juror's views, attitudes, experiences, or connections to the parties or the case would make it difficult or impossible for the juror to be fair and impartial. In a number of countries, including the United States, attorneys may also remove a limited number of prospective jurors through peremptory challenges, without providing justification. Psychologists have studied the U.S. *voir dire* process and the effectiveness of attorneys' peremptory challenges. In past years, some attorneys systematically used their peremptory challenges in order to remove racial and ethnic minorities from the jury. The Supreme Court outlawed relying on race or gender in peremptory challenges, yet racial patterns in peremptory challenges persist.

From a psychological perspective, the *voir dire* questioning period is far from an ideal vehicle to assess the existence of bias and prejudice. Often the opportunity to question prospective jurors is severely limited. The questions may be close ended, in which only a yes or no response is called for, rather than open ended, allowing citizens to speak more fully about their views. This problem is compounded by the fact that it is usually clear which response is "right" or socially desirable.

Even if the questioning is more expansive, it relies on jurors themselves to admit their own biases. It is difficult for jurors to become aware of their personal biases and admit them to a high status figure such as the judge.

Some scholars have drawn on psychological theory to propose and support recommendations for reforming the questioning of prospective jurors during the *voir dire* process of jury selection. More expansive *voir dire* should assist attorneys and the courts in their quest for impartiality. Incorporating both close-ended and open-ended questions during pre-trial questioning permits jurors to communicate their views and experiences in their own words, as opposed to answering with the less informative "yes" or "no." An individual rather than a group *voir dire* provides the judge and attorneys with more useful material for exercising challenges. These reforms would likely promote juror candor in disclosing pertinent information and discussing personal private experiences.

### 2.3.2. *Jury Consulting and Jury Selection*

Jury consulting is one tool used by lawyers to assist in identifying attributes of jurors that may negatively impact a case. Psychologists and other social scientists, operating as jury or trial consultants, have developed various techniques to assist in jury selection, collectively labeled scientific jury selection. Jury consulting is a fast-growing industry, with clients willing to pay large sums of money for this method of trial preparation assistance. It is largely an American phenomenon, employed only rarely in other countries.

Two prime methods used in scientific jury selection are telephone surveys and mock trials. In telephone surveys, members of the public are asked questions about their background characteristics, beliefs, attitudes, and views of the upcoming case. Analysis shows whether significant relationships exist between individual variables and case preferences. Jury consultants use these data to create profiles of favorable and unfavorable jurors.

Mock trials, ranging from the presentation of short statements of the evidence in front of focus groups to elaborate trial simulations, are also employed. Groups of community members observe the mock trial, discuss the evidence, reach a mock jury decision, and answer questions about the strengths and weaknesses of the case. The mock trial tests lay reactions to the evidence within the case and the arguments anticipated by the prosecution and defense. By analyzing mock jurors' reactions, the jury consultant can assist the attorney in preparing opening statements, trial presentation materials, and closing arguments.

Augmenting these methods are pretrial investigation of the jury pool, in-court jury selection assistance, witness preparation, and the use of “shadow juries” during the trial. A shadow jury consists of mock jurors who are chosen to mimic the actual jurors in the trial. They sit in court while the trial is going on, and periodically report their views of the evidence to the lawyers. In sum, research undertaken by jury consultants can assist attorneys in jury selection and the conduct of the trial, or even whether to go forward with a trial.

**2.3.2.1. Limitations of Scientific Jury Selection** The effectiveness of jury consulting is debatable. It has been argued that there is no convincing empirical evidence of the impact of scientific approaches to jury selection. Predicting jury decisions on the basis of telephone polls and mock juries is an inexact science. Critics point out the many differences between mock juries and real juries, who sit through the entire case, listen to all the evidence presented, hear the arguments by both sides, and have the responsibility for a final verdict. That is imperfectly captured by the empirical techniques that jury consultants employ. Many commentators believe that jury consulting about witness preparation and evidence presentation is likely to have more impact on trial outcomes than suggestions for jury selection.

**2.3.2.2. Ethical Issues** When scientific jury selection links jurors’ demographic characteristics to trial evidence and arguments, ethical issues surface. First, any peremptory challenges based on the most obvious demographic characteristics, gender and race or ethnicity, contravene the U.S. Supreme Court’s *Batson v. Kentucky* (1986) line of decisions. Categorizing individuals into particular groups based upon how they are predicted to behave or think establishes, enforces, and supports stereotypes and biases. A final ethical concern is related to the fact that scientific approaches to jury selection are expensive. In most cases, only wealthy clients can afford the full range of services. As a result, the disparity between the trial resources of the rich and poor is further aggravated.

### 3. THE PSYCHOLOGY OF EVIDENCE

#### 3.1. Psychological Models of Juror Decision Making

The critical role of evidence in jury decision making has led to extensive study and debate. Many critics

have charged that jurors misunderstand or incorrectly apply complex evidence in reaching their verdicts. To evaluate the merits of these criticisms, psychological research has examined the different ways that jurors comprehend and organize evidence.

Researchers have identified three distinct models of juror decision making: a Bayesian approach, an information integration model, and the story model. Each attempts to elucidate the process by which jurors interpret, evaluate, and combine evidence to arrive at a final verdict.

##### 3.1.1. Bayesian Models

Bayesian models entail a formal approach to examining jurors’ fact-finding process. This perspective focuses on how jurors update their beliefs about the likelihood of a certain event occurring based on the presentation of new evidence. As new evidence is presented throughout a criminal trial, jurors combine the pieces of evidence together on a single dimension that represents the probability of a defendant’s guilt or innocence. Bayesian theory has been described as the “mental meter” model because it theorizes that jurors adjust their subjective probability of an event occurring as new evidence is presented. However, Bayesian models have been criticized for being too simplistic and for failing to capture the multidimensionality of thought processes during decision making.

##### 3.1.2. Information Integration

The second popular model used to examine how jurors make decisions is the information integration model. Unlike Bayesian models, which are based on probability theory, the information integration model is an algebraic modeling approach. Information integration theory specifies how discrete information is combined into a final verdict. During trial, information may come from a number of different sources, including pre-existing biases of the jurors, jurors’ observations of the defendant, the content of the testimony, the judge’s instructions, and the attorneys’ arguments. Jurors place a value and weight on each bit of important evidence. The evaluative information is integrated into a single judgment about the defendant’s guilt or innocence.

##### 3.1.3. Story Model

The story model posits that as jurors hear evidence, they actively construct it into a story or narrative. The

coherence of the story is shaped by the consistency, plausibility, and completeness of the evidence. The story may include character evaluations, motives, important events, and states of mind. Jurors actively fill in gaps in the testimony to complete the narrative through the incorporation of prior knowledge and preconceptions. When all these aspects of evidence are combined, they create a coherent story, and from this story jurors determine the appropriate verdict. Psychological research conducted on juror decision making has revealed a relationship between constructing a narrative and selecting a verdict that best fits the narrative account.

Importantly, jurors may ignore or reject information that is contradictory to their developing narrative or story. Jurors' rejection of evidence that is contrary to their expectations or individual beliefs is not only a critical component of decision-making models, but also a key to understanding and improving how jurors perceive evidence. For example, evidence may be presented in a way that is difficult to understand or inconsistent with expectations, resulting in jurors disregarding it and focusing on more understandable evidence. However, if legal actors are aware of this potential, the presentation of evidence can be altered to increase juror understanding and comprehension.

### **3.2. Perceptions of Evidence**

The presentation and comprehensibility of different types of evidence also merit investigation. Jury perceptions of three types of evidence have been examined: general evidence, eyewitness evidence, and expert evidence.

#### **3.2.1. General Evidence**

In terms of general evidence, one of the consistent concerns among critics is jury comprehension. Critics charge that the jury is unable to understand simple and complex trial evidence. However, research studies of jury decision making are largely at odds with this conclusion. The classic early study of jury decision making, conducted by Kalven and Zeisel (1966) in the 1950s, examined judicial opinions about jury verdicts in thousands of trials. The study revealed that the majority of judges agreed with the jury's verdict in the case. When judges disagreed with jury verdicts, they did not do so more frequently in difficult cases, suggesting that case complexity was not the source of the disagreement. A high rate of judicial agreement with jury verdicts has also been found in subsequent studies.

Nonetheless, concerns about jury competence have persisted. Research on jury performance confirms that even in fairly complex cases, jurors are generally able to understand enough of the evidence to reach a competent verdict. Research on comprehension of legal instructions, however, is more negative about the jury's ability.

The timing of evidence or legal instructions has been examined for its impact on juror decisions. Traditionally, the judge gives legal instructions to the jury at the end of the trial, right before deliberations. This was based on the assumption of a recency effect, in which immediate past events are remembered better than more distant events. On the other hand, the primacy effect posits that early information has great impact. Jurors usually form some opinions about the trial before the closing remarks by attorneys. Therefore, the primacy effect suggests that giving jurors instructions at the beginning of the trial would have greater impact. Introductory legal instructions at the start could help jurors assess legally relevant evidence throughout the rest of the trial. In addition, applied research shows that jurors who receive instructions at the beginning of the trial report greater satisfaction with the process and a better ability to evaluate evidence according to the proper legal principles.

#### **3.2.2. Eyewitness Evidence**

Eyewitness identification is another example of evidence that may have a significant impact on jury decision making. Although eyewitness identification is perceived as the most influential evidence in jury decision making, research has revealed it to also be one of the most unreliable forms of evidence. Studies have shown that an eyewitness's confidence is only weakly associated with an accurate identification; however, the greater an eyewitness's confidence, the more likely the jury will perceive the identification to be accurate. In addition, the status of the eyewitness has an effect on jurors' beliefs about witness credibility. For example, jurors are more likely to give weight to police officers' identifications than those of regular citizens, believing that training promotes police officers' accurate identifications. However, research reveals that police officers are no better at eyewitness identification than regular citizens.

Psychological research reveals important flaws in eyewitness testimony and the types of procedures that police use to confirm eyewitness identification. Yet, many of these flaws are not common knowledge among the jurors. Because of the importance of eyewitness testimony, the courts have come to recognize

the value of informing jurors about them so that eyewitness identification is allocated an appropriate weight. Psychologists have helped to reform police identification procedures, and now routinely testify as experts in jury trials, educating juries about the research findings concerning eyewitness evidence.

### **3.2.3. Expert Evidence**

The use of expert testimony is another important factor affecting juror decision making. The use of complex evidence during trials has greatly increased within the last few decades. As a result, there has been an increasing reliance on expert witnesses to explain the evidence. Therefore, the question of how juries handle expert evidence on scientific and technical topics has become of great interest.

It has been claimed that expert witness credentials overly impress jurors, resulting in too much weight being given to their testimony. On the other hand, juror comprehension of expert testimony is also criticized on the theory that if it is too complex jurors will simply dismiss it. Also, critics and legal actors fear that permitting expert testimony will result in opening the door to trials becoming a battle between expert witnesses.

Research has revealed that jurors do have difficulty comprehending complex evidence. However, lack of juror comprehension can often be attributed to inadequate presentation of evidence and testimony, not the inherent inabilities of jurors. Improving instructions and evidence presentation will in turn increase juror comprehension. In addition, research has cast doubt on a number of assumptions that depict jurors as rubber stamps who passively accept expert testimony. According to empirical studies, instead of being unduly influenced by experts, jurors critically and adequately evaluate both the credentials of the experts and the quality of their testimony. Finally, even if some complex evidence is beyond the ability of the average juror, case studies indicate that jurors can usually understand enough of it to be able to arrive at a justifiable verdict.

## **3.3. Applications**

### **3.3.1. Promotion of Juror Competence through Trial Reform**

Understanding how jurors comprehend and utilize different types of evidence has led to a number of proposed reforms of the traditional adversary trial, several of which have been adopted by multiple jurisdictions.

Adhering to the traditional idea of jurors as passive observers has contributed to jurors' difficulty in comprehending evidence and trial procedures. One general reform is first to accept jurors as more active participants in the trial process, acknowledging the fact that jurors are not blank slates and that they begin to assimilate and integrate evidence long before the end of the trial.

To maximize juror comprehension and decision making, reformers suggest permitting jurors to take notes during trial, allowing jurors to ask questions of witnesses through the judge, simplifying the language used in legal instructions, giving instructions at the beginning versus the end of the trial, giving written instructions to each juror, and giving jurors multi-purpose notebooks to better organize large amounts of evidence and information during complex trials.

Many judges and attorneys are apprehensive about trial reform and concerned about changing traditional aspects of the adversary jury trial. Yet, some of these fears, when tested, have proved to be unfounded. For example, the idea that juror note taking distracts jurors, causes them to miss important information, and offers an advantage to jurors who are better note takers is not borne out. Research shows that jurors are not distracted and do not miss important evidence due to note taking. In fact, note taking encourages jurors to stay alert during trial, serves as a very good memory aid, and increases juror satisfaction with the trial process.

## **4. THE JURY AS A GROUP**

### **4.1. Group Decision Making: Theory and Research on Group Performance**

#### **4.1.1. Decision-Making Process**

During deliberation, the door to the jury room is locked. No one other than the jurors is present. Therefore, what transpires behind the closed door of the jury room has been, until fairly recently, something of a mystery. Researchers have acquired information on the deliberation process by interviewing actual jurors following their jury duty and also through mock jury experiments.

Once a foreperson or jury leader has been selected, typically one of two types of discussion follows. In an evidence-driven discussion, jurors first take up a general discussion of the evidence in the case. Jurors attempt to develop a joint account of the events related to the trial. Once the evidence has been thoroughly discussed, the

group takes a vote about the verdict preferences of individual jurors. In contrast, in a verdict-driven discussion, jurors begin by taking a vote at the start of deliberations. Discussion is subsequently organized around jurors' votes (e.g., innocent or guilty) and evidence is typically discussed in terms of supporting one side or the other. Another dynamic that occurs during deliberation is the tendency for jurors to align themselves with others holding similar positions.

Psychological research has found the amount of group participation to be associated with individual characteristics. For example, men typically speak more than women during deliberation. Individuals with more education and higher-status occupations tend to dominate group discussion. The foreperson is also quite active in the deliberation process and may serve as an influential voice during deliberation. Studies have shown that most juries have several nonparticipating jurors during the deliberation process. They speak rarely during the group discussion and have little impact on the group; their only contribution may consist of taking part in voting. As a result, during deliberations the jurors who dominate discussions have more potential to influence the views of other jurors. At the same time, jurors who withdraw from critical discussion may be less influential in deliberations.

In most deliberations, the first vote, or ballot, will often reveal what the verdict will be. The majority opinion is usually the final verdict. However, minority votes can sometimes prevail. In criminal trials, they are often more successful in persuading the majority to overturn their votes when arguing for an acquittal as opposed to conviction. Minority jurors can also influence the majority in civil trials. If the jury is asked to award damages, minority jurors can be influential in creating a compromise on the verdict or award amount.

Jury deliberation serves critical functions. Research shows that it promotes jury competence by providing the opportunity for a collective pooling of information and the correction of mistaken memories or conclusions. In addition, deliberation also clarifies and solidifies jurors' positions. The consensus about the evidence and the verdict that typically emerges also increase the legitimacy of the jury's eventual decision in the case.

#### **4.1.2. Persuasion during Deliberation**

Social psychologists have used the vehicle of jury deliberation to examine social influence and persuasion processes in small groups. In turn, theoretical advances

have been applied to jury phenomena. Theories about group phenomena, including conformity effects and group polarization, have been examined in the context of the jury deliberation process. Conformity effects of the majority's initial opinions have been documented in jury deliberations, where the initial vote distribution is a strong predictor of the jury's final verdict.

Informational influence and group polarization found in mock jury deliberations suggest that juries operate in much the same fashion as many other task groups. Social psychological theories about group process have been used to assess the impact of different structural variables related to the jury such as composition, size, and decision rule, as discussed below.

#### **4.1.3. Hung Juries**

Many factors can impact why a jury deadlocks, or is "hung," including individual, group, evidentiary, and institutional factors. Individual factors could be incompetence, attitudes, and lack of social skills. An ineffective foreperson or group leader can impede deliberation, for example, by not permitting everyone to speak, allowing one or two jurors to dominate the deliberation, and ignoring arguments by minority jurors. An ineffective decision-making process, by balloting at the start of deliberations as opposed to thoroughly discussing the evidence prior to voting, can also influence vote outcome. Evidentiary factors are also significant; the ambiguity of the evidence can influence the propensity for a jury to deadlock.

Jury research has found that members of hung juries tend to vote earlier than members of verdict juries. Their deliberation appears to be verdict driven, as opposed to evidence driven. When juries do hang, group dynamics are at work. There may be a conflict on the jury, there may be unreasonable jurors, or one or two jurors may dominate the deliberation. Jury members may not be open minded, and may spend too much time and effort attempting to persuade and convince fellow jurors. Numerous factors act and interact with one another to influence jury verdict outcomes, from individual characteristics to group dynamics.

## **4.2. Applications**

Understanding the jury decision-making process in terms of group interaction has important implications and applications for real-world situations. For example, psychological research has examined the impact of decreasing jury size, altering the decision rule

requirement for juries, allowing jury anonymity, and improving jury decision making through jury reforms.

#### **4.2.1. Impact of Jury Size and Decision Rule**

The jury's size and its decision rule have gone through a number of controversial changes. With respect to jury size, what was once long assumed to be a constitutional guarantee, a trial by a 12-member jury, was dismantled by a number of important U.S. Supreme Court cases during the early 1970s. The U.S. Supreme Court decided the number 12 was a "historical accident," finding that state court juries of six did not violate the Constitution. Decreasing the requirement from a 12- to a six-member jury was seen as a way to decrease expenses, save time, and increase efficiency at no cost to the jury trial system.

However, psychological research has identified a number of negative consequences of reducing jury size. The larger the jury size, the more likely it is to be representative, and in turn the more likely it is to reflect community views. In reverse, the smaller the jury size, the less likely it is that minorities in the community will be represented on the jury. In addition, the unpredictability of jury verdicts increases as jury size decreases. The smaller juries used in civil trials are more likely to result in a wider range of jury awards.

Furthermore, six-person juries appear to deliberate differently than larger juries. Larger juries have more people to rely upon to recall evidence with greater accuracy, which decreases the likelihood of relying on incorrect information. Jurors in larger groups are more likely to be diverse in terms of backgrounds, experiences, and opinions. This diversity of perspectives increases the overall quality of jury deliberations and increases the validity of the final verdict. One juror within a six-member jury proposing an opposing position is in a significantly weaker position compared to two jurors in a twelve-member jury.

In addition to changing jury size requirements, countries have modified the long-held requirement that juries must be unanimous in their verdict. During the 1970s, the U.S. Supreme Court decided that unanimous juries in state courts were not constitutionally mandated. The United Kingdom switched from requiring unanimity to permitting majority verdicts as well. These changes were again based on unsupported assumptions about how juries deliberate. Subsequent research shows that eliminating jury unanimity has negative effects on the quality of juror deliberations. For example, jurors with majority rule verdicts instead of unanimous verdicts are more likely to be verdict driven than evidence driven.

Although verdict-driven deliberations may occur in jury trials that require a unanimous verdict, they are more likely to occur in majority rule juries and are more likely to decrease the quality of the deliberation process.

Jurors deliberating under a unanimity rule requirement are more likely to be evidence driven, and evidence-driven juries are more likely to have longer and more thorough deliberations. A unanimity requirement increases the likelihood that all viewpoints will be appropriately addressed during the decision-making process. Majority rule juries are more likely to be verdict driven, resulting in deliberations that are less thorough and significantly shorter. Verdict-driven deliberations are shorter and less thorough because once the majority requirement is reached, there is no need to listen to the remaining minority opinion, resulting in a termination of deliberations. The quality of deliberation decreases because minority opinions do not get as much attention; when they are addressed, they are not sufficiently discussed. As a result, it should not be surprising that jurors on non-unanimous juries report less satisfaction with the process and the verdict, while jurors on unanimous juries report greater satisfaction with both the process and the final verdict.

#### **4.2.2. Anonymous Juries**

Another controversial technique used to improve the decision-making process is juror anonymity. Juror anonymity means withholding the identities of the jurors from the public, the media, and the attorney. There are a number of reasons why judges have felt it necessary to order that the jury members be anonymous: to prevent corruption, to decrease fear of criticism from unpopular decisions, and to decrease fear of retaliation or harassment in high profile cases.

Many critics have claimed that jury anonymity decreases juror accountability, the ability of lawyers to adequately investigate juror bias during voir dire, and public confidence in the fairness of jury trials. However, some of these concerns can be managed. With anonymous jurors, lawyers have access to the same questionnaires used in regular trials and also maintain the ability to question jurors; the only difference is that jurors are identified by numbers instead of names. Preserving the questioning process during voir dire, which uncovers important background and attitudinal information of a potential juror, seems more important for creating an unbiased jury than knowing the juror's name.

Juror anonymity is also viewed as improving the overall deliberation process. For example, jurors in high-profile

cases may fear retaliation or harassment from the defendant, the lawyers, the media, or the public. As a result, maintaining anonymity may increase general participation from the population and also increase the quality of deliberations, due to decreased fear of retaliation. Anonymous jurors may be more comfortable discussing unpopular or controversial topics, resulting in an increased richness of deliberations. Also, anonymous juries may be freed from the pressure of following popular opinion in high-profile trials. There are many potential benefits to anonymous juries, psychological and otherwise, that could increase the quality of the overall decision-making process, and in turn increase the fairness of jury verdicts. They deserve systematic examination and study.

#### **4.2.3. Active Jury Reforms**

Psychological research indicates that modifying jury size and decision rule may produce some negative consequences for jury decision making. In addition, psychological research has suggested that other jury reforms may improve the overall trial process and jurors' individual and group decision making. One controversial reform proposition is based on the idea of active jurors, allowing jurors to discuss evidence throughout the trial, instead of waiting until the end of the trial as is commonly the case. Permitting trial discussions is based on the belief that jurors form opinions about evidence and testimony throughout the trial, and discussing evidence and questions as they arise will help dispel inaccurate interpretations or beliefs and improve the quality of final deliberations.

The state of Arizona is one place where the concept of the active juror has been vigorously pursued. Following substantial jury reforms, Arizona now permits juror note taking, question asking, and trial discussions, all with the aim of improving juror competence. Research studies of some of these reforms show strong support among jurors and judges, but less enthusiasm for the reforms among attorneys. Attorney fears about distraction and prejudice, however, have not been borne out in systematic studies of the operation of the reforms.

### **5. COMPARATIVE RESEARCH ON JURIES**

Much of the empirical jury research to date has taken place within the United States, England, and Canada,

studying how the jury within a single country operates. Yet, countries that have adopted the jury have molded the institution to fit their unique circumstances and cultures. Researchers have only recently begun to take a comparative approach, analyzing how different jury systems function.

Examining how different procedures in various countries affect jurors' decision-making ability is an emerging area for psychological research. For example, although both the United States and England use adversarial models of procedural justice, the degree of involvement of judges and lawyers varies significantly. English judges, who are permitted to comment upon the evidence, may be more influential in the typically more civil and cooperative British jury trial, while U.S. juries may be more distracted by the greater adversariness of the typical American trial.

Another important question that may be answered through comparative research is how juries measure up to other methods for injecting lay sentiment into law, such as lay judges, lay magistrates, and mixed tribunals, which combine lay judges with law-trained judges. One key finding is that when lay judges decide cases together with law-trained judges, the legally trained judge tends to dominate, minimizing the input of lay perspectives. In contrast, the jury decides in isolation, relatively protected from the direct influence of the law-trained judge. This new comparative work, now in its infancy, holds promise for understanding the consequences of different types of lay participation systems.

### **6. CONCLUSION**

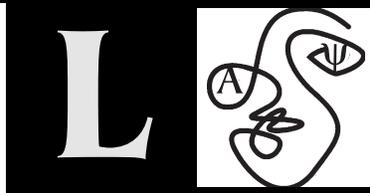
The jury represents an important vehicle for the expression of public perspectives on justice. Critics of the jury system would prefer to weaken or silence the public's voice. Yet psychological research offers empirical evidence that has revealed the basic soundness of jury decision making while simultaneously pointing toward effective reforms of the jury system. Acknowledging jurors as active yet imperfect actors in the trial process allows legal practitioners and researchers to propose jury trial reforms that have the ability to improve the presentation of evidence and maximize the quality of deliberations to not only increase the satisfaction of jurors, but also reinforce the legitimacy of their final verdict.

#### **See Also the Following Articles**

Decision Making ■ Eyewitness Testimony

### Further Reading

- Abramson, J. (2001). *We, the jury: The jury system and the ideal of democracy*. Cambridge, MA: Harvard University Press.
- Dann, B. M. (1993). 'Learning lessons' and 'speaking rights': Creating educated and democratic juries. *Indiana Law Journal*, 68, 1229–1279.
- Diamond, S. S. (1990). Scientific jury selection: What social scientists know and do not know. *Judicature*, 73, 178–183.
- Hans, V. P. (2000). *Business on trial: The civil jury and corporate responsibility*. New Haven, CT: Yale University Press.
- Hans, V. P., & Vidmar, N. (1986). *Judging the jury*. Cambridge, MA: Perseus.
- Hastie, R. (Ed.). (1998). *Inside the juror: The psychology of juror decision making*. Cambridge, UK: Cambridge University Press.
- Horowitz, I., & Bordens, K. (2002). The effects of jury size, evidence complexity, and note taking on jury process and performance in a civil trial. *Journal of Applied Psychology*, 87, 121–130.
- Kalven, H., Jr., & Zeisel, H. (1966). *The American jury*. Boston, MA: Little, Brown, & Co.
- Munsterman, G. T., Hannaford, P. L., & Whitehead, G. M. (Eds.). (1997). *Jury trial innovations*. Williamsburg, VA: National Center for State Courts.
- Vidmar, N. (Ed.) (2000). *World jury systems*. Oxford: Oxford University Press.



# Labor Turnover

Peter W. Hom

Arizona State University, Tempe, Arizona, USA

1. Introduction
  2. Identifying Who Quits
  3. Theories about Why Employees Quit
  4. Methods for Controlling Turnover
  5. Conclusion
- Further Reading

## GLOSSARY

**ecological fallacy** The assumption that the relationship between two variables is the same across different levels of analysis (e.g., individual vs organizational).

**employee turnover** Voluntary cessation of organizational membership by members who receive monetary payment for their participation and labor.

**expectation-lowering procedure** Orientation for new recruits that deflates initial job expectations without providing a realistic preview about a particular job.

**fit** A component of job embeddedness that represents employee compatibility with the company or environment.

**image violation** An experienced event that violates personal ideals or aspirations, inducing thoughts of quitting.

**job avoidance** Less extreme forms of avoiding the workplace.

**job embeddedness** A composite construct representing the web of connections that keep employees from leaving.

**job satisfaction** The favorability of an employee's attitude toward the job.

**links** The formal and informal relations that an employee forms with members in the organization and community.

**matching script** A preexisting plan to exit the job once some preconditions are met.

**met expectations** The alignment between the prehire job expectations of beginning employees and their later work experience that affects their socialization to the job and retention.

**organizational commitment** Employee identification with the values and goals of the employing institution and feelings of loyalty toward that institution.

**realistic job previews** Full communication of the positive and negative aspects of the job to new entrants during hiring or orientation to the job.

**sacrifices** A component of job embeddedness that refers to material or psychological benefits that an incumbent forfeits when he or she vacates the job.

**shocks** Jarring events that trigger thoughts of quitting.

**weighted application blank** A detailed application form that is converted into a selection test to screen out job applicants with a high propensity to quit.

**work-family interference** Organizational interference with employees' ability to participate in family or extraorganizational activities.

Employee turnover is the voluntary termination of membership in employing institutions that is initiated by employees. Over the years, scholars have identified which employees tend to sever their employment ties, why they do so, the process by which they leave, and how they might be prevented from quitting. Although important strides toward the understanding and prediction of turnover have been made, several areas of inquiry remain unresolved or unexplored.

## 1. INTRODUCTION

Employee turnover is the voluntary cessation of membership in an organization by an individual who receives monetary compensation for participation in that organization. This phenomenon has long been an inquiry of investigation for industrial and organizational psychologists, with more than 1000 studies having appeared over the years. This article briefly reviews the theoretical and practical significance of this behavior, investigations of the identities of leavers, theory and research about the causes of employee turnover, and the prevention and control of this behavior.

Turnover attracts the attention of organizational scientists and employers alike because this highly visible behavior may signal poor quality of working life and ineffective organizational processes. In particular, turnover often imposes significant financial costs to firms. When employees quit, employers must expend funds to find and train replacements. To illustrate, human resources professionals project that the cost of one turnover incident ranges between 93 and 200% of the leaver's salary. Apart from direct financial costs, turnover may engender broader costs. For example, excessive personnel attrition may impair customer service and/or harm patients. Moreover, many chief executive officers of leading corporations, such as Microsoft and General Electric, believe that the flight of top talent jeopardizes firm performance. Finally, U.S. corporations worry that excessive attrition among minorities and women may hamper their efforts to diversify their workforces, especially top management teams.

## 2. IDENTIFYING WHO QUITS

### 2.1. Demographic and Personality Correlates of Turnover

Turnover researchers have long sought to identify who quits. According to a comprehensive meta-analysis, Griffeth and colleagues observed that employees who are prone to quit generally are young and single, are not married to working spouses, have no (or few) children, and/or have short tenure in the organization. Contrary to popular opinion, this meta-analysis detected no racial or gender differences in quit rates. Nonetheless, these conclusions should be regarded with caution. Griffeth and colleagues' meta-analysis included only seven studies on the race–turnover relationship and combined all racial minorities together (ignoring racial

differences in quit rates) while failing to examine subgroup differences among women (e.g., educated women and uneducated women quit at different rates). In comparison, the Attrition and Retention Consortium, which is composed of 25 Fortune 500 corporations that share quit statistics on managers and professionals for benchmarking, found that women quit more than do men. Based on a combined sample of 81,346 racial minorities, this consortium further observed that Hispanic and African Americans leave at higher rates than do European and Asian Americans during the early period of employment. Because such findings may expose consortium participants to litigation or embarrassment, such data are not available from the academic literature (and so their omission biases demographic meta-analyses).

In addition to demographic traits, turnover researchers have examined other personal attributes of leavers. Specifically, a meta-analysis of the “Big Five” personality dimensions reported that long-tenured employees tend to be conscientious, agreeable, and open to experience. Supporting the old maxim that the “best predictor of future behavior is past behavior,” it was further demonstrated that employees with a history of unstable employment exhibit a greater risk of quitting.

### 2.2. Performance–Turnover Relationship

By far, the personal characteristic attracting most turnover studies is job performance because whether high or low performers tend to quit greatly determines the costliness of turnover. Organizations may even welcome personnel attrition if those who are exiting are their less valuable or less valued job incumbents. Testing the performance–turnover relationship, several meta-analyses have concluded that substandard performers are more prone to quit, whereas outstanding performers are more stable.

Despite such impressive consistency, some researchers have witnessed inverted U-shaped turnover–performance relationships, such that superior performers (because they have greater employment opportunities) and poor performers (because they are unhappy or face impending dismissals) terminate employment more often than do moderate performers. Conventional wisdom had long assumed that greater flight among top performers occurs only in occupations where achievements are publicly known (e.g., professional athletes, university scholars). Because their professional accomplishments are visible

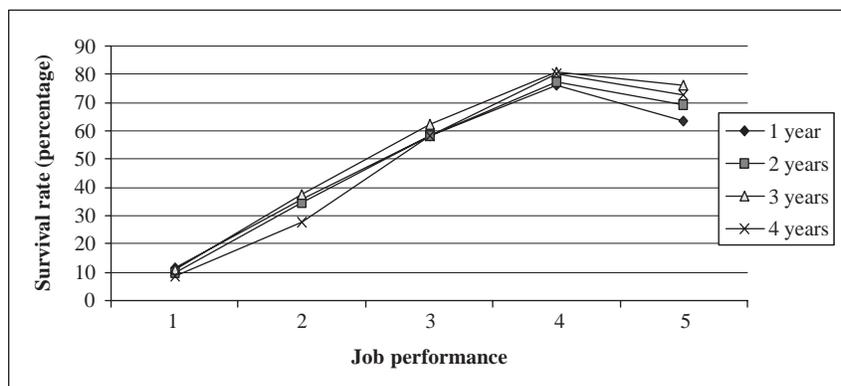


FIGURE 1 Job survival rates as a function of job performance and firm tenure.

signals to the external marketplace of their ability, superior performers can readily secure alternative employment elsewhere (including receiving more unsolicited job offers from outside employers). Recently, a few studies have detected that this trend toward greater job mobility among productive incumbents is more common, reporting higher turnover rates among high-performing managers and Swiss bankers (Fig. 1). As a result, these recent findings qualify traditional observations that low performers exit voluntarily at greater rates than do high performers. When professional accomplishments are externally visible or objectively verifiable (e.g., fast-track promotion history for managers), curvilinear (rather than inverse) relationships between performance and turnover may well emerge.

### 3. THEORIES ABOUT WHY EMPLOYEES QUIT

#### 3.1. Theories about Intermediate Linkages between Job Dissatisfaction and Quits

More than 45 years ago, March and Simon pioneered a theory about why employees quit that continues to dominate current perspectives. They conceptualized two prime turnover determinants—"ease of movement" (e.g., job opportunities that allow employees to quit) and "desirability of movement" (e.g., job dissatisfaction that motivates employees to leave)—that persist in recent theorizing. Extending this line of thinking, Mobley first elaborated the various stages of the withdrawal process. Specifically, he theorized that job dissatisfaction prompts employees to begin

thinking about quitting. Dissatisfied incumbents next evaluate their prospects of obtaining employment elsewhere and the costs of quitting (e.g., loss of seniority benefits). If attractive alternatives exist and turnover costs are low, disgruntled employees form intentions to search for other jobs and actively pursue other work. After securing other job offers, these employees compare them with their current positions. If this comparison favors an alternative, they will ultimately quit.

Since its inception, Mobley's model had dominated theory-testing research throughout the late 1970s and 1980s, yet empirical tests failed to sustain Mobley's causal structure. Instead, in 1991, Hom and Griffeth validated an alternative sequence of structural linkages among withdrawal antecedents (Fig. 2).

In 2001, Hom and Kinicki validated an extended version of this model (Fig. 3) with a national retail store sample. This version additionally considered local unemployment conditions, work-family conflict (i.e., job interference with employees' ability to attend to family and/or extraorganizational pursuits), and job avoidance (i.e., less extreme forms of avoidance such as absenteeism and tardiness).

In this theory, interrole conflict increases dissatisfaction and withdrawal cognitions, whereas unemployment conditions directly lower quit rates. Furthermore, job dissatisfaction increases job avoidance, which in turn augments withdrawal cognitions.

#### 3.2. Structural Determinants of Turnover

Apart from theoretical accounts of how dissatisfaction translates into exits, Price and Mueller introduced another framework that also initiated a long research

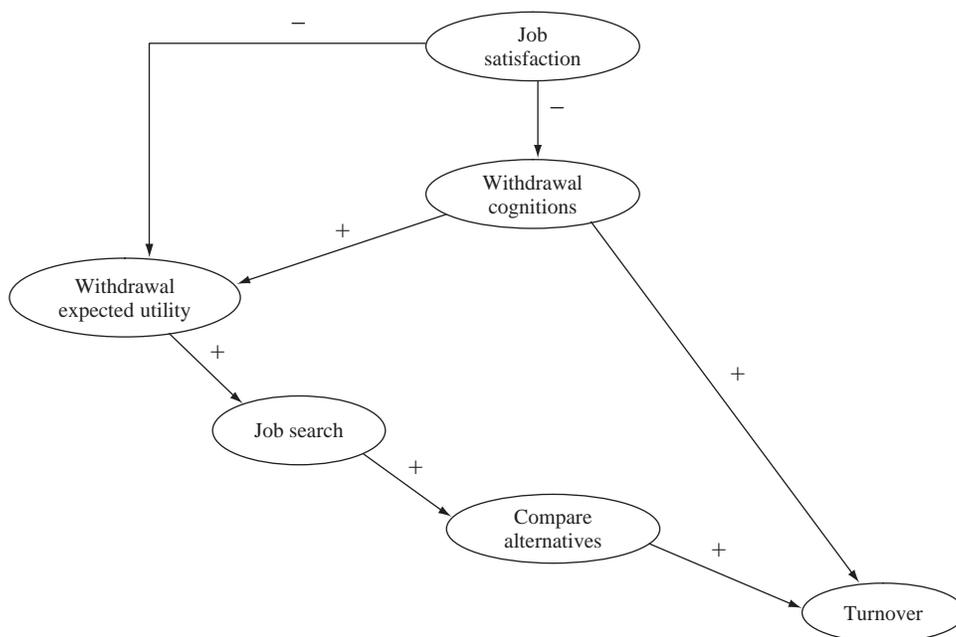


FIGURE 2 Hom-Griffeth model of dissatisfaction-initiated quits.

stream. Unlike psychologists (who focus on micromediation among psychological constructs), these sociologists identified a broad array of more distal structural antecedents of turnover (e.g., characteristics of the work setting, patterns of social interaction within the workplace) as well as influences external to the workplace

(e.g., occupational commitment, family obligations). Price and Mueller theorized that role overload, routine work, centralized decision making, work-related communications, social integration (i.e., friendships at work), pay, fair reward allocation, and promotional opportunities affect job satisfaction, whereas

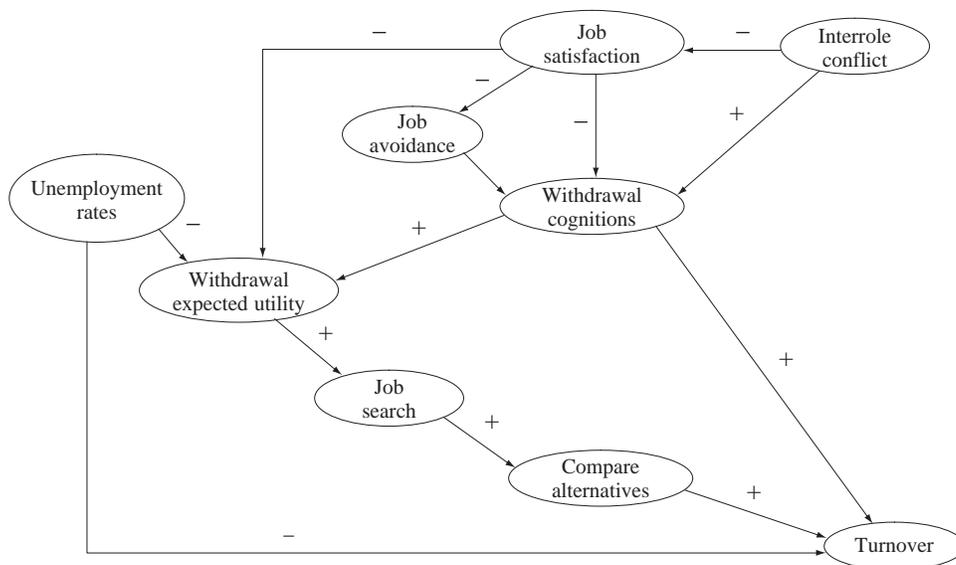


FIGURE 3 Hom-Kinicki model: Expanded theory of dissatisfied quits.

professionalism (i.e., commitment to one’s professional standards and norms), general training, and “kinship responsibility” (i.e., family responsibilities and relatives living nearby), as well as job satisfaction, underpin organizational commitment (i.e., firm identification and internalization of firm values). Commitment and perceived job opportunities, in turn, stimulate the decision to quit, and this decision, as well as perceptions of job opportunities, affects turnover. Although studies have not validated every pathway in the Price–Mueller model, studies have broadly established that structural determinants affect either job satisfaction or organizational commitment—or both—and that environmental antecedents affect quits (although whether their effects are direct or indirect via commitment or quit intentions remains indeterminate).

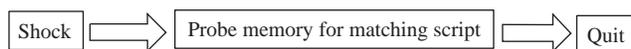
### 3.3. Unfolding Model

Marking a radical departure from traditional perspectives, Lee and Mitchell proposed a provocative new theory to overcome the shortcomings of these viewpoints. They argued that conventional thinking wrongly assumes that job dissatisfaction is the prime impetus for quitting and that prospective leavers quit only for other employment. Prevailing schools of

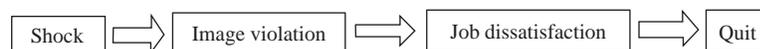
thought also presume that prospective leavers evaluate alternatives using a rational analysis of their costs and benefits. Instead, Lee and Mitchell’s “unfolding model” posits that nonattitudinal causes—namely “shocks” or jarring events that stimulate thoughts of quitting—instigate quits more often than do attitudes toward the job (Fig. 4). Their formulation also holds that some leavers exit without jobs in hand (including, e.g., exiting the workforce to return to school) and that leavers often make quick “compatibility” evaluations of alternative jobs rather than subject them to rational subjective expected utility analyses.

The unfolding model further proposes that leavers follow one of several pathways when exiting. Some leavers follow Decision Path 1, which is precipitated by a shock (typically some expected personal event). This shock causes incumbents to probe their memory for a “matching script” or preexisting action plan, and if one is detected, this can lead to automatic withdrawal. For example, a woman may quit when she becomes pregnant (the shock) because she had already planned that she would stay at home to raise her first child on becoming pregnant. Other leavers follow Decision Path 2 because they experience a shock that triggers an “image violation.” Lee and Mitchell defined three types of images: value image (e.g., personal

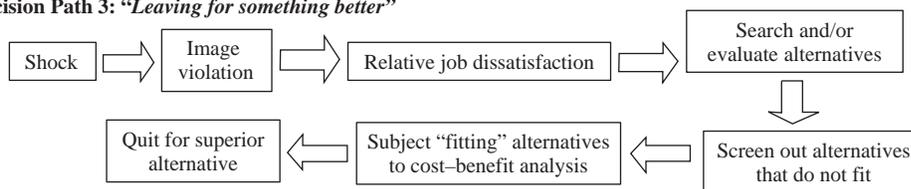
**Decision Path 1: “Following a plan”**



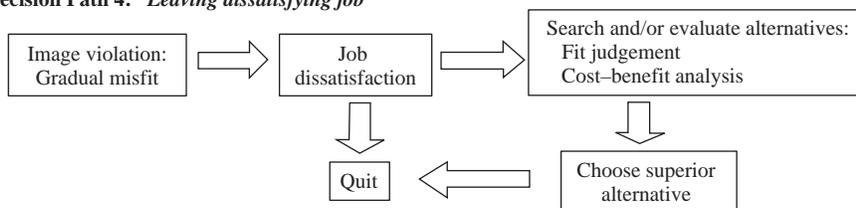
**Decision Path 2: “I’m outta here”**



**Decision Path 3: “Leaving for something better”**



**Decision Path 4: “Leaving dissatisfying job”**



**FIGURE 4** Lee and Mitchell’s unfolding model.

values), trajectory image (e.g., personal goals), and strategic image (e.g., behavioral strategy toward goal attainment). In this turnover path, a shock violates employees' images and causes them to feel dissatisfied and to directly quit. For example, a hospital may shift to team-based nursing (the shock), which violates a nurse's image because the nurse values individual-based nursing. The nurse becomes unhappy and leaves the hospital without engaging in a prior job search.

Still other leavers follow Decision Path 3, which is initiated by some unsolicited job offers or inquiries (the shock). This shock leads job incumbents to compare these new jobs with their current ones and leads to "relative" dissatisfaction if the alternatives are superior (i.e., incumbents experience "opportunity costs" by staying). The prospects of "greener pastures" elsewhere motivate incumbents to pursue and evaluate other alternatives (including the unsolicited jobs). They first screen alternatives using a quick compatibility test to determine fit with their personal images, rejecting those that lack such a fit. Next, the job seekers engage in more deliberate cost-benefit analyses among alternatives meeting the fit test and ultimately quit to take the "best" alternative. To illustrate this path, Lee and colleagues described a nurse who received a job offer from a private clinic (the shock). She had been frustrated for a while but had not contemplated leaving until someone offered her a better position. She then investigated this clinic position thoroughly before exiting the hospital.

Finally, some leavers take Decision Path 4 (regarded by Lee and Mitchell as the conventional path, similar to that portrayed by the Hom-Kinicki model). Here, leavers gradually become disenchanted with their jobs as they progressively experience image violation. Some may quit as soon as their dissatisfaction reaches some threshold. This is what Jim Manzi, chief executive officer of Lotus Development, did when IBM bought his company. Manzi became unhappy with the direction and vision for the newly merged companies and left without a new job in hand. Alternatively, some leavers who take Decision Path 4 seek and judge other employment options and exit once they secure better positions.

Using qualitative methodology, empirical tests of the unfolding model interview former employees to delineate the routes by which they withdrew from their previous jobs. More than 92% of leavers followed one of the four decision paths depicted by the unfolding model. Moreover, the studies employing these tests report that 25% of leavers quit without searching for employment elsewhere and that shocks prompt more turnover than does job dissatisfaction. Finally, 39% left

due to external shocks such as pregnancies, unsolicited job offers, and spousal relocations. Despite these provocative findings, retrospective reports, which may be distorted by postdecisional dissonance and faulty recall, provide inadequate validation for the unfolding model. Further verification of this theory requires predictive longitudinal designs—a standard methodology for verifying turnover models.

### 3.4. Job Embeddedness

In 2001, Mitchell and colleagues proposed explanations for why people stay. Existing formulations explain why people leave but fail to account for why people stay, possibly implicating different psychological processes. Mitchell and colleagues introduced "job embeddedness" as a composite construct representing the web of connections that keep employees on the job. Job embeddedness is composed of three dimensions: links, fit, and sacrifice. Links represent the formal and informal connections between employees and institutions and others. Thus, friendships at work and in the community, as well as extended family in the community, may keep employees from abandoning their workplaces. Fit represents employee compatibility with the company and environment. If employees have values compatible with those of the workplace, enjoy recreational and entertainment activities in the community, and belong to local community organizations (e.g., churches), they may stay. Finally, sacrifices correspond to the costs of material and/or psychological benefits that are forfeited by resignations. Thus, employees stay if they have nonportable job benefits (e.g., pensions), seniority-based benefits, and community advantages (e.g., live in safe neighborhoods close to work, have prime seating at sporting events). Testing this construct, Mitchell and colleagues documented that job embeddedness accounts for unique variance in the prediction of turnover that is not explained by conventional predictors.

## 4. METHODS FOR CONTROLLING TURNOVER

In contrast to the substantial explorations of turnover theories, examinations of ways in which to reduce turnover remain limited. In 2001, Griffeth and Hom discussed the more promising approaches that have been supported empirically.

### 4.1. Employee Selection

The most widely validated selection tool is the weighted application blank (WAB), that is, a detailed application form converted into a selection test. To develop a WAB, an employer would compare application blank responses of former employees (leavers) and current employees (stayers). When these groups answer a particular question differently (e.g., stayers might report higher incomes on their last jobs than do leavers), a scoring key is generated for that question and “scores” the leavers’ response (low wage) higher than the stayers’ response (high wage). Scoring keys are developed for all questions that discriminate between stayers and leavers. The sum of the special scores for all items yields an overall score for quit propensity. When used for hiring, the WAB essentially compares job applicants’ answers on the application blank with those of leavers (former employees). If certain applicants closely resemble leavers in their application responses, these applicants would receive high WAB total scores and would be disqualified from hiring. Despite the simplicity and proven success of WABs, few employers use them because questions about certain background attributes may disproportionately reject minority or women applicants (and prompt discrimination lawsuits) or may constitute an invasion of privacy.

In addition to assessing background attributes to identify applicants at risk for quitting, organizations might consider personality testing. However, conventional personality scales are susceptible to faking by job seekers who (mis)portray themselves in a favorable light to gain employment. To overcome such social desirability, Bernardin suggested a personality-like measure of quit propensity using a forced-choice format. In this approach, respondents choose a descriptor (about themselves) from a pair of descriptors that are matched on equal discomfort, although only one is a valid answer. To create this test, employers must interview current employees and their superiors to identify discomforting aspects about a job (e.g., “your schedule for work changes from day to day”). They then write statements based on these events (questionnaire items). Next, employers also write statements about discomforting situations that are irrelevant to the job (e.g., “buying a new television and later learning that it goes on sale next week”). Then, a set of judges is recruited to rate the discomfort levels of relevant and irrelevant statements. Then, employers pair one relevant statement with an irrelevant statement that is matched for discomfort. Once many sets of paired statements are identified,

they are included in the final personality test. When taking this test, prospective employees would select a statement from each pair depicting an event most distressing to them. If they choose the “valid” statement (representing an actual negative job experience) from a given pair, they would earn a “higher” score for that answer. The test would indicate that applicants are a poor fit for a job if they accumulate high test scores (based on summing their responses to all statement pairs). Put differently, these individuals would likely be upset by discomforting events that are part of the job. They would ultimately become dissatisfied and quit. Bernardin and colleagues established that this personality inventory predicts turnover among customer service representatives and motion picture theater personnel.

Finally, selection interviews have potential for identifying prospective leavers if they are properly designed and executed. Notably, Schmidt and Rader described a promising structured telephone interview. To develop this screening device, employers must carry out in-depth interviews with outstanding performers in a particular job to identify their behavioral tendencies and traits (i.e., “themes”). Next, potential interview questions assessing each theme are written out as potential interviewees would be asked how they exhibit such traits. During the next step, two groups of employees—30 high performers and 30 low performers—are interviewed with these questions individually, and their remarks are transcribed. Questions differentiating between high and low performers are retained, and a scoring guide that credits interview responses indicative of the presence of high-performer themes is created. During telephone interviews, job applicants would be asked to respond to these questions (and to provide examples). Applicant responses would also be recorded and later scored by trained analysts to derive scores that assess how well applicants exhibit the high-performer themes. Summarizing multiple validity studies, Schmidt and Rader estimated that this interview predicts employment stability moderately accurately ( $r = .39$ ).

### 4.2. Interventions for Reducing Turnover

Turnover scholars have most validated realistic job previews (RJPs), that is, comprehensive profiles of a job—portraying its positive and negative aspects—that are communicated to new hires during hiring or orientation. Meta-analyses conclude that RJPs modestly reduce exits among beginning employees in a wide range of

jobs (e.g., nursing, military) with a variety of media (e.g., booklet, audiovisual). To account for the effectiveness of job previews, theorists have often cited met expectations and self-selection as underlying mechanisms. According to met expectation theory, RJPs reduce reality shock by decreasing the gap between newcomers' initial (often inflated) expectations about a new job and subsequent work experience. RJPs clarify what the job entails and increase the odds that newcomers' expectations will be met on the job, thereby improve retention. The self-selection mechanism posits that RJPs help job applicants to make more informed job choices, in turn helping them to better decide whether or not their needs would be satisfied by the job. Those who self-select themselves for the job (despite the RJPs) are better suited for the job and will likely stay, whereas those who are "misfits" will decide not to accept the job (and avoid becoming future turnover casualties). Despite the popularity of these explanatory schemes, empirical corroboration of RJP explanations remains limited.

In 1998, Buckley and colleagues introduced the expectation-lowering procedure (ELP), which reduces job expectations without providing actual information about the job. After all, expensive job analyses are needed to properly design RJPs, and they may lose their efficacy if job content changes. To offset these limitations, Buckley and colleagues designed a special orientation for new hires, who are instructed in the theory of met expectations (and are explained the causes and consequences of unmet expectations) and share their own personal experiences with reality shock. In an experimental test, Buckley and colleagues showed that the ELP reduced turnover among new assembly line workers as much as did an RJP.

Finally, turnover scholars have validated job enrichment—increasing the intrinsic rewards of a job by expanding skill variety and empowering incumbents—as another means to promote job loyalty. Indeed, a meta-analysis concluded that job enrichment is more effective than are RJPs for promoting retention, supporting Wanous's maxim that RJPs cannot substitute for good jobs.

## 5. CONCLUSION

Organizational studies have greatly expanded our understanding of the turnover phenomenon. Nonetheless, much theoretical and empirical work remains. Specifically, empirical verification of turnover formulations has lagged theoretical advances. Conventional static designs typically assess turnover causes at one time and

use them to predict turnover at a later time. This standard methodology overlooks the implicit causal assumption inherent in most theoretical frameworks that changes in turnover's antecedents foreshadow its occurrence. To better corroborate causality, more rigorous evaluations of prevailing theories should apply latent growth modeling or hierarchical linear modeling to longitudinal data (i.e., multiple waves of predictor assessments).

Moreover, conceptual models of the withdrawal process have been validated primarily in North American samples. Yet American transnational firms increasingly employ global workforces and face the daunting challenge of how to bind their workers to employment. Because retention strategies based on ethnocentric domestic theories may falter abroad, it is imperative that researchers validate turnover models (and loyalty-inducing practices) in various cultural settings. Furthermore, theorists have narrowly focused on explaining why individuals quit and derive organization-wide practices for reducing aggregate quit rates from individual-level findings. Yet inferring company-wide prescriptions from individual-level data and theories risks committing the "ecological fallacy." Accordingly, future research should develop (and test) formulations that account for why rates of personnel attrition vary across organizations.

## See Also the Following Articles

Conflict within Organizations ■ Organizational Culture and Climate ■ Person-Environment Fit ■ Personnel Selection ■ Traits

## Further Reading

- Bernardin, H. J. (1987). Development and validation of a forced choice scale to measure job related discomfort among customer service representatives. *Academy of Management Journal*, 30, 162-173.
- Buckley, M. R., Fedor, D. B., Veres, J., Wiese, D., & Carraher, S. M. (1998). Investigating newcomer expectations and job-related outcomes. *Journal of Applied Psychology*, 83, 452-461.
- Griffeth, R., & Hom, P. (2001). *Retaining valued employees*. Thousand Oaks, CA: Sage.
- Griffeth, R., & Hom, P. (2004). *Innovative theory and empirical research on turnover*. Greenwich, CT: Information Age.
- Griffeth, R., Hom, P., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26, 463-488.

- Hom, P. W., & Griffeth, R. (1991). Structural equations modeling test of a turnover theory: Cross-sectional and longitudinal analyses. *Journal of Applied Psychology, 76*, 350–366.
- Hom, P., & Griffeth, R. (1995). *Employee turnover*. Cincinnati, OH: South-Western.
- Hom, P., & Kinicki, A. (2001). Toward a greater understanding of how dissatisfaction drives employee turnover. *Academy of Management Journal, 44*, 975–987.
- Lee, T., & Mitchell, T. (1994). An alternative approach: The unfolding model of voluntary employee turnover. *Academy of Management Review, 19*, 51–89.
- Lee, T. W., Mitchell, T. R., Wise, L., & Fireman, S. (1996). An unfolding model of voluntary employee turnover. *Academy of Management Journal, 39*, 5–36.
- March, J., & Simon, H. (1958). *Organizations*. New York: John Wiley.
- Mitchell, T., & Lee, T. (2001). The unfolding model of voluntary turnover and job embeddedness: Foundations for a comprehensive theory of attachment. In B. Staw, & R. Sutton (Eds.), *Research in organizational behavior* (Vol. 23, pp. 189–246). Oxford, UK: Elsevier Science.
- Mitchell, T., Holtom, B., Lee, T., Sablinski, C., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal, 44*, 1102–1121.
- Mobley, W. (1982). *Employee turnover*. Reading, MA: Addison-Wesley.
- Price, J., & Mueller, C. (1986). *Absenteeism and turnover of hospital employees*. Greenwich, CT: JAI.
- Schmidt, F., & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology, 52*, 445–464.
- Villanova, P., Bernardin, J., Johnson, D., & Dahmus, S. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnover of motion picture theater personnel. *Personnel Psychology, 47*, 73–90.
- Wanous, J. P. (1980). *Organizational entry: Recruitment, selection, and socialization of newcomers*. Reading, MA: Addison-Wesley.
- Wanous, J. P. (1992). *Organizational entry* (2nd ed.). Reading, MA: Addison-Wesley.



# Leadership and Culture

Roya Ayman

Illinois Institute of Technology, Chicago, Illinois, USA

1. Introduction
  2. Culture
  3. Leadership Research across Cultures
  4. New Topics in Culture and Leadership
  5. Summary and Conclusion
- Further Reading

## GLOSSARY

**accommodating strategy** When one party in a conflict gives all of his or her rights for another's to win.

**collaboration strategy** When two parties focus on the problem and the best solution in which both can win.

**competitive strategy** When one party in a conflict is focused on him- or herself to win at all cost.

**compromising strategy** When parties negotiate a settlement.

**consideration** Behavioral style that is friendly, open to input, respectful to others, etc.

**contingency approach** Model of leadership that describes leadership in context.

**derived etic approach** The coalescence of theories or measures about a concept that are developed in two different cultures.

**emic approach** A method of studying culture theories, models, and measures as evolved from inside a culture.

**endogenous researchers** Researchers who conduct research within their own culture.

**etic approach** A method of studying culture in which existing theories, models, and measures in one country or culture is validated in another.

**exogenous researchers** Researchers who conduct research outside their own culture.

**global leaders** Leaders who perceive complexity and are inclusive, respecting people of all cultures equally, and who are concerned about what is the best solution in the long run.

**idealized influence** Behavioral styles that provide vision and unify the team behind a common goal.

**imposed etic approach** Approach using existing theories or measures from one culture in another culture with almost no modification other than translations.

**individual consideration** Style of behaviors such as attentiveness to diversity of needs and values of team members and addressing each on their own basis.

**initiating structure** Behaviors that focus on getting the job done, such as goal setting, clearing expectation, criticizing wrong behavior, rewarding good behavior, planning, etc.

**inspirational motivation** Behavioral style that exudes energy and emotion in the actions to be taken.

**intellectual stimulation** The process of getting team members to think critically.

**international leaders** Leaders who have the tendency to impose their own cultural values and policies on others.

**leader-member relationship** The leader's certainty that the team is united and supportive of him or her.

**position power** The leader having a title, and the legitimacy and resources to reward and punish team members.

**situational control** A leader's amount of control and influence in a given situation.

**task structure** A leader's knowledge of how to get information or conduct a task.

**withdrawal** When parties to a conflict leave the relationship.

Globalization of work economy has made the interplay of culture and leadership an important topic of consideration. This article covers the topic of leadership and culture, demonstrating a symbiotic relationship between the two. It is shown that the definition and methods used in the study of culture in leadership can affect the results. Culture as determined by economic-political conditions had the most impact in determining differences in leadership across cultures. Accordingly, managers in countries that are similar in terms of industrialization and modernization have more similarity in their values and even their behaviors. Through the analysis of studies on behavior of leaders across cultures, it became evident that the definition of leadership is highly cultural dependent. More specifically, the meaning of the behavior, the effect of the behavior, and the level of the behavior varied across cultures. One outstanding finding was that Asian cultures do not seem to consider task-related and people-related behaviors as two separate leader behaviors. Additionally, more recent issues of leadership, such as working in multicultural settings and global leadership, are addressed.

## 1. INTRODUCTION

Globalization is an inevitable and challenging evolution in today's society. Leaders in any social and economic institution are at the helm of this evolution of multicultural interdependence. Therefore, a deep understanding of the concept of leadership within cultures, across cultures, and globally is of paramount importance. To gain an appreciation of the interaction between the leader and culture, consider the following anecdote: An Asian manager in the United Kingdom noticed that one of his employees was absent from work due to ill health. The manager visited the employee at home, bringing flowers and a box of sweets, as was customary in the manager's culture. In his mind, he was showing individualized consideration and social support for his employee. Nevertheless, the employee, acting according to his own cultural norms, filed a complaint, charging the manager with distrust and invasion of his privacy. The issue here is not who was correct, but the elucidation of the role that culture plays in interpreting behaviors.

The preceding anecdote illustrates the fact that leadership is embedded in culture, and that the two have a symbiotic relationship. To explore this, this article presents a brief definition of culture and the cross-cultural methods. Next, different leadership definitions and the

role of culture in their universality is discussed. The topics of leadership as seen in some countries, the leader in cultural exchange, and global leadership conclude this presentation of culture and leadership.

## 2. CULTURE

### 2.1. Definition of Culture

"Without men, no culture, certainly; but equally, and more significantly, without culture, no men." Clifford Geertz made this powerful if not precise definition of culture, highlighting the importance of culture in the study of human beings. Scholars concur that culture refers to a constellation of a group's collective values, expectations, attitudes, and behaviors. In addition, the symbiotic relationship between culture and humans is interactive. While cultural environments model humans, humans also affect the culture. In other words, culture refers to an all-encompassing and long-lasting depiction of a human social entity.

Regardless of the definition of culture used, in cross-cultural psychology, researchers focus on two distinctions of culture: cultural values and physical characteristics. Cultural values primarily include (1) collectivism and individualism, (2) power distance, (3) uncertainty avoidance, (4) masculinity or industriousness, (5) Confucianism, and (6) time orientation. Physical parameters include country of origin or residence, as well as physical attributes of individuals, such as gender and color.

### 2.2. Cross-Cultural Methodology

Because results of cross-cultural research are dependent on the method used to study leadership, this section defines the various alternatives to and factors in cross-cultural research. The methods used to incorporate culture into psychological research have evolved since almost 50 years ago. Traditionally, two types of approaches have been recognized: (1) when existing theories, models, and measures in one country or culture are validated in another, or the etic approach, and (2) when theories, models, and measures evolve from inside a culture, or the emic approach. The etic approach is further expanded to include imposed etic and derived etic. Imposed etic refers to using existing theories or measures from one culture in another with almost no modification other than translations. Derived etic, on the other hand, refers to an

approach in which a theory or measure developed in two different cultures coalesces; it is that which is common between the two cultures.

In addition to the different approaches to cross-cultural research, the researcher's cultural background can impact the formation of the research question and the interpretation of results. Therefore, cross-cultural researchers can be categorized by their level of involvement with a culture as either exogenous or endogenous investigators. Exogenous researchers are those who conduct research outside their own culture. Therefore, they are not entangled with the culture, and high degree of objectivity is the outcome. Danger, however, lies in the fact that exogenous researchers' assumptions may be derived from their own culture or their stereotypic view of the culture they are studying. Endogenous researchers, those who conduct research within their own culture, may be less objective but are more sensitive to nuances and meanings of their own culture.

Thus, cross-cultural research usually starts with emic measures in one culture, which then can be tested in the other cultures as imposed etic measures. The items that best describe both cultures will form the derived etic measure. Most research on leadership thus far can be categorized as imposed etic, particularly the validation of theories developed in the United States in other cultures. Additionally, U.S. scholars have conducted most of the research and developed most of the main theories of leadership; hence, the existing knowledge about leadership is heavily influenced by North American values and expectations. Therefore, given the extent to which exogenous researchers from one specific culture have dominated the theories and conceptualizations of leadership, it is mostly representative of one perspective. For this reason, multicultural research teams that include endogenous researchers are the new trend and the recommended form of cross-cultural research.

### 3. LEADERSHIP RESEARCH ACROSS CULTURES

Concern for leadership has a long history. In China in 5000 BC, Confucius advised the Chinese emperors; in 12th-century Persia, Unsur-u'l-Ma'ani wrote an essay advising the Ziari rulers, and in 15th-century Europe, Machiavelli wrote *The Prince*, guiding the rulers of the time. In the late 19th century, Carlyle formulated the Great Man theory, later known as the leadership trait

approach, which had the most impact on the majority of research on leadership in the United States.

Initially, leadership was approached with a focus on either the leader alone, or the leader as part of a social context. When the attention was focused on the leader, his or her traits and values were considered as well as his or her leadership behaviors. The models of leadership that describe leadership in context are referred to as contingency approaches. In the following sections, the etic approaches to leadership highlight the effect of culture on the universality of the leader's traits and behaviors and the validity of contingency approaches to leadership across cultures.

#### 3.1. Etic Approach: The Leader's Characteristics

The trait approach to leadership is the oldest approach that initially incorporated both dispositional characteristics, including values, attitudes, and traits (e.g., intelligence, locus of control, confidence, optimism, and adaptability), and physical characteristics (e.g., height and attractiveness). More than two-thirds of the 20th-century research on leadership focused on identifying dispositions that best describe a leader. Intelligence, adaptability, adjustment, aggressiveness, dominance, independence, originality, self-confidence, and optimism were dispositions that most studies identified as separating leaders from non-leaders.

Cross-cultural comparison has most often focused on managers' attitudes and values. Two approaches are used to examine the effect of culture on managers' similarity of values. One approach examines the values of managers in different countries; the other compares managers from a similar country of origin living in different countries.

When different nationalities were compared, it accounted for approximately 28% of the variance in managers' attitudes. In contrast, when the effect of similarity among managers from Australia, India, Japan, Korea, and the United States was assessed, and nationality versus industrialization was compared, it was the closeness in the level of the societies' industrialization, language, and traditionalism that accounted for most of the leaders' value similarity, not the country they lived in. Although managers from Australia and the United States shared similar values, the Japanese and Korean managers did not show a similarity of values. Such differences were attributed to varying levels of industrialization of the two countries. Therefore, it is not the country or nationality per se, but the proximity in

socio-economic conditions of those countries that determines similarity between managers/leaders.

To isolate the impact of culture from national origin, researchers have contrasted the effect of social/ethnic culture and country of residence. In these studies, the values of managers from various countries were compared to the values of managers who lived in the United States. The result showed that those managers who came from countries that were at similar levels of modernization as the United States had similar values. So, for example, the U.S. managers of both Caucasian and Chinese background had values more similar to those of managers from Hong Kong and more different from those of Mexican and Mexican American managers. This result not only illustrated issues related to difference in industrialization between the United States and Mexico, but it also alluded to the potential reason that Mexico's adaptation to industrialization is relatively more difficult.

Likewise, the results of a study that compared managers with Chinese heritage from three different countries (the People's Republic of China [PRC], Singapore, and the United States) showed that the values of the Chinese managers who lived in Singapore and the United States were quite similar to each other, and were quite different from those who lived in the PRC. Again, the role of modernization and industrialization of the country of residence seems to contribute to the similarity between the leaders' values.

In short, the aforementioned studies illustrated that industrialization and education are powerful societal forces in unifying managers' values across countries. No one specific set of values defined all managers. However, the conditions in which the managers functioned did have an impact in making them more similar.

The majority of research and theories on leader behavior in the United States relies heavily on subordinate perception of the leader's behavior. The content of these behaviors has been as specific as conflict management strategies and influence, and as broad as transactional and transformational leadership. All the various theories have reported some cross-cultural validity in predicting outcomes and some level of universality in the presence of these behaviors.

To review this aspect of leadership and the role of culture, first the behaviors are briefly defined. Subsequently, three cross-cultural concerns that have been presented in testing the universality of these behaviors are discussed. These concerns are (1) comparative level of occurrence of these styles across cultures, (2) difference in behaviors determining the styles across cultures, (3) their functionality across cultures.

Theorists and researchers have identified many leader behaviors with a plethora of empirical research. The oldest leadership behavioral styles studied in U.S. research are consideration and initiating structure. Consideration refers to behaviors that are friendly, open to input, respectful to others, etc. Initiating structure includes behaviors that focus on getting the job done, such as goal setting, clearing expectation, criticizing wrong behavior, rewarding good behavior, planning, etc. These were identified as two distinct categories of leadership behavior that help a group meet its goals of maintaining harmony and accomplishing tasks successfully.

Subsequently, leadership researchers acknowledged that these are transactional-type behaviors, and they recommended that effective leaders go beyond the status quo. Thus, the transactional and transformational leadership paradigm was presented. Behaviors included in transactional leadership style were active or passive monitoring of performance and giving feedback, also the act of giving rewards when due. Transformational leadership included four behavioral styles, intellectual stimulation, individual consideration, inspirational motivation, and idealized influence. Intellectual stimulation is getting team members to think critically. Individual consideration includes styles of behaviors such as attentiveness to diversity of needs and values of team members and addressing each on their own basis. Inspirational motivation consists of behavioral styles that exude energy and emotion in the actions to be taken. Finally, idealized influence includes behavioral styles that provide vision and unify the team behind a common goal. Theorists such as Bass and Avolio have been explicit in clarifying the relationship between the transformational and transactional styles. The issue is not that the leader should engage in one or the other style, but rather that transformational style augments the effects of transactional style. Overall, men and women leaders do not differ greatly in these behaviors; however, women leaders are perceived to be slightly more participative than men leaders. Additionally, women leaders are perceived to be slightly more transformational and to give more contingent rewards as compared to men leaders.

Conflict management and influence tactics are significant leadership behaviors; many scholars define leadership as a process of influencing others to do what they would not do otherwise. Rahim's conflict management model identifies two adversaries and the strategies they may take in managing the conflict between them. When one party is focused on winning at all costs, this is known as competitive strategy. When one party gives all his or her rights for another's to win, this is

considered accommodating strategy. If both parties leave the relationship, it is known as withdrawal. If the parties negotiate a settlement, this is a compromising strategy. Finally, if the two parties focus on the problem and the best solution in which both can win, this is referred to as collaboration strategy. Influence tactics are behaviors identified empirically when people were asked how they influence others. The tactics most commonly supported in U.S. research were rational persuasion, ingratiation and coalition tactics, inspirational and personal appeals, exchange, and assertiveness.

Overall, each of these behavioral paradigms in the United States has received support across settings and populations. That is, both the reliability of the measures assessing the styles and the functionality of the styles in relation to organizational outcomes have received empirical support. Next, the effect of culture on these behaviors is examined.

The first question that many cross-cultural researchers have had when using these theories across cultures is, Do all managers behave similarly across cultures? Empirical studies have found that the magnitude of the manifestations of these styles vary across cultures. This was established by comparing the mean scores of the frequency of occurrence of the various styles in different countries. For example, it was found that New Zealander managers were described as more transformational than transactional, whereas the Taiwanese managers were seen as behaving equally in both transformational and transactional styles. It was also reported that in a manufacturing telecommunication transmission equipment industry, the U.S. managers, in comparison with the German managers, were described as more charismatic and were reported as being more inspirational when motivating their employees.

Some cross-cultural leadership researchers became aware that specific behaviors that define a style might vary from culture to culture. For example, it is reported that in Indonesia for leaders to be able to inspire and motivate their subordinates, they must first persuade their subordinates that they are competent leaders, so leaders boast about themselves. This behavior, on the other hand, is not acceptable in Japan or in some other countries. In another study, comparing Taiwanese responses to those of Americans, it was found that the value system of the respondent had an effect on which specific behavior was related more to transformational or transactional leadership. Also, in a recent multicultural study, it was found that the behaviors contributing to each dimension of transactional and transformational leadership styles varied

when the respondents were not culturally homogeneous. However, in comparing U.S. managers to those from Bangladesh, the results showed that the five-conflict management styles, as defined in the United States, were also valid for Bangladesh. Therefore, the meanings of the conflict management styles is the same across these cultures.

In regard to universal functionality of leader behavior, we can examine two studies assessing the relationship between leader behavior and organizational outcomes. A five-country (Japan, South Korea, Taiwan, Mexico, and the United States) comparison illustrated that charismatic, supportive, and contingent rewards had a positive impact across cultures on employee work-related attitudes, whereas participative, directive, and contingent punishment in relation to these organizational outcomes varied by country.

In an experimental study in the United States, two confederates acted as either a transformational leader or a transactional leader, and they both led a group of Asian Americans or Caucasians. It was found that the Asian groups were more collectivist while the Caucasian groups were more individualistic; the collectivist groups (Asian) generated more ideas when the leader was a transformational leader than when the leader was transactional. However, the individualistic groups (Caucasian) generated more ideas with a transactional than with a transformational leader. These studies demonstrate that the team members' cultural values and experiences moderate the functional relationship between transactional and transformational leadership and organizational outcomes. Examining the effect of culture on the effect of influence tactics cross-culturally, two cross-cultural studies with managers from Hong Kong, the PRC, Switzerland, and the United States were examined. The findings supported the fact that cultural values affected the effectiveness of managers' choice of influence tactics. The Western managers, as compared to the Chinese managers, rated direct task-oriented influence more effective, and they rated influence tactics that were based on personal relations less effective. More specifically, Schmidt and Yeh quoted the Chinese saying, "Give benefit, but show authority." This quote is a good example of the impact of a society with high power distance and high collectivist cultural values.

The cross-cultural evidence shows that although the notions of the U.S.-driven concepts of leadership are present in all cultures, they do not manifest at the same level or in the same way, or have similar effects on employees. Therefore, it is necessary to recognize that

the cultural context in which a leader is functioning will impact the behavioral choices he or she makes; it also impacts the interpretation of their behavior. Ultimately, culture moderates the effectiveness of these leader behavior styles. On the other hand, the managers from cultures with similar level of socio-economic development seem to be more similar compared to those from societies with various level of development. Thus, it is possible that as the gap between level of education and advancement shrinks, the similarity among leaders across cultures will also increase.

### **3.2. Etic Approach: Contingency Approach to Leadership**

Three main contingency models of leadership are recognized in leadership research: Fielder's contingency model and leadership effectiveness, Mitchell and House's path-goal theory, and Vroom's leadership decision-making theory, which has the most empirical research. Because the path-goal theory of leadership has received mixed support, this article focuses on the role of culture in two models: Fiedler's contingency model of leadership effectiveness and Vroom's normative decision-making theory.

#### **3.2.1. Contingency Model of Leadership Effectiveness**

The contingency model of leadership effectiveness is a leadership model that combines the leader's trait and its match with the situation in predicting leadership effectiveness. The results of past research have found support for the model across various industries and populations. The situation in this model is defined by assessing the leader's amount of control and influence in the situation, referred to as situational control. In high and low situational control conditions, leaders who are more task-motivated than interpersonal are more effective; whereas in moderate situational control, leaders who are more relationship-motivated rather than task-motivated are more effective.

Fielder identified three factors that contribute to situational control. In order of importance, they are the leader-member relationship, the task structure, and the position power. The leader-member relationship refers to the leader's certainty that the team is united and supportive of him or her. Task structure refers to the leader's knowledge of how to get information or conduct a task. Position power refers to the

leader having a title and the legitimacy and resources to reward and punish team members.

The trait of the leader is assessed with the least preferred coworker (LPC) scale. This construct assesses the leader's motivational orientation, that is, whether the leader gives priority to task accomplishment or to people's feelings. High score is indicative of relationship orientation and low score is indicative of task orientation.

The model has been tested in across multiple cultural settings, and in many it has shown to be successful in predicting a leader's effectiveness. This model is sensitive to cultural milieu, as the culture of the situation can be assessed by the situational control factor in the model. Thus, the model has not only gained cross-cultural support, it has also demonstrated that it can incorporate culture as part of its assessment of the situation.

The LPC score, depicting the leader's trait in this model, has a normal distribution in the United States with a mean of 65. The range of leaders' LPC scores varies across cultures. For example, the average score for Mexican managers on this scale was higher than the U.S. samples, and the average for Iranian managers was higher than both the United States and Mexico. On the other hand, the managers from England, Thailand, and the United States were not different on their scores on LPC scale. Also, no gender difference was found when comparing men's and women's LPC scores. Furthermore, the distributions of the score vary across cultures. For example, almost all respondents from Navaho nation (a Native American tribe) scored so high on this scale that it was not possible to identify any individual with low scores (task oriented).

The model, therefore, allows for the assessment of the interference of culture on leadership effectiveness, and it has shown varying degrees of cross-cultural validity. However, due to varying ranges and distribution of the LPC scores, it could be concluded that across the respondents from the United States, Mexico, Iran, and Navaho nation, the U.S. managers tend to have the prepotency to give higher priority to task accomplishment than to their relationships compared to the managers from other cultures. Nevertheless, the premise that task-motivated leaders do better when they feel in control, and relationship leaders do better when engaging team members, seems to be valid across cultures.

#### **3.2.2. Normative Model of Decision Making**

Based on group dynamic and decision-making principles, this prescriptive model of leadership decision-making

strategy was developed. The model states that leaders typically are confronted with five decision-making strategy options that range from very authoritative, to consultative, to participative. This model assists the leader in assessing the situation and choosing the most effective decision-making strategy. The situation in this model is assessed through answers to questions on the leader's knowledge, the team's receptivity to the decision made, and the time available to make the decision. In brief, when leaders have the information and are not dependent on the subordinates to carry out the decision, they can choose more autocratic decision-making styles and spend less time on this process. However, if the leader either does not have all the facts or is dependent on the subordinates' support, he or she needs to take more time and use more consultative and participative styles of decision making.

The validity of the model was tested on both men and women leaders. It was found that male leaders who were perceived as autocratic were evaluated as moderately positive. In contrast, female leaders who were perceived as autocratic were rated negatively.

Cross-cultural studies using this model showed that the choice and success of decision-making style is dependent on politico-economic conditions. For example, when this model was tested in Austria and Poland, it was found that Polish managers had a harder time following the model's prescription to use participative decision-making styles when necessary. This became more evident as the decision became more important. Furthermore, results showed that only after the fall of communism in Poland did Polish managers show a trend toward more participative practices. Therefore, these two studies demonstrated the impact of politico-economic condition on the leader's decision-making choice. Also, when the ideal leader's decision-making style was compared across the three German-speaking countries of Austria, Germany, and Switzerland, the results showed that the managers in all three countries preferred a team-oriented leader. However, the Germans and Austrians favored a "lone fighter," whereas the managers from Switzerland did not support this style.

Thus, the model is cross-culturally supported. The aforementioned studies demonstrate that the political and economic conditions should be considered as part of the situational factors. Some social situations give leaders more choices in using different decision-making strategies. As can be noticed in the case of women and the Polish managers before the fall of communism, the leader is devalued if he or she does not behave according to the expectations of the majority.

Different definitions of leadership that was developed mostly in North America and was validated in various countries or across men and women in leadership positions have been reviewed, and the influence of culture on leadership has been discussed. In summary, it seems that the validity of a universal characteristic of leaders is questionable. In contrast, contingency or situational leadership models showed more cross-cultural validity. However, culture must be considered as a situational factor, a moderator that influences the relationship between the leaders' traits, behavioral styles, and effectiveness.

### 3.3. Quasi-Emic and Emic Approaches: Implicit Leadership Theory

The previous sections reviewed the results of etic studies, in which the leadership theory, model, or measure was transported from one culture and applied to another to validate it. In these studies, however, implicit leadership theories of people of various cultures have not been the focus of research. It is through the emic approach that cultural uniqueness of leadership may come to light.

To illustrate the importance of subjective meaning of leadership in various cultures, the word that describes the activity in the workplace in different languages can be compared. For example, in English, the word "business" is used as a generic term for what goes on in the workplace. This word—denoting the condition of being busy—implies a sense of activity and energy to accomplish goals. In Spanish, however, the term used to refer to the same exchange is *negocio*. Etymologically, the root of this word is similar to negotiation, and it brings to mind exchange, social interaction, and give and take. Therefore, the comparisons of words used in different cultures that refer to activities in the work place may reflect the perspectives and priorities of these cultures. That is, in one culture the main focus may be getting the job done, in another, establishing relationships. Thus, cultural nuances are depicted by slight variations in the priorities on task versus relationships.

In cross-cultural leadership, two sets of emic studies can be identified: first, those that were inspired initially by Western conceptualizations and then arrived at an emic concept, which are referred to as quasi-emic; second, those studies that emerged empirically from within the culture, known as emic.

### 3.3.1. Quasi-Emic Approach

Three main leadership behavior models have emerged from the studies of leadership behavior in Iran, India, and Japan. These models did not support the two-factor model of leader behaviors (consideration/people orientation and instrumental/task orientation) that were present in North America. We refer to these endeavors as quasi-emic, because they were originally based on theories or measures that were inspired by researchers in North America. By examining the findings in each Asian country, similarities and uniqueness can be noticed.

In Iran, Ayman and Chemers reported that the behaviors describing consideration and initiating structure behavior style were both highly correlated with "My leader is like a father to me" and "My leader is effective"; the two styles were inseparable. Comparing the results of the studies in Iran, in which the leader was either American or Iranian, showed that Iranians maintained a perception that the two types of leader behaviors need to occur simultaneously for a leader to be effective and acceptable. Thus, the label "benevolent paternal" was coined for the Iranian leadership.

To illustrate the significant contrast of acceptable leadership behavior in Iran and the United States we can consider this anecdote: Iranians in the United States have commented with amazement that their bosses are "all about business" and are impersonal, which is judged as inappropriate and rude. For example, it is said that first thing in the morning, the bosses walk into the offices, and without greeting employees and asking about their well-being, state their business, make task requests, and leave the room. However, when Iranians share this experience with their American peers, the response they get is that this is an acceptable behavior at work. That is, in the U.S. workplace, it is acceptable to forego exchange of social pleasantries. In Iranian culture, however, the expectation is for the leader to be personally warm and courteous as well as professional.

Examining the application of quasi-emic approach in India, J.B.P. Sinha and colleagues explored leader behaviors. They used the two leader behaviors of consideration and initiating structure, but also added culturally unique leadership behaviors they labeled as "nurturant-task." As a result, a unique conditional behavior that resembled the benevolent paternal leadership style in Iran emerged. In other words, leaders' behaviors included simultaneously expressed demand for getting the job done and caring for the person.

Concurrently, in Japan, Misumi developed a PM (performance/maintenance) theory of leadership behavior.

He took the concept of performance-emphasis and maintenance-emphasis functions from Cartwright and Zanders' conception of leadership. The concepts resembled other leader behaviors popular in the United States, such as consideration and initiation of structure. However, PM is distinct from the leader behavior models in the United States. Misumi's theory emphasized that the two behaviors of attending to the task and to the people are not polar opposites, nor do they function separately. In addition, Misumi believed that leader behavior cannot be generic but should be put in context of the job. Therefore, not only was the PM model similar to the other Asian studies in the inseparableness of leader behaviors but it also mentioned the importance of context.

In summary, there is similarity across the studies in Asia, where the accepted leadership behavior style is a combination of consideration and initiating structure. In these cultures, focusing on only one aspect of work, that is, either getting the job done or attending to people, is not considered appropriate. The social protocols and pleasantries are significant elements of effective leadership. In these societies, business is about building trusting relationships through which the job gets done, not just about getting the job done.

In addition to the above studies, Schein and her colleagues conducted a series of studies around the world examining the image that people have of a manager compared to the image they have of a woman and a man. The results of these studies have shown that across cultures; there is a male image of a manager. This finding is very significant, as the number of women managers is increasing, and they will be faced with expectations that may be contrary to their style and habit. Interestingly, this may be the first universal finding in leadership that was not moderated by culture.

In summary, each culture cultivates a unique set of expectations of their leaders. The universal concept that emerges from these studies seems to underline the fact that leaders need to attend to task and to relationship. Across cultures, the importance of these two styles and their distinctions seem to vary. In Asia, it seems that effective leaders are expected to engage in task- and people-oriented behaviors simultaneously.

### 3.3.2. Emic Approach to Leadership

Earlier, the various types of cross-cultural approaches and the role of the endogenous and exogenous researcher were discussed. The previous section reviewed the results of the findings of studies conducted

by endogenous researcher. Based on their interpretations, an emic understanding of the leadership process in Asia evolved. This section presents two studies, in which, through a series of interviews and observations, two exogenous researchers explored leadership in China and Japan.

The first study's objective was to understand women leaders in China at the early stage of that country's decision to open itself to a capitalistic economy. Through a series of interviews, Korabik (1990) contended that Chinese women managers, compared to men, attributed their success mostly to interpersonal social skills. On the other hand, most of the men managers, as compared to women managers, attributed their success to skill, training, and experience. Interestingly, both Chinese men and women managers said, "It is possible to be both tough and caring. I try to combine patience and firmness." This finding reconfirmed the quasi-emic models of Asian leadership, in which the two task and interpersonal behaviors needed to occur simultaneously.

Pascale, in a 1978 case study of Japanese managers' style of leadership, presented comments from Japanese managers about how American and Japanese managers differ in their style. As an example, he quoted a Japanese manager as saying, "You Americans are fond of announcing things. It sets everything astir." Pascale described the Western leaders as those who are "out standing," whereas Eastern culture values leaders who stand "in" rather than "out." The general but important difference between Japanese and American managers is that Americans are more explicit in their communication and the Japanese more implicit.

The emic studies of leadership, therefore, portray cultural subtleties in the leadership process. For example, across the world, leaders are expected to act more masculine. Also, if one were to use a continuum in which one end represented attention to performance and the other interpersonal relationship, the emic studies have demonstrated that Asian leaders seem to put more priority on a balance of interpersonal relationship and task accomplishment and see them as intertwined, whereas managers in the United States may prioritize task accomplishment at the possible expense of interpersonal relationship. The culture of men compared to women is typically more focused on task activities than emotions and interpersonal relationship. Hence, one could conclude that in the United States, men managers compared to women managers put more priority on performance than on interpersonal relationships. Also, one could surmise that since Asian managers seem to combine the relationship and task behaviors, their

culture would be more similar to that of American women managers than to American men managers.

#### **4. NEW TOPICS IN CULTURE AND LEADERSHIP**

In recent years, three topics have emerged in the study of culture and leadership. One is related to the method of research in the study of cross-cultural leadership and the call for a multicultural team. Another is leadership when cultures collide, that is, when leaders operate in a multicultural environment. The third topic focuses on the evolving role of a transnational leader, that is, the difference between an international leader and a global leader.

##### **4.1. Multicultural Research Teams**

The majority of studies discussed so far have examined either the validity of one leadership paradigm cross-culturally or the unique meaning and function of leadership in a given country. Several recent research teams have explored culture and leadership. The GLOBE project, a recent cross-cultural research endeavor, has systematically examined similarities and differences of respondents over the question of leadership expectations and prototypes in several industries within each country.

In the late 1990s, House, Hanges, and Ruiz-Quitaniella described the GLOBE project as a multi-phase and multi-method research project that included collaborators representing various cultures around the world. They reported that the project presented data from 62 cultures and included 170 indigenous and exogenous social scientists. In their view, the concept of leadership is defined as "the ability of an individual to influence, motivate and enable others to contribute toward the effectiveness and success of the organizations of which they are members." The key research questions of this project focused on identifying both universal and culture-specific leadership attributes, behaviors, and organizational practices. Additionally, it identified in what way culture affected the type of leadership attempted and the effective leadership, and it presented the comparative standing of various cultures on the cultural dimensions. Finally, the project examined the potential for an underlying theory that could explain both national differences and the universal leadership attributes and practices.

To accomplish this, three measures were developed. One focused on the cultural and societal values attached to an effective leader (i.e., what they should be) and how they observed leaders behaving (i.e., what they are). The second measure addressed the practices of their organization. The third measure examined the attributes that impede highly effective leadership. (A detailed representation of the method and sampling of the GLOBE project is beyond the scope of this article.)

This project will eventually allow the examination of similarities and differences across cultures based on agreed-upon definition of leadership. Nonetheless, the preliminary findings attest to the same phenomenon that has been stated throughout this article. That is, although there may be some general universal qualities of leadership behavior, the acts manifesting them will vary and have different meanings across cultures.

#### **4.2. Cultural Collision and Leadership**

Multicultural work settings have been around for centuries. Immigrant workers and cross-national trade have existed around the globe for centuries. However, the intensity of cross-cultural exchanges has a more noticeable economic impact in today's global society than in the past. In previous centuries, those with resources and military might imposed their ways on others. Now, the competitive global economic market does not allow for a company or a nation to mistreat others for long, as there are competitors and potentially better options. Also, the presence of global communication systems has increased the accountability for social responsibility and human rights in business exchanges.

In today's global economy, expatriates can be migrant workers or expatriate managers of large, multinational corporations. Leader-member exchange theory can assist understanding the impact of culture on leadership in a multicultural setting.

The leader-member exchange defines leadership as a relationship. It has demonstrated that leaders have varying levels of intensity in their relationships with different employees. When the leader and the subordinates are in a close, trusting relationship, they enjoy an enriched relationship in which the leader is more apt to engage in transformational behavior styles. When the subordinates are at the periphery of the team, are not closely involved with the leader, and do not have a strong trusting relationship, the leader tends to behave

more transactionally with them. Leader-member exchange theory, therefore, demonstrates that like any relationship, the leadership process also goes through a maturation process, from being strangers to being confidants. What makes this model appealing to multicultural setting is that at the core it assumes that leaders and subordinates develop unique dyad relationships.

In a multicultural setting, it is conceivable that leaders will develop closer relationships with subordinates who are similar to them; the leader may stay distant from those who are different. The research on this model is still young, yet the opportunities to explore the role of culture in leader-subordinate dyad relationship building are promising. Exploring the cultural factors that contribute to building strong leadership relationships in multicultural settings is necessary.

Independent of this theory of leadership, practitioners and cultural anthropologists have developed cultural guides to help expatriate leaders adjust to working in cross-cultural settings and with people from various cultures. However, in this type of cultural awareness training, two issues are overlooked. The first is that the cultural prototypes are used. That is, in this training, the leader is told what typically may occur in a country or what to expect from people of a certain culture. This can create stereotypes, as most people do not fit the cultural prototypes or stereotypes, and so can lead to misunderstanding. Second, this method of cultural awareness assumes that culture is static. Trompenaars and Hampden-Truner stated that international managers are faced with layers of cultural demands that are fluid and need balancing. In addition, in this type of training there is a tendency for the leader to try to act similar to the natives. This can be a no-win situation, as it has been demonstrated that natives do not trust expatriate leaders who act like them. Consequently, the value of knowing about the norms of the other culture is the awareness of the range of accepted behaviors that can equip leaders with alternative explanations when they observe behaviors that may be strange to their own cultural norms.

In summary, our knowledge about leadership in multicultural settings is still evolving. For years, cross-cultural trainers such as Foa and Triandis have recommended that those interacting in cross-cultural settings defer judgment and view situations through the eyes of the member of the other culture. They also developed the cultural assimilator training programs, which teach the individual faced with a heterogeneous team to be aware of his or her cultural assumptions and potential biases, as well as the beliefs

and assumptions of the others. This technique trains the sojourner or expatriate to be more attentive and to think before acting.

Therefore, it is most helpful for the leaders in multicultural settings to be aware of the cultural norms of their own culture and those of the others. This awareness gives them a perspective of the range of behaviors to be expected. In addition, keeping an open mind and asking questions before judging is the best way to adjust to a team composed of various cultures and styles.

### 4.3. Global versus International Leadership

Many scholars use the terms global and international interchangeably. A number of scholars, however, use the terms to refer to two distinct mindsets. International leaders are those who have the tendency to impose their own cultural values and policies on others. On the other hand, global leaders perceive complexity and are inclusive, respecting people of all cultures equally, and are concerned about what is the best solution in the long run. This strategy is not much different than that referred to earlier as the collaborative approach to managing conflict.

Two forces affect the leaders who cross country boundaries or take part in cross-cultural functions. One is that most managers are faced with a universality of the technology and education in various fields. The other is that the international managers are faced with the diverse expectations and norms of the cultures they visit. The global leader has the ability and mindset to create the needed balance and harmony between these two. The international leader, though, will focus on the universal, and may not be skilled in or aware of the important cultural needs and expectations. The global leader incorporates this diversity when establishing his or her business plan and policies.

Global leadership requires awareness and acceptance of the diversity of the needs present in the world. In global leadership, the cultural interests and values of all are included; this process, due to its complexity, requires more time. Global leadership is inclusive and, through a balancing process, brings a sense of unity in diversity. As this is a new phenomenon, its representation is limited and difficult to examine empirically. But it is the type of leadership needed for the global society to progress and prosper. Some have proposed that perhaps a federation, committee, or

council representing global-minded individuals with different cultural background is the best choice for a global leadership.

## 5. SUMMARY AND CONCLUSION

As this article has demonstrated, the study of culture and leadership is very complex. The progress in exploring this symbiotic relationship was explored through various perspectives, and the importance of including culture in understanding a social phenomenon such as leadership in a diverse and culturally heterogeneous setting was well documented. Imposed etic, emic, and derived etic studies were reviewed, and their findings were summarized. Overall, although the various approaches to leadership were validated in different cultural settings, these findings need to be considered with caution.

It seems that most countries, other than those with Anglo-Saxon background, place importance on the interpersonal relationship aspect of leadership. Also, the most powerful aspect of culture in determining leadership values or behaviors seemed to be economic environment such as urbanization. Additionally, although the various definitions of leadership functions were validated across cultures, the specific behavior that is perceived by one culture as contributing to one style is not the same across cultures. This fact alerts the sojourner managers of a multicultural team to be attentive as to how their specific acts may be interpreted differently by different members of their team.

In conclusion, to understand leadership in various cultures or globally, one must review multiple studies and approaches across time. This article provided a glimpse into the complexity present in the interplay between culture and leadership, highlighting the dangers of oversimplification and overgeneralization of the universality of leaders' behaviors. At the same time, some of the paradigms that have evolved so far in leadership, in principle, seem to be working globally. The challenge of a multicultural and global leader is to be attentive to the nuances present in different cultural groups and be accepting of them; in essence, to pursue a balance in unity and diversity.

### See Also the Following Articles

Conformity across Cultures ■ Cross-Cultural Psychology, Overview ■ Industrial/Organizational Psychology across Cultures ■ Power, Authority, and Leadership ■ Training, Cross-Cultural

### Further Reading

- Ayman, R. (2004). Situational and contingency approaches to leadership. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Eds.), *The nature of leadership* (pp. 148–170). Thousand Oaks, CA: Sage.
- Ayman, R. (2002). Contingency model of leadership effectiveness. In L. L. Neider, & C. A. Schriesheim (Eds.), *Leadership* (pp. 197–228). Greenwich, CT: Information Age.
- Ayman, R., Kreicker, N. A., & Masztal, J. J. (1994). Defining global leadership in a business environment. *Consulting Psychology Journal: Practice and Research*, 46, 64–77.
- Chemers, M. M. (1997). *An integrative theory of leadership*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chemers, M. M., & Ayman, R. (Eds.). (1993). *Leadership theory and research: Perspectives and directions*. New York: Academic Press.
- Early, P. C., & Erez, M. (Eds.). (1997). *New perspectives on international industrial/organizational psychology*. San Francisco, CA: The New Lexington Press.
- Gelfand, M. J., Raver, J. L., & Ehrhart, K. H. (2002). Methodological issues in cross-cultural organizational research. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology. Blackwell handbook of research methods in psychology* (pp. 216–246). Malden, MA: Blackwell Publishers.
- Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
- Lewis, D. R. (1999). *When cultures collide: Managing successfully across cultures*. London: Nicholas Brealey.
- McCall, Jr., M. W., & Hollenbeck, G. R. (2002). *Developing global executives*. Boston, MA: Harvard Business School.
- Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1990). *Human behavior in global perspective*. New York: Pergamon.
- Trompenaars, F., & Hampden-Turner, C. (1998). *Riding the waves of cultures: Understanding diversity in global business* (2nd ed.). New York: McGraw Hill.



# Learning

Gary D. Phye

Iowa State University, Ames, Iowa, USA

1. Academic Learning
  2. Learning as Enduring Change
  3. Defining Elements
  4. Types of Personal Knowledge Construction
  5. Declarative Knowledge
  6. The Role of Practice
  7. Procedural Knowledge
  8. Student Self-Efficacy: A Sense of Competency
  9. Strategic Knowledge
  10. Implications for Knowledge Assessment
  11. Developmental and Individual Differences
  12. Conclusion
- Further Reading

## GLOSSARY

**declarative knowledge** Involves knowledge of facts, concepts, vocabulary, and the usual bits of information that are stored in memory.

**formative assessment** The observation or measurement of student learning outcomes prior to and during the learning experience. Formative evaluation provides a starting point for determining change as well as periodic assessment during the learning episode.

**no child left behind (NCLB)** Federal legislation that was enacted to redefine the manner in which federally supported programs are held accountable for the impact of educational interventions on all students in the classroom. Target behavior for defining impact is the improvement of academic achievement in mathematics, reading/literacy, and science.

**pedagogy** The art and science of teaching.

**problem-solving transfer** The ability of students to apply prior knowledge in the form of academic problem-solving strategies to new and novel problem situations in academic subject matter areas.

**procedural knowledge** This type of knowledge is observed in the classroom when students demonstrate the ability (knowing how) to combine, refine, incorporate, and accommodate (in the Piagetian sense) declarative knowledge so that it can be used in a course of action.

**self-efficacy** Personal beliefs concerning one's capabilities to organize and implement actions necessary to perform behaviors at designated levels (efficacy expectations).

**self-regulation** The process whereby students personally activate and sustain behaviors, cognitions, and affects, which are systematically oriented toward the attainment of learning goals.

**strategic knowledge** This type of knowledge is observed in the classroom when students demonstrate "knowing when" to use prior knowledge that is in the form of declarative and procedural knowledge. This type of knowledge is observed in problem-solving situations.

**summative assessment** The assessment that occurs at the end of an educational intervention or teaching episode. A comparison of formative and summative assessment is the basis for the determination of change.

Academic learning is receiving increased attention in the early 21st century. The educational shibboleth is "no child left behind." This federal mandate has impacted the nation in a way that will continue to reverberate for several years. At the classroom level, teaching practices are being impacted by the emphasis on accountability

and impact measures that are being implemented. At the school district level, the accountability and impact data are to be used for policy analysis and to guide educational reform. The professional educator involved with the evaluation of educational practices and programs must produce credible data that meet the criteria of scientifically based evidence. In all cases, the target behavior for these efforts is the improvement of student achievement in reading, mathematics, and science. Assuming that we send our children to school to learn these academic subjects, we must also ask the questions, “what do we mean by academic learning and what are the processes by which students develop their knowledge of academic subjects?” These two questions serve as the focus of this article. Coming from the psychology side of educational psychology, the context is current research findings from the memory, cognition, and learning literature that have potential application to educational practices. Thus, it goes without saying that the perspective is “learner centered” and the unit of analysis includes both the individual and the group, as implied in the phrase “no child left behind.”

## 1. ACADEMIC LEARNING

Learning as a proper noun has so many meanings that it is almost impossible to communicate effectively without first adding a qualifying term. For example, in the applied setting we call the classroom, there are different types of learning theories and models. Some theories describe and explain motor learning that serves as the basis for motor skill development. In other cases, some learning theories focus on learning that serves as the basis for social skills development. However, the focus of this article is academic learning that relies on cognitive learning theories and information processing models to describe and explain the development of personal knowledge about the subject matter we teach in the classroom.

I use the term academic as a descriptive adjective to communicate in one sense something about the nature of the thinking and remembering processes a student must engage in to successfully interact with classroom academic demands. In a second sense, I also draw attention to the fact that the curriculum content we commonly refer to as literature, mathematics, and science has a logically organized structure that must be understood by the student. In other words, in addition to simply acquiring information in the form of vocabulary, facts, or concepts, these building blocks for personal knowledge must be organized in a manner consistent

with the epistemological demands of the subject matter being taught. Examples include knowledge of the syntactical structure of the language in which a third-grade student is writing, the logic for translating and representing a fifth-grade math story problem as an equation to be solved with the use of appropriate solution algorithms, and the formal prepositional logic involved with the ability to use combinations and permutations when solving ninth-grade algebra problems. From these examples, it is readily apparent that academic learning is quite different from social learning or motor learning, which also play a major role in students' efforts to adjust and cope with the demands of schooling.

## 2. LEARNING AS ENDURING CHANGE

Having identified academic learning in terms of a “type” of learning activity, a closer look at a commonly accepted definition of learning will serve as a means of considering what is frequently referred to as the learning process. In the third edition of *Learning Theories: An Educational Perspective*, Dale Schunk offers a suggestion. Although experts disagree about the precise nature of learning, the following is a general definition of learning that is consistent with a cognitive focus and captures the criteria most educational professionals consider central to learning: “Learning is an enduring change in behavior, or in the capacity to behave in a given fashion, which results from practice or other forms of experience” (Shuell, 1986, as cited in Schunk, 2000, p. 2).

There are three elements to Schuell's definition of learning that deserve further attention (change, endurance, and practice) because these elements help define the level of analysis involved in the present use of the term personal knowledge. This level of analysis is also driving the new “no child left behind” (NCLB) legislated mandates. These mandates emphasize the idea that all children can learn, and the key to success according to NCLB mandates is in the assessment and analysis of learning from a scientifically based perspective. Within this context of academic learning our discussion of the construction of personal knowledge can be critically examined.

## 3. DEFINING ELEMENTS

Behavioral change or change in the capacity for behavior is basic to any attempt to distinguish between information

processing and learning. Daily, each of us processes information through contact with our environment as well as information retrieved from memory in the form of ideas, recollections, etc. However, some of this information processing is at a level that does not initiate change in our behavior or our capacity to adapt. Academic learning involves the development of new behaviors or capacities, not simply the replication of previous habits. Furthermore, the demonstration of change is not left to conjecture. Rather, change can only be demonstrated by measuring the behavior at two different points in time. Typically, this is prior to and following instruction.

Enduring change must have some permanency about it. In other words, the important elements of instruction to which a student is exposed should be remembered. In fact, our curriculums in the public schools are designed around this primary assumption. Unfortunately, the empirical evidence for enduring change is scanty at best. In our discussion of personal knowledge, we focus on the individual learner as the unit of analysis, not the class as a whole. This is in fact the reason that the NCLB mandates are framed within a consideration of the individual student (no child left behind). This in no way ignores the fact that academic learning occurs in a group context. Rather, it focuses attention on both the individual and the collective we call a classroom. This makes common sense when one considers that the basis for knowledge permanency is the individual student's memory system.

Practice is an interesting concept in the academic community. A basketball coach teaching motor skills will tell you that "practice makes perfect," a music teacher will insist that students practice between lessons, and the school administration will stress the practice of civility and social behavior. However, the classroom teacher responsible for academic content frequently views practice as a "hit-or-miss" proposition that gets attention during seatwork or homework. Interestingly, although all these practice episodes can be observed in any school, the episode that typically lacks any organization or rationale is seatwork or homework.

Together, these three basic elements of our definition of academic learning stress the idea that academic learning is not a one-shot instructional experience. It is a process of acquiring new information, the refinement and organization of what is already known, and the successful use of that knowledge. Academic learning is the product of practice that provides the basis for relatively long-term change in one's personal knowledge. Ideally, this change increases students' ability to be successful as they move from grade to grade where

the curriculum requires more complex and specialized forms of personal knowledge.

#### 4. TYPES OF PERSONAL KNOWLEDGE CONSTRUCTION

The following discussion is based on a common theoretical assumption that pervades educational psychology: Cognitive change must occur as a student progresses through the curriculum in order to be successful. Jerome Bruner, in his discussion of the spiral curriculum, made this point. Still valid today, this assumption is frequently presented in the form of a hierarchical schema, from simple to the complex, in terms of both the cognitive processing demands made on students and the nature of the subject matter to be learned. The following discussion takes this form and reflects a prevailing attitude among educational psychologists that the construction of personal knowledge can be identified as falling within three major typologies: declarative knowledge (concepts and facts), procedural knowledge (strategies, procedures, and schemata), and strategic knowledge (expertise in academic problem solving).

The study of knowledge can take many forms. Within the context of academic learning, the following analysis is functional in nature and involves an information processing perspective. The functional nature of the following discussion involves addressing the "what" and "how" questions of personal knowledge construction. Phye (1997a) stated,

A particular value of the information processing perspective for the classroom teacher is the assumption that, while each student in the classroom will have a unique set of prior experiences that influences entry learning level, rate of learning, style of learning, and so forth, all nondisabled students in the classroom will be approaching the learning assignment with essentially the same set of processing components. Consequently, when learning is not successful, an assessment can be quickly conducted in order to determine cognitive processing strengths and weaknesses. (p. 53)

#### 5. DECLARATIVE KNOWLEDGE

Basically, declarative knowledge involves those things teachers refer to as facts, vocabulary, concepts, etc. These bits and pieces of information are the building blocks for academic learning. These are the "basics" that

cut across all curriculum content areas. For the student, the construction process involves not only attending and remembering but also the bringing of meaning to this newly acquired information. By connecting new information with prior knowledge or using new information to refine what is already known, meaning is changing from the concrete to the complex.

Three examples make the point. Outside the classroom as well as in the primary grades, a commonly accepted mathematical fact is that  $2 \times 2 = 4$ . However, as one progresses through the mathematics curriculum, a part of any student's academic learning involves the following refinement of prior knowledge. The  $2 \times 2 = 4$  mathematical fact applies only in a base 10 mathematical system. Thus,  $2 = 2 = 4$  may not be a fact in other mathematical systems. This mathematical fact is refined or deemed to be "true" only within a specific context. For most students, this refined knowledge would have little or no relevance outside the classroom.

The second example is a common classroom learning experience that is observed when dealing with concepts in science. During the early grades, the concept of "animal" is introduced and students typically use the term to mean a large group of "nonpeople living things" of one type or another. However, as science lessons progress, the student is challenged to refine this general concept into a more specific organizational system that includes people within the subcategory of mammals and that mammals are still animals. This bit of declarative knowledge refinement has both high academic relevance and relevance outside the classroom.

Having touched on examples from mathematics and science, the last example comes from early reading. As a child progresses from kindergarten to the fourth grade, vocabulary building, fluency in reading, phonetic awareness, word attack skills, and comprehension are reading skills that are practiced in oral reading groups. As these basic reading skills become integrated, a child's reading performance improves. A primary goal of learning to read involves the ability to comprehend what is being read. One part of this process is taking individual words and their meanings and putting them together into sentences. Interestingly, in English many words have different meanings depending on whether they are used as a noun or a verb in a sentence. The word "desert" used as a noun means a dry, barren, often sandy region that can naturally support little or no vegetation. Desert used as a verb means to forsake or leave, especially when most needed. Obviously, the key to understanding the meaning of the word desert depends on having both

meanings available and selecting the appropriate meaning based on the context (rest of the sentence).

Students process information every waking moment of their lives. Some of the information is organized and some is fragmented or organized around real-world survival skills. However, academic knowledge is a special case in which the organizational principles are unique to the subject matter being taught. Consequently, remembering becomes a valuable cognitive skill for any student. Thus, as we go back to our basic definition of academic learning, the emphasis on enduring change becomes more apparent. As learners, we remember facts and concepts not only so we can demonstrate the acquisition of new declarative knowledge but also so this new knowledge becomes prior knowledge to be added to and refined at a later point in time. The reading example described previously also demonstrated the vital role that memory plays in academic learning. The student who encounters the word desert must remember two different meanings and successfully retrieve from memory the appropriate meaning, depending on the cues provided by the rest of the sentence. The ability to do this fluently requires a great deal of practice so that the process of remembering meanings occurs almost automatically.

Reading fluency is a good example of an academic behavior that improves with practice and is a critical academic survival skill. By the intermediate grades, students are asked to take increasingly more responsibility for their own learning. Most classroom environments change at the middle school level and students are expected to acquire a great deal of their own declarative knowledge by reading texts, working through Web-based instructional modules, etc. This expectation on the part of educators suggests that students should be practicing their reading skills during the primary grades so that they can successfully adapt to the demands of the middle school.

## 6. THE ROLE OF PRACTICE

Although practice is an essential element of the learning process, a frequently asked question is, "Does practice make perfect?" It goes without saying that practice of some type is typically more beneficial than no practice. However, not all practice is equally effective. This statement is simply the recognition that practice plays an important role in two quite different ways. In terms of acquiring academic knowledge, repeated exposures or repetition provide the basis for memory storage that is critical for remembering. In terms of reading

vocabulary, the association of the word desert with its meaning and the remembering of this pairing will require several repetitions in order for the information to be stored in long-term memory. This may be accomplished with a simple practice strategy that involves developing a set of what were called “flash cards” when I was a youngster. However, to successfully retrieve and use the appropriate meaning across a number of different reading passages requires the successful application of prior knowledge. The ability to successfully apply one’s knowledge is really a disposition or personality characteristic. This disposition is also acquired through practice and has been identified by Albert Bandura as a sense of “self-efficacy” or competence. In other words, successful practice strategies involve both multiple exposures to the task and actively seeking potential applications of prior knowledge. This is a simple example of “training for transfer.”

## 7. PROCEDURAL KNOWLEDGE

Within the context of academic learning, procedural knowledge is demonstrated when a student can combine, organize, refine, or accommodate declarative knowledge so that a course of action can be taken. This is “knowing how” to use declarative knowledge (what one knows). Procedural knowledge in the classroom is what we commonly strive to teach and we ask our students to learn.

In teaching circles, constructivist or student-centered approaches to instruction place a great deal of emphasis on promoting procedural knowledge. This is frequently the case in the areas of science education and mathematics education. The caveat is that students must have a solid background of declarative knowledge in order to successfully move to a more complex level of cognitive processing. The cognitive processing that serves as the basis for procedural knowledge is domain specific. For example, if we think of algebra as a part of the mathematics domain, there are basic mathematical facts and concepts (declarative knowledge) that have been learned in earlier grades that serves as the foundation for success in algebra. However, in algebra we are also asked to take our prior knowledge about mathematics and use that knowledge to develop new rules, concepts, strategies, algorithms, etc. This is really a case of taking what we know and learning how to use that knowledge to develop new and more complex ways of thinking.

For example, by the time students reach middle school, where they usually take algebra, the teaching

of reading shifts from a focus on teaching basic reading skills to using reading skills to promote self-directed learning in literacy. At this point in students’ academic careers, emphasis is on procedural knowledge involving “knowing how to read” in order to successfully perform other educational tasks (write a poem, give a persuasive speech, read a science or mathematics textbook, etc.). These complex academic tasks are built on the declarative knowledge any student remembers from the earlier grades. This is why remembering (memory storage and retrieval) is the key to academic learning because complex subject matter that students are asked to learn is always built of declarative knowledge and simpler forms of procedural knowledge.

These descriptions of our expectations for students as they progress through the primary and middle school grades again draw our attention to the key academic learning element of enduring change that is couched in practice (seatwork, homework, study hall, etc.). Interestingly, there is little evidence in the research literature that teachers systematically teach study skills. This means that teachers should not only present information in the classroom but also teach students how to use the information being conveyed. Thus, teachers must teach not only declarative knowledge (what) but also procedural knowledge (how). A position frequently taken by educational psychologists is that domain-specific procedural knowledge strategies must be a part of any teacher’s instructional repertoire.

## 8. STUDENT SELF-EFFICACY: A SENSE OF COMPETENCY

Self-regulation or self-regulated learning emphasizes the motivational side of human behavior. As such, it is a complementary view of the information processing approach to academic learning. Metaphorically, cognitive theory and motivational theory are two sides of the “coin” called academic learning. Research efforts by educational psychologists studying self-regulation emphasize how learners direct their thoughts, feelings, and actions toward the attainment of their academic goals.

Bandura’s concept of self-efficacy was discussed previously within the context of study or practice. Generally speaking, self-efficacy is the exercise of control by the individual in the daily task of adapting to a complex environment. Self-regulation and the development of a sense of self-efficacy grow out of a complex interaction of behaviors, interpersonal factors,

and environmental demands. In the classroom, a simplified example involves teacher variables, curriculum demands, and students' sense of success with the curriculum being taught (math, reading, etc.). This sense of success translates on a personal level for the student as being competent in the subject being taught. Furthermore, the success breeds a sense of self-efficacy: "I can do this, I understand how to take what I already know and use it to learn more about this because I understand what it takes to get better." "I am a good reader." "I am good at math."

Self-efficacy was introduced earlier within the context of practice because of the position taken by Bandura that it is the demonstration of being competent (a good reader or good at math) that precedes the sense of self-efficacy. Over the long term (from kindergarten to grade 12), it is the sense of self-efficacy in academic subject matter areas that provides the motivational basis for successfully making the transition between procedural and strategic knowledge construction by a student.

This sense of self-efficacy is what sustains a student when academic problem solving becomes complex and difficult. When confronted with a complex academic problem, a successful problem solver has sufficient declarative knowledge to bring to the task; sufficient procedural knowledge to synthesize, integrate, and evaluate the declarative knowledge; and can strategically (knowing when) translate this into a successful solution for a specific problem. It is self-efficacy, the sense of being able to exercise personal control over the process of knowledge construction in an academic setting, that characterizes strategic knowledge.

## 9. STRATEGIC KNOWLEDGE

Strategic knowledge is viewed as an extension of declarative and procedural knowledge. Whereas declarative knowledge involves learning facts, concepts, etc. and procedural knowledge involves learning how to use declarative knowledge to construct domain-specific rules, algorithms, etc., strategic knowledge involves knowing when to use prior knowledge. In other words, when confronted with a novel or problem-solving situation, emphasis is on the retrieval of prior declarative and procedural knowledge that has been stored from long-term memory. When this type of information processing is successful, we observe academic learning behaviors such as problem solving in mathematics, critical thinking in science, or creative thinking in the writing of fiction. This last type of knowledge demonstrated

as a learning outcome emphasizes the active construction of a learning outcome and places the student at the center of the teaching–learning process. In addition to reflecting a competency dimension, primary consideration is given to motivational aspects of academic learning. This involves helping students develop a disposition frequently referred to as intellectual curiosity or academic expertise.

The development of strategic knowledge may be based largely on student characteristics that are very difficult to teach. Furthermore, some students may simply have difficulty with this level of cognitive functioning because of their cognitive limitations or personality characteristics.

Regardless, a teacher should provide opportunities that foster the development of strategic knowledge for those students who are capable. Strategic knowledge is a primary learner outcome of a self-regulated learning process we frequently identify as academic problem solving. Academic problem solving is observed when students can take primary responsibility for demonstrating the learning outcomes of problem identification, problem representation, strategy construction, and solution verification (evaluation).

Emphasis is on the student taking responsibility for identifying a problem, representing the problem, developing a strategy for solving the problem, and then evaluating the success of the strategy employed. This academic problem-solving perspective has been a major contribution to the teaching literature that is based to a large extent on the research efforts of Richard Mayer. These stages of the problem-solving process can be taught in any of the previously identified academic areas (mathematics, science, and literature). Any self-regulated learning process places a premium on remembering—the use of prior knowledge. The student is responsible for knowing what, how, and when. As demonstrated by Barry Zimmerman and colleagues, if we are successful, some of our students will become self-regulated learners. The choice is theirs, and the reward will be ours.

## 10. IMPLICATIONS FOR KNOWLEDGE ASSESSMENT

The implications for pedagogical practice are again derived from our definition of academic learning. Consistent with the NCLB legislation, the focus is on every child. This calls for the recognition of individual

differences in the classroom. In a single classroom on any given day, some students will be dealing with the development of declarative knowledge in a subject area while others will be grappling with issues involving procedural knowledge. A few may even be working at the strategic knowledge level of construction. On a daily basis, this means that classroom teachers will have to be well grounded in both the subject matter being taught and effective teaching strategies. In addition, since change is the focus of our definition of academic learning, this means that when appropriate, teachers are responsible for guiding students from one level of knowledge construction to another. This is a long-term view and involves change over the academic year within a grade. Consequently, for future change to occur, the student must remember today's lesson, which is tomorrow's prior knowledge. The implications for practice are obvious since the effective use of memory storage and retrieval is not easily accomplished.

Formative and summative assessment of academic learning is required since enduring change can be verified only to the extent that it can be measured. As mentioned previously, measurement at a minimum of two points in time is required for the assessment of change. Furthermore, a classroom environment in which student progress is monitored effectively would require multiple assessment techniques. The major obstacle facing most teachers is the assessment issue. There are numerous sources available for teachers that provide the basis for developing assessment activities that measure declarative knowledge. Most of these activities involve the assessment of memory retention using standard teacher-made tests.

The assessment of procedural knowledge and strategic knowledge also requires the assessment of memory retention, but it must also include the assessment of a student's ability to use the knowledge in a manner that reflects the transfer of higher order thinking skills (e.g., Bloom's taxonomy). This means that the assessment will require evidence of strategy transfer or problem-solving

transfer on the part of students. Thus, as we promote the development of procedural and strategic knowledge through the use of instructional practices that foster such thinking, we must not forget that assessment practices must align with the instructional efforts and the learning expectations we have for our students. In simple terms, there are two assessment questions we must ask ourselves as teachers: Do our students remember what we taught, and can they effectively use what they remember?

## 11. DEVELOPMENTAL AND INDIVIDUAL DIFFERENCES

A model for knowledge assessment is provided in Table I. As mentioned previously, declarative knowledge can be assessed in the traditional manner and is frequently a one-time occurrence. In contrast, procedural knowledge and strategic knowledge require both a measure of memory retention and a measure of transfer. This dual-assessment procedure is required because transfer is based on prior knowledge. If we simply measure transfer performance and transfer performance is poor or nonexistent, as teachers we cannot determine if our students fail because they cannot remember or because they cannot use their declarative knowledge. In other words, memory retention is a necessary but not sufficient condition for transfer. Thus, teaching for transfer is based on teaching for declarative knowledge but also involves helping students understand how and when to use their knowledge.

Having made a distinction between an assessment procedure for declarative knowledge and procedural or strategic knowledge in terms of single or dual assessment, a distinction is also to be made between the assessment of procedural and strategic knowledge. In contrast to procedural knowledge, in the discussion of strategic knowledge, the volitional nature of strategic knowledge was emphasized. During procedural

TABLE I  
Model for Functional Knowledge Assessment

<i>Type of learning outcome</i>	<i>Memory retention performance</i>	<i>Transfer performance</i>	<i>Kind of knowledge</i>
Nonlearner	Poor	Poor	None
Non-problem solver	Good	Poor	Declarative
Guided problem solver	Good	Good	Procedural
Self-directed problem solver	Good	Good	Strategic

knowledge assessment, if students can transfer procedural knowledge “when asked to do so,” guided understanding is demonstrated. From a problem-solving transfer perspective, these students are demonstrating the last three stages (problem representation, strategy selection, and solution verification) of the aforementioned problem-solving process. Unfortunately, we lack information about problem identification. The assessment of strategic knowledge, where no teacher guidance is provided, provides information about students’ ability to identify problems. Here, students must volitionally construct an awareness of what knowledge is appropriate, how that knowledge can be organized into a solution strategy, and when the strategies are to be employed. Such students not only demonstrate the learning outcome of understanding but also the characteristics of successful problem solvers. This distinction between the assessment of procedural and strategic knowledge must always be considered when developing performance assessments in the classroom.

## 12. CONCLUSION

The multiple assessment approach to promoting understanding and problem solving in the classroom is at the heart of the NCLB legislation. No child should be left behind as a nonlearner. Every child can learn to a level that is determined by the individual differences in the academic abilities each of us exhibit. The hope is that as teachers, we facilitate to the best of our ability

the personal knowledge construction efforts of all our students.

### See Also the Following Articles

Educational Achievement and Culture ■ Gender and Education ■ Intelligence in Humans ■ Learning Disabilities ■ Learning Styles and Approaches to Studying ■ Transfer of Learning

### Further Reading

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bruner, J. S. (1966). *Toward a theory of instruction*. New York: Norton.
- Haskell, R. E. (2000). *Transfer of learning: Cognition, instruction, and reasoning*. San Diego: Academic Press.
- Mayer, R. E. (2002). *The promise of educational psychology. Vol. 2: Teaching for meaningful learning*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Phye, G. D. (Ed.). (1997a). *Handbook of academic learning: Construction of knowledge*. San Diego: Academic Press.
- Phye, G. D. (Ed.). (1977b). *Handbook of classroom assessment: Learning, adjustment, and achievement*. San Diego: Academic Press.
- Schunk, D. H. (2000). *Learning theories: An educational perspective* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research*, 56, 411–436.
- Zimmerman, B. J. (1990). Self-regulated academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, 2, 173–201.



# Learning Disabilities

William D. Tilly III

Heartland Area Education Agency 11, Johnston, Iowa, USA

1. History of Learning Disabilities
  2. Definition
  3. Treatments for Learning Disabilities
- Further Reading

programs and their “response to the intervention” is observed and used in decision making.

## GLOSSARY

**formative evaluation** The process of using data collected frequently and repeatedly over time to determine the effectiveness of a treatment program.

**learning disability** As used in this article, a heterogeneous group of disabilities that result in deficits in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself as an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculation.

**modality instruction** The practice of assessing certain characteristics of the learner (i.e., modalities) such as whether the individual is a visual, auditory, or kinesthetic learner, and then matching treatments to the individual's strengths.

**perceptual-motor training** The most frequent form of “process training” in special education, . it assumes that children with learning disabilities, or at least some of them, have problems in sensory integration to which some or all of their learning problems can be attributed. The purpose of the training is to remediate these deficits.

**response to intervention (RTI)** An approach to assessment and treatment of persons with learning problems that can, in many cases, assist in identifying individuals with learning disabilities. In this method, students with specific learning problems are placed in a validated intervention

This article describes the development of our understanding of learning disabilities in the United States. It begins with a brief history of learning disabilities, then the struggles for a universal definition are described, along with the federal legal definition and the emerging consensus on definition within the field of learning disabilities. Finally, a sampling of both proven effective and proven less-effective treatments for learning disabilities is presented.

## 1. HISTORY OF LEARNING DISABILITIES

The existence of learning disabilities has been documented for approximately 200 years. The origins of what are currently called learning disabilities began with early attempts to match functioning within certain areas of the brain to human behavior. Early physicians noticed that patients with certain types of brain injury experienced specific types of behavioral problems. Most relevant to learning disabilities were patients with injuries to the left hemispheres of their brains, who in many cases experienced problems such as slow, laborious speech and the inability to name objects or persons. In the late 1800s, a more specific set of cases was identified in which persons' ability to speak and write remained

intact, but their ability to comprehend spoken or written words was impaired. These conditions were termed word deafness and word blindness.

From these studies, during the early 20th century, physicians began to notice children and adults who experienced what was at the time considered to be defects in their ability to recognize words and letters, despite no obvious brain damage or injury. The inference was made that perhaps these individuals suffered from some sort of neurological disorder that mimicked the effects of people with brain damage. Moreover, a number of cases appeared to run in families, providing further evidence for a possible physical cause. The medical diagnosis for this condition was congenital word blindness.

Beginning around 1920, and building on this early work in medicine and neurology, a group of American pioneers including Samuel Orton, Grace Fernald, Marion Monroe, and Samuel Kirk began landmark work on understanding reading disabilities. While still clearly linked to foundations in medicine, reading disabilities were attacked by these researchers and clinicians as much from an educational perspective as from a clinical one. Indeed, some of these pioneer's names have become synonymous with specific approaches to educational treatment for students with learning disabilities. For example, the Orton-Gillingham approach to treatment is a phonics-based, multisensory approach to instruction that uses the kinesthetic, auditory, and visual modalities for teaching decoding and spelling.

Parallel to developments in educational understanding of students with these disorders, additional work was being pursued between about 1920 and 1960 on the perceptual, perceptual-motor, and attentional abilities that often accompanied damage, injury, or dysfunction in persons' brains. Physicians and psychologists such as Kurt Goldstein, Heinz Werner, Newell Kephart, and William Cruikshank conducted many research studies documenting deficits such as hyperactivity, perseveration, figure-ground confusion, and distractibility. Again, it was reasoned that if persons with documented damage to their brains exhibited perceptual and attentional problems, persons with presumed damage to these brain areas might be expected to have these problems as well.

It was between approximately 1960 and 1975 when the field of learning disabilities began emerging as a discipline unto itself. On April 6, 1963, another pioneer in the field of learning disabilities, Samuel Kirk, told a parent advocacy group for children with "perceptual handicaps" that "Recently, I have used the term 'learning disability' to describe a group of children who have disorders in development, in language,

speech, reading and associated communication skills needed for social interaction. In this group, I do not include children who have sensory handicaps such as blindness or deafness, because we have methods of managing and training the deaf and blind. I also exclude from this group children who have generalized mental retardation" (Kirk, 1975, p. 9). Most educators attribute coinage of the term "learning disability" to Kirk. Moreover, some of the parents who heard Kirk's speech in 1963 established the Association for Children with Learning Disabilities, which has now evolved into the Learning Disabilities Association of America. This organization is widely known as one of the most influential learning disabilities advocacy organization in the United States.

## **2. DEFINITION**

### **2.1. Historical Definitions**

One of the biggest challenges in the field has been coming up with a definition of learning disabilities that a large majority of people can agree with. Part of the challenge has been the heterogeneous nature of problems associated with learning disabilities. That is, persons with learning disabilities do not all have the same problems, and the types of problems experienced are very broad. While most persons with learning disabilities have problems with reading and the use of language (roughly 80%), not all do. Some have problems with memory, others do not. Still others have perceptual difficulties. And the list goes on.

In the years between 1960 and about 1975, there was considerable disorganization and disagreement on the specific nature of what constitutes a learning disability. Though Sam Kirk and his student Barbara Bateman had offered a series of definitions in the early and middle 1960s, definitions created by different advocacy and professional groups continued to stress different components of the disability. It was at this time that the federal government began taking interest in learning disabilities and began working toward definitions of its own.

### **2.2. United States Legal Definition**

A series of federal task forces on learning disabilities and legislative actions occurred throughout the 1960s and into the early 1970s. These set the stage for the federal government's enactment of Public Law 94-142, the Education of the Handicapped Act, in 1975 (EHA). Coming on the heels

of the civil rights movement, EHA provided a mechanism for the federal government to provide funding to states to support the provision of special education services to students with disabilities. Based on the law, children with disabilities could not be excluded from receiving an education solely on the basis of their disability. Major components of the law included ensuring students with disabilities a right to a free, appropriate public education, heightened access to due process and procedural safeguards to ensure appropriate provision of services, and provision of an Individual Education Program that was reasonably calculated to provide benefit to students with disabilities (in this case, learning disabilities).

EHA was an impressive and ambitious step by the federal government. At the time, however, the promise of EHA could not be realized unless an acceptable definition of learning disabilities could be written. Recall that at the time, there was considerable professional disagreement on the nature of learning disabilities. Congress and the U.S. Office of Education considered much testimony, research, and advocacy in selecting the definition of learning disabilities that finally made its way into federal law. Indeed, many components of historical definitions were considered, and the final definition of learning disabilities was finally written. According to federal law, a learning disability is as follows:

**General.** The term learning disability refers to a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself as an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculation, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Disorders not included: The term does not include learning problems that are primarily the result of visual, hearing, or motor disabilities or mental retardation, of emotional disturbance, or of environmental cultural or economic disadvantage [U.S.C. 300.7(C)(10)].

In addition to the definition provided in the U.S. federal law, the U.S. Department of Education provided regulations that further direct states and territories in determining the presence of specific learning disabilities. As stated in the regulations:

- a. A team may determine that a child has a specific learning disability if:
  1. The child does not achieve commensurate with his or her age and ability levels in one or more of the areas listed in paragraph a.2 of this section, if provided with learning experience appropriate for the child's age and ability levels; and

2. The team finds that a child has a severe discrepancy between achievement and intellectual ability in one or more of the following areas:
  - i. Oral expression
  - ii. Listening comprehension
  - iii. Written expression
  - iv. Basic reading skill
  - v. Reading comprehension
  - vi. Mathematics calculation
  - vii. Mathematics reasoning
- b. The team may not identify a child as having a specific learning disability if the severe discrepancy between ability and achievement is primarily the result of:
  1. A visual, hearing, or motor impairment
  2. Mental retardation
  3. Emotional disturbance
  4. Environmental, cultural, or economic disadvantage

Based on the language in the law and the language in the regulations, each of the 50 states and U.S. territories needed to establish specific procedures that would be used to identify individuals with specific learning disabilities.

### 2.3. Converging Professional Perspectives on Definition

The federal definition of learning disabilities set the stage for greater stability within the learning disabilities professional community and general movement toward agreement on a definition. One reason for this progress between about 1975 and 1985 was the considerable research conducted across the country that was funded by the U.S. Office of Education. Five learning disability research institutes were funded at major research universities across the country, each focusing on a different component related to understanding and treating individuals with specific learning disabilities in the schools. The institutes included the following:

- The Columbia University Institute, directed by Dale Bryant, focused primarily on information-processing problems of students with disabilities. Within this program, memory skills, arithmetic, study skills, basic reading, and reading comprehension (among other areas) were studied.
- The University of Illinois at Chicago Institute, directed by Tanis Bryan, focused primarily on the social competence of children with learning disabilities and their attributions of success or failure.
- The University of Kansas Institute, directed by Donald Deshler, focused primarily on school-based

interventions for secondary (adolescent) students with learning disabilities.

- The University of Minnesota Institute, directed by James Ysseldyke, focused their energies on the decision-making process related to identification of students with learning disabilities as well as methods for monitoring academic progress for students with learning disabilities. It was from the Minnesota Institute that curriculum-based assessment and curriculum-based measurement technologies were developed initially.

- The University of Virginia Institute, directed by Daniel Hallahan, focused their research efforts on students with learning disabilities who also had attentional problems. Additionally, the Virginia site also conducted research on the use of academic strategy training to improve the performance of students with learning disabilities.

During the 10 years following the work of the learning disabilities institutes, a series of new definitions of learning disabilities was offered by professional organizations and federal task forces. These definitions, though having some different nuances, all describe learning disabilities as (1) intrinsic to the individual, of presumed neurological dysfunction, (2) related to the use of language such as reading, writing, and reasoning, and (3) able to impact the individual in other areas of their lives such as social skills, mathematics, and attention.

Throughout the years between about 1985 and 2000, much additional research was conducted on the nature of learning disabilities. While there are many significant and important findings in this literature base, it appears that there are two large strands of research emerging related to the definition of learning disabilities. The first major area is in phonological processing. In 2000, a group of distinguished researchers (called the National Reading Panel) came together to examine and synthesize what is known about the acquisition of reading. In their report, these researchers synthesized the major skills that a person must have to be able to read proficiently. These include the following:

*Phonemic awareness:* The ability to hear and manipulate the sounds in spoken words and the understanding that spoken words and syllables are made up of sequences of speech sounds.

*Phonics:* Includes both the understanding that words are composed of letters that represent sounds and the ability to use letter-sound correspondence to pronounce an unknown printed word.

*Vocabulary:* Includes both expressive vocabulary, which is the ability of a speaker or writer to produce a

specific label for a particular meaning, and receptive vocabulary, which requires a reader to associate a specific meaning with a given word (either read or spoken).

*Fluency:* The ability to read words with no noticeable cognitive or mental effort. It is having mastered word recognition skills to the point of overlearning.

*Comprehension:* This has been defined many ways by many authors. Comprehension includes active and intentional thinking in which meaning is constructed through interactions between the text and the reader.

The importance of these five “big ideas” in reading is that we now have both a roadmap for directly measuring the subcomponents of reading and specific strategies, linked with each of these ideas, that can be used to directly teach the skill. As noted before, a majority of students with learning disabilities have problems with reading, and this new knowledge provides us both a very specific way to identify specific skill deficits and a sensitive way to identify students with problems. As research progresses, we can expect this strand of research to provide increasingly effective treatments for students with learning disabilities.

The second major strand of learning disabilities research has to do with the identification of a biological basis for learning disabilities. Since the early years of learning disabilities identification, there has been a presumption of a neurological basis for these disabilities. Not until the late 1980s and 1990s has our technology caught up with our intuition and allowed us to study and document these findings. There are a number of different converging sets of data at this point. We know now from postmortem studies that the brains of persons with reading disabilities differ from those of persons without reading disabilities. We know from neuroimaging studies that there are differences in critical brain functions between proficient readers and nonproficient readers. We know also through systematic evaluation of heredity that about 40% of first-degree relatives of students with reading disabilities, speech and language disabilities, and spelling disorders also experience these problems.

## 2.4. Proposed Future Definitions

While all of these new research studies and findings were being created, the federal legal definition of learning disabilities (1975) has remained virtually unchanged. Our new research base, coupled with the practical and political realities associated with learning disabilities, has created a context for reexamination of the federal definition. A major political and practical

challenge is that since the inception of the Education of the Handicapped Act, both the number and percentage of students identified with learning disabilities in U.S. schools has consistently and steadily risen. This rise creates both fiscal and practical challenges. Currently, approximately 50% of all children with educational disabilities served under the IDEA are students with learning disabilities. There are many reasons for this rise, including that (1) the assessment and identification process in schools has been in many cases unreliable, (2) the use of technically inadequate tests is widespread, and often little functional information about how to teach a child with a learning disability is provided from the comprehensive evaluations required by federal law, and (3) different states use very different practices and criteria for determining the presence of learning disabilities. Moreover, some experts contend that learning disability numbers are increasing due to an increasing number of students who are what has been called curriculum casualties. That is, they have not learned because either the curriculum is not geared to meeting their learning needs or their teachers are not skilled in providing instruction in reading, language arts, and/or mathematics. When curriculum and instruction fail, special education is there to fill the gap. In short, the identification and special education service delivery process worked somewhat well from an administrative perspective of allocating resources, but it did not work well from the standpoint of remediating critical skill deficits nor in creating a consistent identification system for learning disabilities.

Added to our research base and the practical challenges we have experienced, additional evidence has been added in very recent years to our understanding of learning disabilities. Headed up by Reid Lyon, Chief of the Child Development and Behavior Branch of the National Institute of Child Health and Human Development of the National Institutes of Health, and a national team of researchers from many major universities, critical insights and definitive studies have begun to emerge that are causing researchers, policy-makers, and practitioners to reconsider the way we identify and serve individuals with learning disabilities in our schools. Perhaps the most critical addition to our common knowledge base is that the prevailing practices used in our schools to identify and serve students with learning disabilities are problematic from the standpoint of improved outcomes for students. Our emerging knowledge is compelling.

1. We know that most students with learning disabilities have important problems with the use of language

and acquisition of reading skills (approximately 80%) and that these problems begin early in life.

2. We know that in American society there are serious, important, and life-long consequences of not learning to read.

3. We have learned that there is a critical period for children to develop language and early reading skills. There are five major component skills that proficient readers master, typically in kindergarten to grade 3. They are phonemic awareness, phonics, vocabulary, fluency, and reading comprehension. All of these skills are directly teachable.

4. We know that the most effective interventions for reading problems are those that begin early and teach students in these critical skills.

5. The learning disabilities identification system in practice in most U.S. schools does not allow the identification of students with learning disabilities until measures of academic achievement can demonstrate a severe discrepancy between measured potential and academic achievement. Typically, this is not possible until second or third grade, past the time when intensive instruction has the best chance of remediating students' early deficits. In short, the system is problematic because we have delayed intervention to a point where it may be less effective and until we are playing catch-up.

As of 2004, the federal government is actively looking at making changes to the definition of learning disabilities in the federal law. A national center for research on learning disabilities was established in 2002 that is a collaborative effort between Vanderbilt University and Kansas University. One task that this center has undertaken is the examination of alternative models for the identification of learning disabilities. An emerging approach to identification of learning disabilities is called response to intervention, or RTI. In short, RTI is an approach in which individuals with learning problems often experienced by students with learning disabilities are placed in high-quality treatment, research-validated intervention programs across relatively short periods of time. The student's response to these interventions is monitored. Students whose performance is both far below age/grade-level peers and who do not respond positively and sufficiently to the validated treatment approaches may be considered for learning disabilities services. Much research is ongoing with regard to RTI approaches to identification of learning disabilities, and it is anticipated that successful implementations of RTI will inform and influence the learning disabilities definition that will be a part of federal law in the very near future.

### 3. TREATMENTS FOR LEARNING DISABILITIES

As important as defining learning disabilities has been the process of conducting research on effective treatments. The literature on treatment for learning disabilities is broad and extensive, so this summary will necessarily be abbreviated. To focus this section further, primarily school-based treatments will be discussed, though individuals with learning disabilities experience challenges beyond the school setting. Finally, instead of attempting to chronicle every researched school-based approach, this summary is organized around three topics: treatments without widespread research support, general principles undergirding generally effective treatments, and selected examples of effective approaches for educating students with learning disabilities. At a general level, the treatments reported in this section reflect well-established intervention methodologies with significant research examining their effectiveness.

#### 3.1. Treatments That Do Not Have Widespread Research Support

##### 3.1.1. Modality Instruction

Modality instruction is the practice of assessing certain characteristics of the learner (i.e., modalities), such as whether the individual is a visual, auditory, or kinesthetic learner, and then matching treatments to the individual's strengths. It has long been thought that matching treatment to measured processing strengths can provide one mechanism to improve learning for persons with learning disabilities. The assumption is that if we teach individuals through their preferred modality, then they will learn more and/or better than they would if we did not tailor our instruction in this way. These approaches, sometimes called aptitude-by-treatment interactions or diagnostic-prescriptive teaching, have great intuitive appeal. They sound logical and reasonable. Unfortunately, the effectiveness of these approaches has not been empirically proven on a wide scale. For these methods to work, we need to be able to demonstrate that individuals with certain characteristics learn better with instruction matched to their strengths. These effects have been very difficult to find. Instead, treatments that are generally effective for students with learning disabilities also tend to be effective for students without learning disabilities.

##### 3.1.2. Psycholinguistic Training

Psycholinguistic training is an approach to training people in processes that they are believed to be deficit in. It was at one time a major intervention approach for students with learning disabilities. Training approaches were often based on results from a very popular test called the Illinois Test of Psycholinguistic Abilities (ITPA). This test measured integrative, receptive, and expressive linguistic abilities through presenting test subjects with information through visual and auditory channels. The assumption was made in this treatment approach, as in others, that discrete psycholinguistic abilities can be measured directly and then remediated. This assumption has been the subject of intense research scrutiny throughout the 1970s and 1980s. There were those who claimed with data that psycholinguistic training was generally not effective, and others with data who refuted these claims, stating that psycholinguistic training provided discrete benefits. The claims and counterclaims become a bit confusing. However, based on a review of the existing literature by this author, it appears that at least some of the areas measured by psycholinguistic assessments can be enhanced by psycholinguistic training. Particularly in the "expressive" areas of manual expression and verbal expression, evidence appears to confirm that these areas can be improved moderately through direct training. In the other 10 subareas measured by the ITPA, it appears that the effects of training are more modest.

Despite these moderately positive findings, a number of researchers have questioned the practical utility of gains of the magnitude reported. That is, it may be possible that gains in specific psycholinguistic variables are possible. The question is, do these gains translate into important gains in functioning in other areas of a person's life such as reading or language use? In the absence of these types of evidence, we must question whether these psycholinguistic interventions should be the highest priority for persons with learning disabilities.

##### 3.1.3. Perceptual Motor Training

Probably the most frequent form of "process training" in special education has been the use of what is generally called perceptual motor training. The theory goes that children with learning disabilities, or at least some of them, have problems in sensory integration to which some or all of their learning problems can be attributed. These problems may include disorders in the postural, ocular, and bilateral integration, a disorder in the ability

to plan and execute motor acts, a disorder in form and space perception, auditory language problems, and/or tactile defensiveness. Numerous clinical measures intended to measure these problems have been created over the years. For children with such problems, the idea is that direct sensory motor training could remediate these children's underlying problems, and hence mitigate the effects of their disability. There have been many testimonials to the effectiveness of perceptual motor training, and a number of clinical trials have fanned the popularity of these treatments. Research support on a broad scale, however, has not emerged for these therapies. It may be that persons with learning disabilities have problems in these areas, but consistent, effective treatments that result in consistent, large-scale positive effects have not been identified and validated on students' perceptual/sensory motor skills, their academic achievement, or their cognitive performance or aptitude.

### **3.1.4. Placement in Special Education**

Perhaps the most frequent treatment for children with learning disabilities in U.S. schools is placement in special education (as allowed by the IDEA). It is difficult to make blanket statements about special education as an enterprise, because so many different practices exist in these settings. Moreover, there are methodological problems associated with measuring the overall effectiveness of such a diverse enterprise. Despite these challenges, a number of studies have examined the effectiveness of special education placement for students with mild disabilities such as learning disabilities. In general, these studies have found weak effects for special education placement as a treatment. These findings, however, must be tempered by the awareness that many of the research-based effective practices identified in the last section of this article are most frequently carried out in special education placements. As such, it is probably most accurate to say that practices in special education are effective or ineffective, as opposed to placement. The extent to which any particular placement will be effective or ineffective for students with learning disabilities is a direct result of the effectiveness of practices used in that setting.

## **3.2. Characteristics of Generally Effective Instructional Interventions**

Researchers have identified a series of components that characterize generally effective instructional

interventions. These components do not guarantee that a specific program or approach will be effective with individuals with learning disabilities. However, when these principles are implemented in classrooms, children generally learn better and more. Some of the principles of effective instruction are listed below.

### **3.2.1. Teacher Organization**

Instructional groups should be organized for efficiency and sustained focus on the instructional content. This means that students in small group instruction should be seated so that the teacher can work with each student, transitions from one instructional activity to another should be quick (which may require instruction on how to transition), and lessons should be started quickly at the beginning of lesson time.

### **3.2.2. Student Engagement**

Students should not just be attending to academic material, but should be actively responding to material that is geared to their instructional skills. Students should be placed in materials that are challenging, in which they can respond briskly and in which they will not make large numbers of mistakes. They should experience high rates of success in the materials.

### **3.2.3. Opportunities to Respond**

Students should receive many opportunities to respond during instruction and should get clear feedback on the correctness of their performance from the teacher.

### **3.2.4. Overlearning**

Skills should be taught to a level of overlearning, with new skills being introduced gradually. That is, students need to be taught to a level at which they are able to perform skills almost automatically, without great effort.

## **3.3. Treatments for Learning Disabilities with General Research Support**

In addition to treatments that have not demonstrated high levels of effectiveness and broad research support, there are a host of treatments for learning disabilities that have been found to be generally effective. This section reviews some of the most widely used and

familiar treatments in schools for learning disabilities. The list is not exhaustive, but is intended to highlight some of the most effective approaches.

### **3.3.1. Early Intervention**

The goal of early intervention is the early detection and remediation of negative effects of disability and the prevention of developmental and learning problems for children who are at risk of developing these problems. There have been some very impressive results reported from the implementation of early intervention programs, though methodological issues have made it difficult to make large-scale, conclusive statements. In general, however, children who participate in early intervention programs have demonstrated gains in IQ, language development, motor development, and academic achievement.

One of the most exciting recent developments in learning disabilities is improvement in our understanding of early language development and early literacy. At the same time, assessment technologies have been developed that allow us to identify and remediate the problems of students with learning disabilities early, before they grow into larger ones. With these developments, learning disabilities advocates and professionals have moved one step closer to realizing their long-standing desire to provide meaningful, effective treatments for persons with learning disabilities early, when interventions can be maximally effective. We now know many critical “building blocks” of early reading and we have direct measures of these skills and each is directly teachable. Initial results for directly teaching children these building block skills appear promising. That is, many children who are at risk of learning failure seem to maintain and build upon their early gains after experiencing direct and focused reading instruction early in their educational careers.

### **3.3.2. Direct Instruction**

Direct instruction is a method of teaching that was developed by Siegfried Engelmann and Wesley Becker at the University of Oregon. These researchers married the positive effects seen in using behavior modification with a systematic analysis of how to teach different knowledge forms (e.g., facts, rules, concepts). To this knowledge, they applied the principles of effective instructional design and developed a series of curricula for teaching academic skills that have widespread and powerful research supporting their effectiveness. These

include the DISTAR and Corrective Reading series, among others. These programs provide explicit, step-by-step guidance for teachers, strategies for correcting student errors, gradual removal of teacher direction as students become more independent, the use of adequate, systematic practice with many different examples, and consistent review of recently learned materials. They have also generated some of the most widespread and consistent research support on program effectiveness of any instructional methodology.

### **3.3.3. Reading Comprehension Strategy Instruction**

Because many students with learning difficulties have problems with language variables that often translate into reading difficulties, a large number of studies have examined the effectiveness of teaching reading comprehension strategies to students with learning disabilities. Many different approaches have been examined, including the use of advance organizers, vocabulary interventions, the use rules to help comprehend text, self-questioning, the use of previewing and reviewing strategies, and memory enhancement strategies, to name a few. Though the effectiveness of different comprehension strategies varies, the results are generally in the good to excellent range. These findings are especially encouraging because they collectively demonstrate that reading comprehension can be enhanced significantly for students with learning disabilities and that a wide variety of strategies are available that can have significantly positive effects.

### **3.3.4. Behavior Modification**

Behavior modification is an approach to treatment that is based on the science of applied behavior analysis. Basically, this approach works through careful application of behavior change principles to the problems experienced by students with learning disabilities. In behavior modification, the behaviors that are the subject of change are defined precisely, very explicit and complete instructional procedures are used to teach new behaviors, the results are measured directly, and teaching procedures are modified to produce optimal effects. People who use behavior modification consider academic performance to fall in the realm of behavior just as they view social behavior. Indeed, behavior modification procedures have a very strong research base, underlying their use for both helping improve social behavior and developing academic

skill. Behavior modification has demonstrated consistent and strong relationships to improved student performance.

### **3.3.5. Peer Tutoring**

Peer tutoring is a set of strategies in which students serve as both teachers and learners. In general, students work with each other in a 1-on-1 setting, learning together. Typically, students serving as “teacher” at any given time follow directions regarding what and how to interact with the student serving as the “learner.” Peer tutoring has the advantage that it can individualize instruction for students based on individual needs. In many cases, all students participating in peer tutoring get to serve both as instructors and as learners. A range of positive effects have been documented in the peer-tutoring literature, including positive effects on academic achievement in reading, mathematics, and language, increases in students’ attitudes toward school, and improved school behavior. Gains have been documented for both tutors and tutees.

### **3.3.6. Progress Monitoring, Graphing, and Data-Based Decision Making**

One of the few universals that can be identified at the current state of our knowledge about individuals with learning disabilities is that we cannot predict with certainty which strategies will work prior to trying them. As a result of this situation, much research and development has been devoted to identifying ways to measure growth and development for individuals with learning disabilities. The process of monitoring progress across time, graphing data, and using these data for the purpose of determining effectiveness of treatments is called formative evaluation. Recall that development of formative evaluation tools was a major purpose for one of the federally funded institutes for research on learning disabilities in the late 1970s and early 1980s. Not only have progress-monitoring and decision-making tools been developed that are specifically useful with students with learning disabilities (e.g., curriculum-based measurement), but also research has proven that using these technologies to give teachers feedback on instruction improves student performance significantly. The idea is that if teachers can get feedback on the effectiveness of their instruction over a short period of time (say 6 weeks), they will be in a good position to either keep effective instructional strategies or move on to other potentially

effective strategies if what they are doing is not working. Progress monitoring, graphing, and data-based decision making provide teachers these tools.

## **3.4. Maximizing Results for Persons with Learning Disabilities**

In general, the specific problems associated with learning disabilities are lifelong, though many of the problems can be attenuated by instruction and accommodation. Students with learning disabilities can learn to read and can become functional readers. Students with learning disabilities who have memory problems can learn specific strategies to assist with remembering. Students with learning disabilities with organizational problems can learn strategies to assist with planning and keeping themselves organized. To maximize their functioning, however, individuals with learning disabilities may need a variety of supports throughout life.

During the school years, the problems associated with learning disabilities impact students’ performance in academic skill development, primarily in reading, language arts, and mathematics. In turn, students’ grades and academic achievement, two important success indicators in schools, are affected. As students mature and transition into adult life, the problems they experience can affect their lives in different and important ways. Low levels of academic achievement can diminish students’ likelihood of pursuing college, which can translate into less lucrative employment, less independence, or even lower standards of living. Specific skill deficits can preclude access to or participation in certain professions, and for some, deficits in social skills can make it difficult to maintain steady employment. For many individuals with learning disabilities, however, some of the most negative effects associated with their disabilities can be mitigated through instruction, advocacy, and advanced planning. Indeed, there are many, many persons with learning disabilities who have minimized the effects of their disabilities and have excelled personally and professionally.

To maximize the chances of success for students with learning disabilities, a number of scholars have argued that it is critical that increased attention be paid to the issue of preparing students for adult life while they are in school. Supports need to include vocational instruction and experiences, assistance in finding and securing employment, assistance in finding and participating in post-secondary education options, and assistance in learning to use community resources to the students’

advantage. For optimum effectiveness, transition services should include students themselves in planning their transition from school to adult life. Specifically, these services should include the student's family, should begin early (some advocate as early as elementary school), should be sensitive to familial and cultural factors, and must be comprehensive. It is through this careful planning and service delivery that successful results for individuals with learning disabilities can be achieved.

### See Also the Following Articles

Behavioral Assessment in Schools ■ Behavioral Observation in Schools ■ Emotional and Behavioral Problems, Students with ■ Learning ■ Mathematics, Teaching of ■ Neuropsychological Assessment in Schools ■ Reading, Teaching of ■ School Discipline and Behavior Management ■ Teaching Effectiveness ■ Writing, Teaching of

### Further Reading

- Bradley, R., Danielson, L., *et al.* (2002). *Identification of learning disabilities: Research to practice*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hamill, D. D. (1990). On defining learning disabilities: An emerging consensus. *Journal of Learning Disabilities*, 23, 74–83.
- Kirk, S. A. (1975). Learning disabilities: Selected ACLD papers. In S. A. Kirk, and J. M. McCarthy (Ed.), *Behavioral diagnosis and remediation of learning disabilities*. Boston, MA: Houghton Mifflin.
- Lyon, G. R. (1994). *Frames of reference for the assessment of learning disabilities*. Baltimore, MD: Paul H. Brooks.
- Mercer, C. D. (1997). *Students with learning disabilities* (5th Ed.). Upper Saddle River, NJ: Merrill.
- Ysseldyke, J. E., Algozzine, B., *et al.* (1982). Similarities and differences between low achievers and students labeled learning disabled. *Journal of Special Education*, 16, 73–85.



# Learning Styles and Approaches to Studying

Noel Entwistle

University of Edinburgh, Edinburgh, United Kingdom

Elizabeth Peterson

University of Auckland, Auckland, New Zealand

1. Historical and Conceptual Introduction
2. Learning Styles
3. Learning Strategies and Approaches to Studying
4. Measurement Instruments
5. Relationships to Achievement and to Learning Environments
6. Other Applications in Education and Business  
Further Reading

Learning styles are relatively consistent preferences for adopting particular learning processes, irrespective of the task or problem presented but generally relating to educational or workplace settings. Learning strategies and approaches to studying are specific to education and training, being more context-specific ways of tackling learning tasks that involve characteristic combinations of intentions, learning processes, and study activities.

## GLOSSARY

**approach to studying** A context- and content-specific way of carrying out academic tasks.

**cognitive style** A habitual and distinctive way of attending to, and processing, perceptual and cognitive information.

**disposition** A composite concept bringing together ability, inclination, and sensitivity to context.

**learning strategies** The way of learning adopted in tackling more specific types of tasks.

**learning style** A preferred and relatively consistent way of learning, usually related to educational or workplace settings.

**study orientation** A habitual way of tackling tasks presented in educational contexts.

**teaching-learning environment** The instructional activities, learning resources, assignments, and assessments set up to encourage and support learning.

**thinking style** A preferred way of thinking, described in general terms to apply to various contrasting situations.

## 1. HISTORICAL AND CONCEPTUAL INTRODUCTION

The term “cognitive style” was first used in the psychological literature during the 1940s to describe individual differences in perceptual ability, especially Witkin’s identification of “field independence,” that is, the ability to separate an object visually from its background. The meaning of cognitive style has since been widened to refer to other strong, relatively consistent individual differences in the way in which people cognitively process, as well as perceive, information.

Conceptually, cognitive styles have links to both personality and ability. Like personality traits, they are relatively fixed characteristics that affect the way in which people respond across a variety of situations, but their theoretical and empirical underpinnings are much less secure. Like abilities, cognitive styles affect

mental processing, but instead of indicating the content, speed, and effectiveness of mental operations, they describe characteristic processes through which information is handled. In addition, the level of ability (measured on a low-to-high scale) will affect performance, whereas cognitive style is indicated by the position on a bipolar scale, with each pole being equally adaptive but for rather different kinds of learning.

The term “thinking style” represents a concept intermediate between cognitive style and learning style but is seen as having broader applicability than the latter. Distinctive differences between convergent and divergent thinking scores were recognized as contrasting modes of thinking, equivalent to styles, but the term “thinking style” has more recently become associated with the work of Sternberg through his theory of mental self-government.

During the 1960s, the term “learning style” was introduced to describe individual differences in the more specific preferences shown by people learning in educational and workplace settings. Work focusing on study methods and study strategies began earlier within educational psychology, but the term “approaches to learning” emerged from educational research during the mid-1970s and its meaning was subsequently broadened to cover approaches to studying.

Generally speaking, “styles” describe relatively stable personal preferences in how cognitive processing or learning is undertaken, whereas “strategies” and “approaches” are contextually more specific, being influenced by the content and setting within which learning and studying take place.

## 2. LEARNING STYLES

The early work on cognitive style, led predominantly by Witkin, focused on individual differences in perceptual processing, but Witkin’s later work also distinguished broader “articulated” and “global” ways of thinking that influenced more recent work on styles. A vast array of terms has been accumulated in the attempt to describe distinctive preferences, with those associated with education and the workplace being known as learning styles. By 1976, Messick had already found 19 different sets of categories, and more styles have been introduced since then. This vast array of competing conceptualizations, with no widely agreed theoretical framework to connect them, makes it difficult to single out specific contributions, but the work of Kolb was influential in the early conceptualization of learning style. Kolb

identified two dimensions: one differentiating between preferences for either convergent or divergent thinking and the other indicating the relative reliance on the Piagetian mental processes of assimilation or accommodation. Relative closeness to each of the poles of the two dimensions was used to suggest four stylistic “types” that were labeled accordingly.

The subsequent proliferation of concepts describing learning styles, each with its own related learning style instrument, has led to confusion in the essential meaning of the term. However, it is possible to make sense of the literature from integrative reviews in terms of a few recurring dimensions, although with different labels. Recent reviews by Sternberg and Riding, among others, suggest that the main learning styles cover either individual preferences for a particular modality of sensory input (e.g., visual, auditory, kinesthetic) or different forms of mental representation in learning. These latter styles are exhibited in terms of preferences for one or the other pole of dimensions described as analytic/wholist, concrete/abstract, or verbal processing/visual imagery. There has been some suggestion that aspects of learning style may reflect differences in hemispheric brain functions, but the evidence for this is still contested.

Another way in which learning style constructs can be systematized usefully is by contrasting those that focus on the learning process itself (e.g., Pask’s version of analytic/wholist, which he labeled serialist/holist) with others that look at instructional and environmental preferences (e.g., the Dunns’ extensive array of categories related to classroom contexts).

## 3. LEARNING STRATEGIES AND APPROACHES TO STUDYING

Two distinct research traditions have developed for defining and measuring learning and study strategies: one applying preexisting psychological constructs and the other developing concepts from studies within the specific educational context. However, the most recent work can be seen as merging those traditions. After early attempts at using Likert scales to measure study methods and motivation, research in the United States drew heavily on cognitive psychology to describe students’ typical ways of tackling academic tasks, being particularly influenced by the distinction between active elaboration of information received and routine memorization of it. During the early 1980s, the work of Weinstein brought together these ideas with

practical advice on effective studying to link the measurement of study strategies with courses designed to help develop students' studying. Around the same time, Schmeck also used insights from cognitive psychology to develop an inventory of learning processes that described deep and elaborative processing as well as fact retention and methodical study.

More recently, Pintrich introduced several additional motivational constructs into the measurement of study strategies, including aspects of expectancy, value, and affect. Expectancy refers to students' beliefs about whether they can perform a task and involves self-efficacy and control beliefs about learning. The value component indicates why students engage in particular academic tasks and is covered partly by contrasting intrinsic and extrinsic goal orientation and partly by task value. This latter concept introduces affect by exploring the extent to which students find a particular task interesting, useful, and important. Learning strategies, as in the earlier work, focus on elaboration and are seen in relation to explicit and detailed metacognitive activities and self-regulation as well as volition. Interesting interplay among expectancies, motivation, and self-regulation, with consequent influences on achievement, has been reported in the most recent work, with tentative suggestions about possible gender and ethnic variations.

Whereas much of the research on learning strategies in the United States has increasingly been extended to work in schools, the alternative tradition, describing approaches, developed in Europe and Australasia and focused mainly on university students. The early work of Biggs was influenced by information processing theory, but most of the subsequent work has established conceptual frameworks out of interview and inventory studies carried out specifically in the higher education context. Using a distinctive form of qualitative analysis of interview data, Marton and his research team found a marked contrast among students in how they described reading an academic article in a naturalistic experiment. They distinguished a deep approach (which involved reading with the intention of understanding) from a surface approach (which relied on routine memorization to "get by"). The researchers argued that these approaches were specific to a particular academic context and were also dependent on the content of the learning task. However, the successful introduction of inventory measures indicated that approaches had a certain consistency, at least across similar learning environments and subject matter. Some researchers have suggested that such typical approaches should be referred to as "study

orientations," but there is still terminological disagreement in the literature.

Subsequent work by Entwistle and Biggs, working independently, described approaches to learning in terms of a combination of intention (or motive) and processes. For example, the deep approach involved the motive to understand for oneself combined with learning processes that involved an alternation between holist and serialist thinking. These descriptions were based partly on interviews and partly on factor structures of specially designed inventories, but it became clear that a third approach was necessary to explain the variations in studying reported. The strength of students' concern about their grades leads to the adoption of a strategic or achieving "approach to studying." The strategic approach seems to draw on either achievement motivation or a sense of responsibility for achievement, while the processes involve organized studying, effort, and attempts to monitor the effectiveness of studying.

This work was developed further by Vermunt, who introduced additional components relating to students' beliefs about the nature of knowledge, aims, and goals as well as to the balance between external regulation and self-regulation in studying (or the absence of regulation). This trend—seeing learning strategies in relation to epistemological beliefs—can also be found in the work of Pintrich and others in the United States. However, the most recent work by both Vermunt and Entwistle and their colleagues is distinct from the research on learning strategies in the United States in creating constructs that are composites. These cover a series of concepts that hang together so consistently that separating the component parts tends to destroy the sense of the experience reported by students. In this sense, they act rather like the "dispositions," described by Perkins, that bring together ability, inclination, and sensitivity to context.

#### 4. MEASUREMENT INSTRUMENTS

A wide and disparate range of learning style inventories exists. This is due in part to the failure to reach any strong consensus on the theoretical basis of learning styles. Thus, the instrument that is chosen will depend largely on the context to which it is applied and the type of learning style that the investigator wants to assess. Furthermore, as Coffield and colleagues recently argued, the inventories vary considerably in how much evidence of reliability and validity is provided and in the extent to which the evidence is independent of the originator of

the inventory. This is particularly problematic where the inventories have been vigorously marketed, with strong claims for their value in educational or workplace contexts yet without convincing psychometric substantiation. In contrast, there is greater agreement and fewer inventories describing the main learning strategies and approaches to studying, and the psychometric properties of the most fully developed inventories have been firmly established.

However, uncertainties about the instruments used in this area of research do remain because the majority of instruments have been operationalized in self-report formats, usually Likert scales. Thus, the validity of these instruments depends on the extent to which individuals are conscious of the strategies they typically use in their learning and studying and can accurately reflect on them. The items are sometimes quite transparent in meaning and value loading, and students can, to some extent, present whatever impressions they like. However, the style items tend to offer alternatives that can be viewed as equally attractive, and the factor structures of the strongest learning strategy and approaches to studying inventories are sufficiently consistent across both subject area and educational system to substantiate their validity claims. Some instruments have adopted forced-comparison formats or ratio scoring, but both of these methods create difficulties for conventional statistical analysis. Messick recently argued that the uncertainties of measurement that exist in the measurement of styles necessitate the use of normative and ipsative scoring techniques in combination as well as the collection of data from contrasting populations.

## **5. RELATIONSHIPS TO ACHIEVEMENT AND TO LEARNING ENVIRONMENTS**

The underlying purpose in using both learning style and approaches to studying concepts in educational research is to find ways of helping learners to become more aware of their styles and approaches and, thus, to seek ways of improving the effectiveness of their learning and studying. Evidence for the validity of the inventories rests, in part, on the relationships that exist between the scale scores and the measures of academic achievement. The unconvincing or inconsistent relationships often obtained are not surprising given the psychometric problems with several of the style

inventories, but these results could just as well be attributable to the varied definitions of achievement and its sometimes unreliable measurement in educational institutions. However, subscales of both learning strategy and approaches to studying inventories do relate to student achievement more convincingly but are still dependent on the type of learning being assessed. Although few studies have reported consistent gender differences, suggestive cultural and ethnic variations are emerging.

Theoretically more interesting research has been relating both style and approaches or strategies to teaching and learning in various subject areas, contrasting learning environments, and different educational stages. For example, in the style literature, there has been considerable debate about whether it is better to match or mismatch learner preferences with teaching preferences, how this should be done, who benefits from this, and who should decide from among the alternatives. Few large-scale, independent studies have been done to investigate this, but the findings do suggest that when the styles of the teacher and the learner are matched, learning is perceived as more congenial and teachers are rated more positively.

In the approaches literature, it is becoming clear that a deep, self-regulated approach can be encouraged through methods of teaching and assessment that focus explicitly on conceptual development and subject-specific thinking skills and that also require students to reflect on their own ways of studying. But the most recent work from both Biggs and Entwistle has stressed the necessity of thinking about the influence of teaching-learning environments on learning in terms of interacting systems. The students' approaches to studying and perceptions of teaching are part of one system in which the separate constructs used by researchers artificially separate aspects that are integrated within the individual and have their origins in earlier experiences. The idea of a "disposition to learn through personal understanding" would bring together ability, conceptions, intentions, style, motivation, strategies, and approaches, seen as constructs with different levels of generality and relative stability that work together to influence an individual's study behavior. The teaching-learning environment is a system that can be described at different levels as well, concentrating on the specific content and teaching methods devised by teachers, the broader social experiences of students, or the institutional and sociopolitical context within which education takes place. All of these

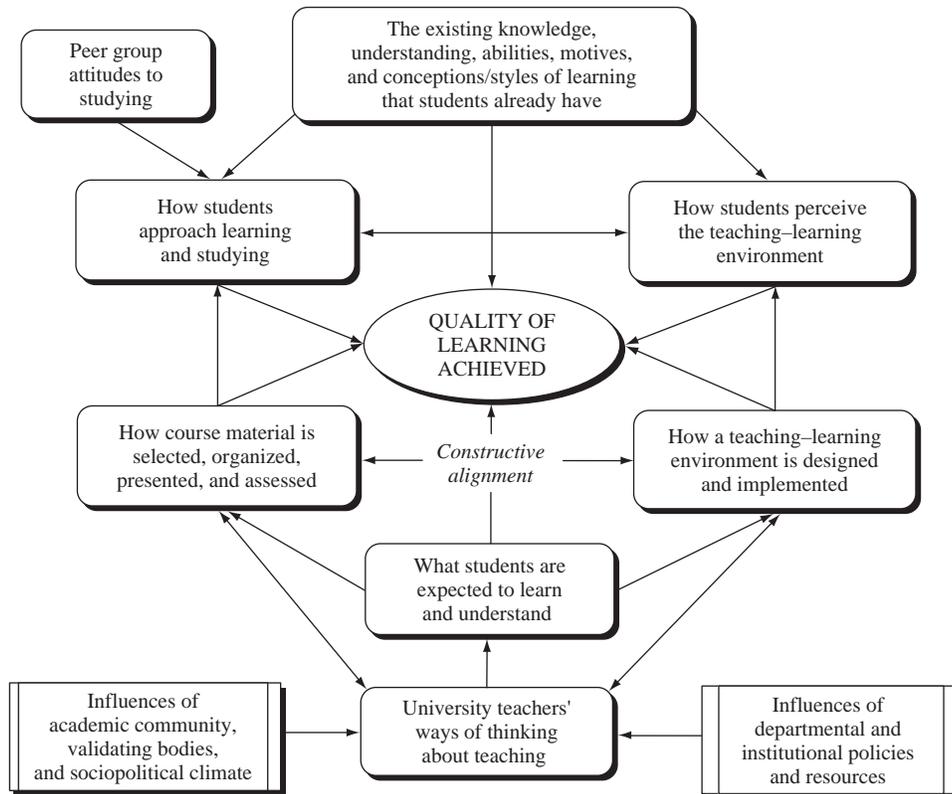


FIGURE 1 Interacting systems affecting the quality of student learning.

act together to influence student learning, but their influence is not uniform due to the different ways in which individual students react to the environment they experience. Figure 1 suggests what such interacting systems might look like in an outline based on ongoing research into teaching-learning environments in higher education.

However, the interacting systems do not show which aspects of each component of the environment are most likely to promote “deep learning.” For that, one must bring together a series of studies in schools such as Project Zero at Harvard University on “teaching for understanding” and de Corte and colleagues’ study in Belgium on “powerful learning environments.” Work at the university level also contributes to the listing shown in Table I, drawing on recent studies by Biggs in Australia on “constructive alignment” and by Entwistle, Hounsell, and their research team in the United Kingdom on ways of “enhancing teaching-learning environments.”

TABLE I  
Teaching Activities Encouraging a Deep Approach to Learning

- Specifying broad generic aims emphasizing understanding and using “throughlines” to remind students of those aims on a regular basis
- Ensuring that teaching methods and assessment procedures work in consort to support those aims, thereby providing constructive alignment
- Using instructional methods that provide clear conceptual explanations, are presented with enthusiasm, and demonstrate empathy with the learners
- Providing authentic tasks and realistic problems that have personal meaning and involve understanding performances
- Using group discussions of both the content and the process of learning and studying
- Promoting students’ awareness of their own cognitive processes and approaches to studying as well as their ability to control their motives, feelings, and efforts

## 6. OTHER APPLICATIONS IN EDUCATION AND BUSINESS

In spite of the lack of convincing evidence about the predictive validity of learning style inventories, they have been used widely in both education and the business world because the underlying conceptualizations tend to make intuitive sense to the users. The process of reflecting on the meaning of the scores then encourages metacognitive activity about learning preferences. However, learning style inventories have also been used for selection and promotion in companies, where any lack of convincing evidence about the instruments' psychometric properties and predictive validity has serious consequences.

Learning strategy and approaches to studying inventories have been used to identify students who reveal ineffective study strategies as well as to monitor the effects of innovations in teaching and learning. Although these inventories do seem to provide valid findings at the group level, their use in guiding individuals is much more problematic and needs to be supplemented by individual discussion of the responses prior to any changes in approach being suggested. Experience in using the research findings to encourage more effective studying strongly suggests that stand-alone teaching of study techniques is much less effective than building repeated suggestions and advice into the curriculum itself, thereby making them specific to the tasks that students undertake on a regular basis.

### See Also the Following Articles

Achievement Tests ■ Assessment and Evaluation, Overview ■ Cognition and Culture ■ Cognitive Skills: Training, Maintenance, and Daily Usage ■ Educational Achievement and Culture ■ Educational and Child Assessment ■ Intelligence Assessment ■ Learning ■ Mathematics, Teaching of ■ Psychometric Tests ■ Reading, Teaching of ■ Teaching Effectiveness ■ Writing, Teaching of

### Further Reading

Biggs, J. B. (2003). *Teaching for quality learning at university*. Buckingham, UK: Open University Press.

- Coffield, F., Ecclestone, K., Hall, E., Meagher, N., & Mosely, D. (2004). *A systematic and critical review of the literature on learning styles and pedagogy in post-16 learning*. London, UK: Learning and Skills Development Agency.
- Collis, J. M., & Messick, S. (2001). *Intelligence and personality: Bridging the gap in theory and measurement*. Mahwah, NJ: Lawrence Erlbaum.
- De Corte, E., Verschaffel, L., Entwistle, N. J., & van Merriënboer, J. (Eds.). (2003). *Powerful learning environments: Unravelling basic components and dimensions*. Oxford, UK: Elsevier Science.
- Dunn, R., & Griggs, S. (2003). *Synthesis of the Dunn and Dunn learning styles research*. New York: St. John's University Press.
- Entwistle, N. J. (2000). Approaches to studying and levels of understanding: The influences of teaching and assessment. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 15, pp. 156–218). New York: Agathon Press.
- Entwistle, N. J. (in press). *University teaching and student learning*. London: Palgrave Macmillan. (See also the project Web site at [www.ed.ac.uk/etl](http://www.ed.ac.uk/etl))
- Jonassen, D. H., & Graboski, B. L. (1993). *Handbook of individual differences, learning, and instruction*. Hillsdale, NJ: Lawrence Erlbaum.
- Kolb, D. A. (1984). *Experiential learning: Experience as a source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Messick, S. (1996). Bridging cognition and personality in education: The role of style in performance and development. *European Journal of Personality*, 10, 353–376.
- Pintrich, P. R., Brown, D. R., & Weinstein, C-E. (Eds.). (1994). *Student motivation, cognition, and learning*. Hillsdale, NJ: Lawrence Erlbaum.
- Riding, R., & Rayner, S. (1998). *Cognitive styles and learning strategies*. London: David Fulton.
- Schmeck, R. R. (1988). *Learning strategies and learning styles*. New York: Plenum.
- Sternberg, R. J., & Zhang, L-F. (Eds.). (2001). *Perspectives on thinking, learning, and cognitive styles*. Mahwah, NJ: Lawrence Erlbaum.
- Vermunt, J. D. (1996). Metacognitive, cognitive, and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education*, 31, 25–50.
- Wiske, M. S. (Ed.). (1998). *Teaching for understanding: Linking research with practice* [Project Zero]. San Francisco, CA: Jossey-Bass.



# Legal Competency

Norman G. Poythress

University of South Florida, Tampa, Florida, USA

1. Introduction: The Structure of Mental Health Laws
  2. Components of Legal Competencies
  3. Contexts for the Assessment of Legal Competency
  4. Issues in Forensic Assessment
- Further Reading

## GLOSSARY

*adjudicative competence* Functional legal abilities that a criminal defendant must have in sufficient degree in order for the adjudicatory process to go forward. Also called competence to stand trial, competence to proceed, and fitness to proceed.

*forensic assessment instrument (FAI)* An assessment test, tool, or device designed to measure or to inform a clinical judgment about specific functional legal abilities required in a given legal context.

*functional legal ability* A legally prescribed capacity that an individual must have in sufficient degree in order to participate in a legal proceeding (e.g., to face prosecution in a criminal court) or to avoid legal intervention into his or her affairs (e.g., to manage one's own financial affairs or to make decisions about one's medical treatment independent of a court-appointed guardian).

*mental state at the time of an offense (MSO)* General term inclusive of various legally recognized mental states (e.g., insanity, diminished capacity, and automatism) that provide the bases for a defense to culpability for criminal behavior.

Legal competence refers to a legally prescribed ability, or set of abilities, that a person must have in sufficient degree in order to enjoy, without government intervention, the rights, benefits, and opportunities to which citizens are ordinarily entitled. Issues of legal competence generally arise in discrete legal contexts (e.g., during police interrogation or court procedures), although some may arise in clinical (e.g., competence to make decisions about one's own medical treatment) or research (e.g., competence to consent to research participation) contexts. The conceptual and contextual framework for assessments of legal competence differ from those that guide more traditional therapeutically oriented assessments, and an array of forensic assessment instruments have been developed to assist mental health professionals whose evaluations address these legally focused competencies.

## I. INTRODUCTION: THE STRUCTURE OF MENTAL HEALTH LAWS

In the United States and many other democratic societies, the government cannot arbitrarily deny citizens the rights, benefits, and opportunities to which they are entitled. However, under either of two legal doctrines, citizens may be deprived of such rights and benefits if it is determined that they lack the requisite mental capacity to exercise such rights. One legal doctrine is police protection, which permits the

government to interfere in the lives of citizens whose behavior poses too substantial a risk to their own safety or the safety of others. This doctrine underlies, for example, involuntary hospitalization laws that permit the government to impose mandatory psychiatric treatment on individuals who pose a substantial risk of harm to others. The other doctrine is the *parens patriae* doctrine, in which the motive for state intervention is a paternalistic one. Procedures such as the appointment of legal guardians are intended to protect individuals from adverse outcomes in day-to-day situations that might result from their own poor judgments that stem from mental impairment. This doctrine underlies, for example, procedures that provide for the appointment of legal guardians to manage a person's personal and/or financial affairs.

Intrusions into citizens' lives on the basis of these rationales are prescribed in a collection of laws, broadly conceived as mental health laws. Associated with each of these laws is a legal competency that people must have in order to avoid state intrusion and to retain the full array of rights to which citizens are ordinarily entitled. As Morse elucidated most clearly in 1978, the structure of such laws involves three components: (i) a predicate mental condition that (ii) "causes" (iii) impairment in functional legal ability.

Most commonly, the predicate mental condition is described in legal terms, such as "mental disease" or "mental defect." Although in most instances there is no direct translation of these terms into contemporary psychological or psychiatric nosology, for more than a century mental health professionals have been utilized by the courts to provide evidence about the legal competence of parties to litigation. Historically, clinicians approached issues of legal competence using diagnostic interviews and general psychological tests (e.g., intelligence tests, general personality inventories, and projective tests) that they routinely employed in, and that were developed for, therapeutic applications. Today, however, the limitations of such methods for assessing legal competencies are widely recognized. Although the information gleaned from diagnostic interviews and traditional tests may be informative to judgments about the first element—predicate mental condition—these methods typically have not been normed or validated on relevant legal populations in ways that help inform the causality and functional impairment elements. In other words, the components of legal competencies involve considerations that go beyond what can be gleaned from conventional clinical assessment approaches.

## 2. COMPONENTS OF LEGAL COMPETENCIES

Grisso identified five features of legal competencies that distinguish them from more general competence constructs: a functional component, a causal component, an interactive component, a judgmental component, and a dispositional component.

### 2.1. Functional Component

The primary concern in forensic evaluations is the person's functional legal ability—the third element in the structural equation of mental health laws. As far as the law is concerned, it is the degree of impaired legal functioning, not the mere presence of mental disorder or generally impaired mental functioning, that determines whether a person's usual rights or privileges may be denied.

Several additional observations are important here. First, the particular functional abilities of concern vary from one legal context to another; to compare a couple of contexts alluded to previously, involuntary hospitalization proceedings focus on (among other things) a person's capacity to provide adequately for his or her basic needs (i.e., food and shelter), whereas guardianship proceedings focus on the capacity to manage one's personal affairs (e.g., to manage money and to enter contracts). Second, although a person's general mental condition is a relevant consideration insofar as a legally cognized predicate condition must be established, diagnosis is of secondary importance to functional legal ability. A person may have no mental disorder or several of them, but state intervention is justified not on the mere finding of mental impairment. Third, for any given legal context the relevant functional abilities are relatively narrow, and it is for this reason that some traditional clinical measures may be limited as a basis for inference about legal incompetence. Intelligence tests can provide information about a person's general problem-solving ability, but the legal questions will often be more narrowly focused and therefore require a more nuanced analysis as a basis for inference about impaired legal functioning.

### 2.2. Causal Component

Judgments about legal competence require causal inferences about the reasons for a person's functional abilities or deficits. Legal criteria for competence, often found in statutes governing a particular type of

proceeding or conflict (although criteria may also appear in appellate cases, administrative regulations, or other legal sources), often specify that the third element—impaired legal functioning—must be “as a result of” the predicate mental condition (first element).

Research that relates specific mental disorders, symptom patterns, neurological impairments, cognitive capacities, and so forth to functional legal abilities provides a basis for bridging the assessment of mental condition and legal competence. A further consideration in legal competence assessments is to distinguish legally relevant “causes” from irrelevant ones. For example, the legal context of adjudicative competence—more familiarly called competence to stand trial—requires (among other things) that a criminal defendant be able to adequately assist his or her attorney. For a defendant who is not able to do so because of impairments attributable to mental disease or defect, the adjudicatory process must be postponed until he or she regains such ability. However for a defendant, even one who has a mental disorder, who is merely unwilling to work with his or her attorney, the process can go forward. Assessments of legal competence may call for mental health professionals to aid the legal system in making such distinctions.

### 2.3. Interactive Component

Judgments regarding a person’s legal competence are not dependent merely on his or her mental capacity or the absolute level of his or her functional ability. His or her particular circumstances within the legal context may also be of relevance. It is the fit or congruence between capacity and demand that is ultimately of importance. From this perspective, one can see that courts (or other legal decision makers) might reach different conclusions about the legal competence of two individuals who have the same (measured) abilities.

Consider the adjudicative competence context, for example, for two individuals of modest mental capacity (e.g., judgment, memory, and general knowledge). One might be facing very simple charges such as driving without a valid license; the only issues of contention in the case are likely to be whether he or she was behind the wheel of a car and whether he or she possessed a valid license at the time. The other individual may be charged with intention to commit securities fraud, for which the defense (and required degree of assistance to counsel) is likely to be much more prolonged and complex. A mentally disordered individual of modest functional capacity might reasonably

be judged “competent to proceed” in the first case but perhaps not in the second.

As this example illustrates, it will be not only the individual’s measured level of psycholegal functioning but also the match of capacity with anticipated demand that will be relevant to the ultimate legal judgment of competence. Thus, to the degree possible, assessments of legal competence may require that the mental health professional assess and take into account the anticipated demands of the situation on the defendant.

### 2.4. Judgmental Component

In considering the adequacy of the fit between a person’s abilities and the demands of his or her legal situation, the law applies a variety of flexible legal tests. As will be evident in the next section, words such as “sufficient,” “reasonable,” or “substantial” are common qualifiers of the degree of capacity that the law demands.

There are at least three important consequences of the use of these qualifiers. First, these terms facilitate individualized judgments by legal decision makers. The competency criteria, although explicit, are not of the “hard-and-fast,” “one-size-fits-all” variety. Such terms permit a reasoned application of complex rules in judging cases that may vary radically in their particulars. Second, these terms reflect that most legal judgments about competence also involve social and moral considerations. In a criminal case involving an insanity defense, for example, such as the notorious case of John Hinckley, the decision about the accused’s mental state at the time of the offense is ultimately one of moral culpability. That is, within the framework of the written definition of legal insanity, was this particular individual’s impairment of the type that would warrant his or her exclusion from moral culpability for criminal behavior?

Following from the first two observations, the third consequence is that mental health professionals are (or at least logically should be) limited in the scope of conclusions that they draw in legal competency evaluations. The social and moral issues alluded to here are of a type that go beyond and stand outside the scope of scientific inquiry. Psychology and psychiatry have no “capaci-meters” for distinguishing the moral gradations that are part and parcel of legal competency judgments. Thus, the challenge to forensic examiners in legal competence evaluations is to provide clinical information that informs the judgment about competency while stopping short of offering conclusory opinions as to what that ultimate judgment should be.

### 2.5. Dispositional Component

Legal outcomes or dispositions hinge on the judgmental question—tersely put, whether or not the individual is judged “competent” (or “insane” or other analogous terms). The preceding section suggests that mental health professionals should not venture categorical conclusions about the competency judgment due to the social and moral nature of such decisions. It follows, therefore, that they should not presume to tell a legal decision maker what the appropriate disposition should be in a given case. Rather, clinical input regarding dispositions is best framed using “if—then” formulations that do not constrain the range of possible legal judgments. Whether, how, and the type of clinical assessment and treatment or other intervention might be provided can be discussed separately for different hypothetical outcomes; for example, “If the court decides X, then . . . . If the court decides Y, then . . . .”

## 3. CONTEXTS FOR THE ASSESSMENT OF LEGAL COMPETENCY

Issues of legal competency arise in many situations, several of which have been alluded to previously. In this section, several of the more frequently occurring legal competency questions are discussed in-depth. The reader is warned that not all legal competency

issues are recognized in all jurisdictions; for example, legal insanity is a recognized defense throughout most of the Western world, although in the United States there are approximately four states in which this defense is not available. Similarly, the formal legal criteria or “tests” of competency or capacity are not identical across jurisdictions. Thus, the presentation here is necessarily more conceptual and generic in nature.

### 3.1. Competencies in the Criminal Justice Context

Table I summarizes several issues of “competence” that arise in the criminal adjudicatory process.

#### 3.1.1. Competence to Waive Miranda

Confessions are the primary evidentiary basis for many criminal convictions. In the United States, the legal basis for challenging the admissibility of defendants’ confessions resides in the Fifth and Sixth Amendments to the Constitution, which protect the defendant from compulsory self-incrimination and guarantee the right to the presence of legal counsel at all relevant stages of the adjudicatory process. Thus, when the accused individual is questioned by the police he or she must be informed of the legal rights that apply at the time of questioning. These include the right to have an attorney present during

TABLE I  
Issues of Legal Competence in the Criminal Adjudicatory Process

Competence issue	Fundamental inquiry	Legal criteria (generically)
Competence to waive Miranda rights	When giving a statement to the police, did the defendant competently waive his or her right against self-incrimination?	Cognitive issue: Was the defendant’s waiver made knowingly and intelligently? Volitional issue: Was the defendant’s waiver made voluntarily?
Adjudicative competence	Can the defendant competently participate in any of several phases of the adjudicatory process?	Does the defendant have a factual and rational understanding of the proceeding, and can he or she adequately assist counsel at the proceeding?
Mental state at the time of the offense	Retrospective inquiry: Was the defendant morally culpable for his or her criminal behavior?	Cognitive issues: Did the defendant know right from wrong? Understand the nature and quality of his or her acts? Volitional issues: Could the defendant control his or her behavior? Was his or her will overcome by an irresistible impulse?
Competence to be executed	Does the offender understand the nature and purpose of the sentence of death?	Effectively the same as for adjudicative competence but applied at the post adjudication phase of sentencing.

questioning, the right to have an attorney appointed for them if they cannot afford to hire their own attorney, and the right to refuse to answer questions altogether. They must also be advised that what they say during interrogation may be used in court to prove their guilt.

Challenges to the competence of a defendant's waiver are almost always made after-the-fact. Defense counsel hopes to have the defendant's prior statement ruled inadmissible at a pretrial "suppression hearing" on grounds that the defendant did not competently waive his or her rights. To the degree that the state's case depends on incriminating information obtained from the defendant, the outcome of such suppression hearings may be critical.

One key feature of clinical assessment of competence to waive Miranda is that the evaluations are almost always retrospective. The attorney finds out days or weeks later that the defendant has given a statement to the police. The mental health expert who consults with the defense is asked to evaluate the nature and circumstances of that prior interrogation and to develop clinical information that informs the judgment of the defendant's competence at that time. Thus, the evaluation often hinges in part on archival information (e.g., transcripts or recordings of the defendant's statement) and the recollections of those present (e.g., interviews with the defendant and police) as bases for reconstructing the defendant's prior mental state.

Another feature is that the legal "test" for evaluating this particular legal competence, unlike most others, does not involve an explicit "mental disease" or "mental defect" type of predicate. Rather, the court is required to consider "the totality of the circumstances" that may have impinged on the defendant's competence to waive his or her rights, including a consideration of the appropriateness of the police behavior during the interrogation.

The criteria, as indicated in Table I, note the relevance cognitive factors—whether the waiver was made knowingly and intelligently. Personality features such as general intelligence and susceptibility to suggestion or influence of authority figures are often relevant in these cases. Similarly, whether one's capacity to make a rational or well-reasoned waiver was adversely affected by drug or alcohol intoxication or symptoms of mental illness is a relevant inquiry. Some standardized tests of Miranda comprehension, modeled after more general vocabulary and verbal comprehension tests, have been developed to assist in making a highly focused inquiry into the relevant cognitive capacities.

Although Table I also indicates that volitional issues—the "voluntariness" of a waiver—are relevant,

the law in this area has developed in a way that limits severely the relevance of clinical input on this issue. The higher courts have generally held that there must be clear evidence of inappropriate behavior by the police (e.g., excessive threats or other coercive behavior and interrogation periods of unreasonable duration) in order to infer that the defendant's waiver was not of his or her own free will. Intrapersonal factors, including the presence of symptoms such as delusional beliefs, obsessive thoughts, or hallucinations that "urge" the defendant to confess, have generally not been accepted as a basis for concluding that a waiver was not voluntary. Because mental health professionals will typically have no particular expertise in judging when and whether the police have exceeded legally acceptable norms in their interrogations, their assessments are typically limited in relevance to issues of cognitive capacity.

### 3.1.2. Adjudicative Competence

Adjudicative competence refers to a defendant's capacity for meaningful participation in the adjudicatory process. Historically, the term *competence to stand trial* was often used, although it seems clear in retrospect that this term was inappropriately narrow. In particular, criminal justice data reveal that more than 90% of defendants resolve their cases by entering pleas rather than by going to trial; thus, competence to stand trial is literally a moot concern in the vast majority of cases, and "competence to enter a plea" is a much more relevant consideration. Although competence for both of these phases—participating in a trial or entering a plea—involves the "understand and assist" test shown in Table I, so too does competence for other adjudicatory activities, including sentencing. Thus, the more comprehensive term adjudicative competence better describes the breadth of this legal competency test.

In the United States, the legal basis for adjudicative competence questions reside in the Sixth and Fourteenth Amendments to the Constitution. The Sixth Amendment guarantees the right to the assistance of counsel and to confront one's accusers; these rights cannot be effectively implemented by an individual whose mental capacities are too eroded by "mental disease or defect." The Fourteenth Amendment, which deals with issues of fairness in legal proceedings, is also implicated. Although these legal foundations reflect the defendant's concerns, the state has its own independent rationale for ensuring that only competent defendants are prosecuted. That concern is with public opinion in the larger world

community about the kind of justice afforded citizens in this country.

Because of these multiple rationales for the adjudicative competence inquiry, questions about the defendant's competence can be raised by either party or by the court on its own motion. When such motions are granted, the adjudicatory process is temporarily suspended pending the result of clinical evaluations of the defendant's competence. Every state and the federal government make provisions for government-funded evaluations of the defendant's competence to proceed; both the state and the defense may also privately retain mental health consultants at their own expense if they want to obtain independent evaluations. When the clinical evaluations are completed, a hearing may be held before a judge or jury (depending on jurisdiction rules) to determine whether the defendant is competent to proceed. If the defendant is found competent, the adjudicatory process continues (e.g., a date for entering a plea, or a trial date, may be set). If the defendant is found incompetent, then he or she is typically referred for psychiatric treatment to "restore" competence. At such time as the defendant's competence is restored, he or she is returned to court and the adjudicatory process resumes.

Unlike competence to waive Miranda, adjudicative competence is a present state evaluation. The focus is on whether the defendant is now (and in the reasonably short-term future) able to understand the proceedings and assist counsel. Also, in most (if not all) jurisdictions it is explicit that only impairments resulting from "mental disease or defect" are legitimate bases for finding the defendant incompetent.

The relevant functional legal abilities for adjudicative competence are shown in Table I. Factual understanding concerns a defendant's awareness and comprehension of how the proceedings (e.g., entering a plea or a trial) are designed to work in principle. Relevant queries include knowledge of the proper roles of participants (e.g., "What is a judge, prosecutor, or jury supposed to do?"), the nature of the proceedings themselves (e.g., "What is the nature of a plea agreement?"), and so forth. Rational understanding is more specifically focused on the defendant's beliefs about his or her own case; that is, even if a defendant has a good "civics book" understanding of how the process is designed to work, is his or her view of how the process will play out in this particular case compromised by delusional beliefs (e.g., that God will intervene to assist in the defense)? Capacity to assist counsel comprises the ability to understand the defendant's role (i.e., as the one who stands accused and whose freedom is in jeopardy); to recognize and put

forth information that will facilitate the defense attorney's efforts to construct a defense; and the ability to participate, with the degree of autonomy that the law requires, in decisions that the attorney is proscribed by law from making on behalf of the defendant (e.g., whether to enter a guilty plea and whether to testify).

A number of measures or tools, most developed with substantial input from mental health professionals, are available for use in forensic evaluations of adjudicative competence. They range widely in design and structure, from some that constitute little more than menus of topics or issues to raise with defendants to others that constitute semistructured interviews with ratings of (in)capacity on explicit dimensions (e.g., knowledge of roles of court participants and comprehension of available pleas) and measures that have standardized inquiries and interpretive norms for selected psycholegal abilities.

### **3.1.3. Mental State at the Time of the Offense**

Retribution and deterrence are two of the primary legal justifications for adjudication and punishment of individuals who violate society's laws. However, dating back to antiquity, various cultures have recognized that it is morally inappropriate to hold certain classes of individuals fully responsible for criminal behavior. Mentally disordered individuals constitute one such class of individuals. Compassion, rather than vengeance, has been recognized as a more appropriate emotional reaction to individuals who were so mentally impaired that they "did not know what they were doing" or "could not help themselves." Furthermore, such individuals would not likely be deterred from future criminal behavior through the usual mechanisms of punishment. Hospitalization and treatment have been recognized as morally more appropriate societal responses to such individuals. Thus the rationale for the insanity defense and other mental state defenses.

The insanity defense is premised on the presence of legally defined mental disease or defect. Although there are no direct translations of recognized psychiatric diagnoses into such categories, empirically it is clear that this defense is usually reserved for individuals afflicted with the more debilitating disorders (primarily psychoses) or other acute symptom presentations that resulted in severe impairments in rational judgment (e.g., "knowing right from wrong" or capacity to control one's behavior) (Table I). The defense usually will not accommodate a defendant who was generally anxious, mildly depressed, disgruntled, or was merely "having a bad day."

The clinical inquiry into a defendant's sanity is a highly focused, retrospective evaluation. As with other legal competencies, the issue is not whether, at the time of the offense, the defendant merely had symptoms of a mental disorder that the trier-of-fact might judge constitutes a mental disease or defect. Mentally ill individuals may commit crimes for the same reasons that non-ill individuals do; having schizophrenia or bipolar disorder does not insulate one from the avarice that can motivate theft or robbery, nor do such disorders protect one from aggression based on anger or vengeance.

Thus, the critical linkage is the causal relationship (if any) between the symptoms of mental disorder and the alleged criminal behavior, and these linkages may be quite subtle or complex. For example, consider the case of a mentally ill person who fires a handgun through the skylight of his apartment because of delusions and hallucinations that space invaders are circling overhead in flying saucers, trying to destroy his brain with invisible radioactive rays that they beam into his home. On charges related to improper or illegal discharge of a firearm, this person may qualify for a judgment of insanity under at least some of the criteria listed in Table I. However, if this person is an ex-convict, then he may also be charged with parole violation because of his being a "felon in possession of a firearm." For this charge, the pertinent inquiry is the nature and quality of his reasoning in relation to obtaining the firearm in the first place. If he obtained the gun many months earlier for reasons unrelated to mental disorder (e.g., "I got it for protection because I live in a bad neighborhood"), presumably the insanity defense would not be relevant to his charge of parole violation.

As should be obvious, the defendant's recollection of his or her thoughts, feelings, and behavior at the time of the offense is critical to mental state at the time of the offense (MSO) evaluations. However, such self-reports may not be completely valid for a variety of reasons. MSO evaluations are often arranged weeks or months after the occurrence of the crime in question or the arrest of the accused. In the intervening period, the defendant's mental condition may have been affected by any number of factors—the natural course of illness, psychiatric treatment received (e.g., in jail or as a result of a finding of incompetence to proceed), or his or her emotional reaction to the offense (e.g., emotional reaction to having killed a loved one). Furthermore, when no other defense is viable, the potential for a defendant to feign insanity in an effort to escape prison must be considered. Consequently, as in the case of other retrospective assessments, the clinical evaluation of MSO often relies

extensively on archival, historical, and third-party information.

### **3.1.4. Competence to Be Executed**

In the United States, the courts have held that it is a violation of the Eighth Amendment protection against cruel and unusual punishment for the state to execute an "insane" person. Perhaps the closest the courts have come to defining the criteria for this particular competence was Justice Powell's concurring opinion in *Ford v. Wainwright*, which stated that the state is forbidden from executing only "those who are unaware of the punishment they are about to suffer and why they are to suffer it." A broader criterion suggested by the American Bar Association includes the provision that a defendant be able to recognize and convey to counsel (or a court) any facts that might constitute a basis for appealing this sentence.

Conceptually, these provisions constitute a particular instance of the general criteria for adjudicative competence—the defendant must understand (rationally and factually) the nature and object of the proceedings against him or her and be able to assist counsel. In terms of assessment, the considerations described previously regarding adjudicative competence are relevant here, with a particular substantive focus on the purpose (retribution) and nature (loss of life) of capital punishment.

For a defendant who is judged incompetent for execution, the proceedings are stayed and he or she is referred for psychiatric treatment to restore competence. Perhaps more than for any other forensic issue, mental health professionals feel ethically compromised when they consider that the outcomes of their evaluations or treatment of such defendants may provide the final legal basis for putting them to death. Consequently, this is one area of forensic assessment that some clinicians exclude from their practices.

## **3.2. Civil Competencies**

Apart from the criminal justice system, there are a variety of other social contexts in which concerns about citizens' legal capacities arise. Table II provides a brief description of several such civil competencies.

### **3.2.1. Guardianship**

Some citizens develop mental disabilities that substantially impair their capacity to make, or to communicate, reasonable decisions in day-to-day situations, such as

TABLE II  
Civil Competencies

<i>Competence issue</i>	<i>Fundamental inquiry</i>	<i>Legal criteria (generically)</i>
Need for a guardian to manage one's personal affairs	Should the individual's right to self-determination in day-to-day affairs be limited or removed due to mental incapacity?	Lacking sufficient understanding or capacity to make or communicate responsible decisions (in any of a wide array of day-to-day aspects of living)
Competence to make treatment decisions	Can the person make decisions about medical or psychiatric treatment that are in his or her best interest?	Understand and/or appreciate information relevant to one's illness and alternative treatment options Rationally weigh the risks and benefits, advantages and disadvantages, of treatment alternatives
Competence to consent to research participation	Can the person consent to research participation?	Adequate understanding of the required elements of informed consent in a research consent disclosure Capacity to make a rational decision to participate in the proposed research study
Testamentary capacity	Is the person competent to make a will?	Knowing the extent of their property and the "natural objects of their bounty" (potential heirs) Know that they are making a will and the manner in which their property will be distributed when the will is executed

attending to food and shelter needs, managing finances, and entering contractual agreements (including marriage). One common situation is that of elderly individuals afflicted with Alzheimer's disease or other dementia; some individuals with chronic mental disorders or substance abuse-related illnesses are also so afflicted. Under the *parens patriae* doctrine, most states have provisions for the appointment of responsible third parties (often relatives) to make decisions for individuals who are substantially impaired in this regard.

Of the various legal competencies, the guardianship issue is potentially one of the most sweeping in its scope. As the previous examples illustrate, the ability to "manage one's affairs" can cover a multitude of decisional domains. In recognition of the potentially extensive degree of intrusion that guardianship judgments may entail, the law in many jurisdictions distinguishes between plenary and specific (or limited) guardianships. In the former situation, the court appoints a legal guardian with the authority to make decisions in all (or nearly all) domains of a person's life; in the latter, only specific

decisions or types of decisions are placed under the authority of a guardian. A special instance of guardianship is conservatorship, in which a guardian is appointed only for purposes of managing the individual's financial assets (e.g., conducting financial transactions). Out of respect for individual autonomy, there is a preference for specific, as opposed to plenary, guardianship arrangements when the clinical circumstances permit. Thus, when mental health professionals participate in clinical assessments to inform guardianship decisions, their task is not only to determine whether and how severely a respondent's decision making is impaired but also to attempt to determine whether there are areas of decision making in which the respondent's capacity for judgment remains relatively intact.

As with other legal competencies, the mere presence of a mental disorder (i.e., diagnosis) is of little value to the legal decision maker. Although a predicate condition may be required, the legal criteria (depending on, and varying by, jurisdiction) often demand that impairment in specific mental functions (e.g., alertness and attention,

memory, and thought process) be demonstrated and the impact of these impairments on specific decisional behavior (e.g., how to manage finances) be elucidated.

### **3.2.2. Competence to Make Treatment Decisions**

A specific decisional capacity that might be considered a special area of application for guardianship is the capacity to make decisions about medical treatment. This issue arises with respect to both medical and mental health treatment, although it is one for which special procedures and venues (outside formal guardianship proceedings) have developed. As with guardianship proceedings, the situation is one in which a person's right to self-determination (autonomy) is the focal point.

As Table II indicates, the substantive issues involve a person's capacity to comprehend, and to act rationally on, information relevant to making a decision about treatment. Such information includes awareness of one's own illness, the array of potential treatments available, the risks and benefits anticipated with each treatment, the likely course of illness if no treatment is accepted, and so forth. In the context of decision making about whether to accept psychiatric treatment, the courts have recognized varying tests of capacity. These tests parallel, to some degree, the factual understanding, rational understanding, and capacity for autonomous decision making discussed previously in the adjudicative competence context. The person should be aware of the symptoms, nature, and anticipated course of the illness with which he or she is diagnosed; appreciate (i.e., acknowledge) that he or she in fact has the illness and that the probabilities regarding course and outcome—with and without treatment—apply to him or her; and be able to identify and compare risks and benefits in a way that informs judgment about which course of treatment (including refusal of treatment) to follow. Grisso and Appelbaum offer an extensive discussion of the legal framework for treatment decision making and a research-based procedure for clinical evaluations of treatment competence.

Unlike the other legal competencies discussed so far, in some jurisdictions treatment competence determinations are not made in a formal legal arena (i.e., in court) by legal decision makers (i.e., judges or juries). Rather, in some jurisdictions determinations of treatment competence regarding the acceptance or refusal of psychiatric treatment (mainly, psychotropic medications) are left to mental health professionals. Although there may be provisions for administrative review or appeal of clinicians' decisions about the person's need for treatment, the initial judgment whether to impose

involuntary treatment may rest with the attending physician and professional colleagues who provide an independent review of that physician's judgment. In these situations, the clinical judgments do involve the judgmental and dispositional components of legal competencies that are left to judges and juries in other contexts.

### **3.2.3. Competence to Consent to Research Participation**

The Nuremberg Code and subsequent guidelines established by governmental administrative bodies, such as the Belmont Report, provide the legal and conceptual frameworks for research involving human subjects. Briefly, no person can be involved as a research participant unless he or she gives competent consent to such participation. In order to give consent, the prospective research participant must be provided information relevant to the decision (whether or not to participate) and demonstrate adequate comprehension of that material. As in the situation involving medical treatment, the research investigator may be liable for a battery against an individual who has not consented to research participation.

As with guardianship and treatment competence, individual autonomy and self-determination are central values at stake in this context. People with sufficient mental capacity to understand research disclosures and the parameters, risks, and potential benefits of research described therein should be allowed to decide for themselves whether or not to participate. However, impaired decision-making capacity can derive from a variety of situational (e.g., stress) or intrapersonal (e.g., mental illness) sources, and some individuals may not be capable of the degree of comprehension and rational judgment necessary to give informed consent. Research suggests that people with dementias (e.g., Alzheimer's patients) and major mental disorders with psychotic features (e.g., schizophrenia) may have the greatest difficulty comprehending research consent disclosures.

Locally, research procedures are approved and monitored by institutional review boards (IRBs). Procedures for eliciting informed consent may vary depending on the type and level of risk that the research entails. Some studies involve minimal or virtually no risk to participants (e.g., an anonymous survey study that does not involve questions about personal or sensitive issues) and in certain circumstances the IRB may waive the requirement that informed consent be obtained. In studies involving greater than minimal risk, investigators are routinely required to obtain written informed consent.

This legal competency is one for which the determination of capacity is not made in a legal forum (e.g., a courtroom) or by a formal legal decision maker (i.e., a judge or jury). Almost universally, the judgment that a person has given competent consent to research participation is made by the study's principal investigator (PI) or research team members trained by the PI in informed consent procedures. The substantive elements of research consent are identified in the U.S. Code of Federal Regulations, and IRBs are responsible to see that PIs include required elements in their written consent disclosures.

There are currently no standardized or objective procedures for evaluating a prospective research participant's level of comprehension of research consent disclosures, which can vary radically in length and complexity from one type of study to another. Miller and Willner recommended brief objective testing of consent disclosure comprehension as one approach to the systematic evaluation of capacity to consent, and Poythress suggested ways that normative assessments of competence to consent might be structured. Currently, research investigators' ad hoc questioning and request for paraphrasing of disclosed material is likely the approach taken with most prospective study participants.

### **3.2.4. Testamentary Capacity**

Planning for the disposition of one's property after death commonly involves making a will. For a will to be valid, the person must have been "of sound mind" at the time that the will was prepared. As Table II indicates, this criterion is ordinarily framed in terms of awareness of the extent of one's property and due consideration of the family members and friends who might be considered the most likely heirs to one's money or estate.

As with other legal competencies, testamentary capacity does not hinge on the mere presence or absence of mental disorder. Mentally disordered individuals may make valid wills, and in order to invalidate a will the courts require a showing that the mental disorder "caused" impairment in the person's ability to evaluate the nature and extent of his or her estate or to consider those to whom the estate might naturally be passed on. Knowledge of one's property does not necessarily have to be detailed, and merely forgetting a distant relative with whom one was marginally involved does not necessarily invalidate a will.

Thus, the assessment will often focus on the individual's capacity to describe the various facets of the estate and to explain the reasons for including, or excluding,

potential heirs or the amount of property to be distributed to each. An individual may exclude or severely limit the property planned for distribution to presumably natural heirs (e.g., children or other close relatives) for reasons based in reality (e.g., residual animosity from prior disputes). In contrast, delusional beliefs about such individuals might be judged as a wrongful basis for excluding an otherwise deserving relative or friend, and evidence of such delusions might persuade a court that the person was not of sound mind when the will was prepared.

Challenges to testamentary capacity are likely to be brought by individuals, often family members or friends, concerned about the anticipated distribution of the individual's estate. Such challenges may be brought at the time the person prepares to make a will, in which case the evaluation is a present state assessment. Challenges brought after a will has been made and filed, including challenges after the person is deceased, result in necessarily retrospective evaluations in which the focus is on the testator's mental state at the time the will was made. As in other retrospective evaluations discussed previously, third-party information that helps to reconstruct the testator's thoughts, feelings, and attitudes from the relevant prior point in time may be essential to the assessment.

## **4. ISSUES IN FORENSIC ASSESSMENT**

As should be clear from the foregoing discussion, assessments of legal competencies differ in significant ways from more traditional psychological and psychiatric evaluations that target primarily the diagnosis of mental disorder and the development of treatment plans. The following sections highlight some of the important ethical and assessment features that distinguish evaluations of legal competencies.

### **4.1. Ethical Issues**

Clinicians who participate in the assessment of legal competencies face a number of ethical issues with which they must be prepared to cope. An excellent resource in this regard is the work of Greenberg and Shuman.

One primary issue is that of role clarification. The person whose legal competence is at issue is rarely the person who contacts the clinician for a consultation; rather, the court or an attorney is generally the referring agent, and it is this entity to whom the mental health

professional owes certain obligations. These obligations may need to be discussed and explained in detail with the individual whose competence is being evaluated to ensure that this person does not mistakenly view the clinician as “my doctor” or as having a primarily therapeutic purpose for conducting the assessment. Although there may be some legitimate exceptions, generally mental health professionals should decline to conduct evaluations of legal competency on individuals with whom they have a preexisting therapeutic relationship.

The rules that govern privacy and confidentiality are different in forensic assessments, and these must also be carefully explained to the individual, who may otherwise assume that the usual doctor–patient relationship rules govern subsequent disclosures of information gathered during the assessment.

Another ethical consideration, and one perhaps easily overlooked by forensically naive practitioners, is that of scope of practice. As the foregoing discussion indicates, assessments of legal competence often require an awareness of particular legal tests or criteria, as well as competence in the administration of forensic assessment instruments, that may not be part and parcel of standard mental health training. Thus, some additional training and/or supervised experience in evaluating legal competencies may be required before one can claim adequate education, professional preparation, or competence in conducting such evaluations.

#### 4.2. Instrumentation

It should also be clear from the previous discussion that the assessment approaches and formal instrumentation employed to evaluate legal competencies may vary considerably from those routinely used in therapeutically focused evaluations. Although conventional assessment devices (e.g., EEGs, brain imaging equipment, tests of intelligence or neuropsychological functioning, and personality tests) may have some role in forensic evaluations, legal competencies often involve a more narrow and nuanced inquiry concerning particular functional abilities for which traditional tests were neither developed nor normed. Conventional testing will more likely be helpful in illuminating the background question of an existing mental impairment that might constitute a basis for impaired functioning; independent assessment of the relevant psycholegal abilities and functional capacities is often required and may best be approached using forensic assessments designed especially for such applications. Grisso provides perhaps the most comprehensive review of forensic assessment instruments.

A unique feature of some legal competencies, such as competence to waive Miranda, MSO evaluations, and (some) assessment of testamentary capacity, is their retrospective focus. More often than in therapeutically oriented evaluations, mental health professionals will need to obtain and evaluate archival or third-party information that may be helpful in reconstructing the person’s prior mental condition.

#### 4.3. Communicating with Attorneys and Courts

Because evaluation reports from legal competency assessments are written primarily for consumers in the legal community (judges and attorneys), mental health professionals must also learn to write reports that are appropriate for this audience. Perhaps most obvious is the fact that judges and attorneys will not be familiar with the jargon or terms of art that are commonly found in reports written for other mental health colleagues. Clinicians must learn to present clinical material in more of a “plain language” format to ensure that they are understood. Also, as noted previously, clinical diagnoses—diagnostic labels in particular—are less important in forensic contexts. Because of the need to explain whether and how mental impairment affects functional legal capacity, a greater emphasis is needed on description of particular symptomatology and the ways that such symptoms affect competence. Guidelines for communicating with legal professionals can be found in Melton *et al.* and other sources.

#### See Also the Following Articles

Forensic Mental Health Assessment ■ Psychology and the Law, Overview

#### Further Reading

- Bonnie, R. (1990). Dilemmas in administering the death penalty: Conscientious abstention, professional ethics, and the needs of the legal system. *Law and Human Behavior, 14*, 67–90.
- Committee on Ethical Guidelines for Forensic Psychologists. (1991). Specialty guidelines for forensic psychologists. *Law and Human Behavior, 15*, 665.
- Greenberg, S., & Shuman, D. (1997). Irreconcilable conflict between therapeutic and forensic roles. *Professional Psychology: Research and Practice, 28*, 50–57.
- Grisso, T. (2003). *Evaluating competencies: Forensic assessments and instruments* (2nd ed.). New York: Kluwer/Plenum.

- Grisso, T., & Appelbaum, P. S. (1998). *Assessing competence to consent to treatment: A guide for physicians and other health professionals*. New York: Oxford University Press.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer/Plenum.
- Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers*. New York: Guilford.
- Miller, R., & Willner, H. S. (1974). The two-part consent form: Suggestions for promoting free and informed consent. *New England Journal of Medicine*, 290, 964–966.
- Poythress, N. G., Bonnie, R. B., Monahan, J., Otto, R. K., & Hoge, S. K. (2002). *Adjudicative competence: The MacArthur studies*. New York: Kluwer/Plenum.
- Rogers, R. (1997). *The clinical assessment of malingering and deception* (2nd ed.). New York: Guilford.



# Leisure and Work, Relationship between

Howard E. A. Tinsley and Diane J. Tinsley

Mukilteo, Washington, USA

1. Research on Leisure
  2. Emergence of a Psychological Perspective
  3. Attributes of Leisure Experience
  4. Prerequisites of Leisure Experience
  5. Benefits of Leisure Experience
  6. Applications of Psychology of Leisure
  7. Summary
- Further Reading

## GLOSSARY

**benefit production process** The psychodynamic mechanism that intervenes between a leisure experience and the psychological benefit the individual receives from that experience.

**compensation hypothesis** A theory that work activities provide only some of the benefits that most people desire, so they engage in leisure activities that provide the benefits their work activities do not provide.

**dual-process theories** Theories that derive from Wundt's theory of hedonic tone, which postulates that stimulation has both a pleasant and an unpleasant effect, and the individual's evaluation of the stimulus is a function of the relative salience of these pleasant and unpleasant components.

**flow** A transcendent experience postulated by Csikszentmihalyi to occur as a consequence of an optimal match between the demands of an activity and the skills possessed by the participant.

**idle hands theory** A widely held belief, rooted in the Protestant work ethic, that people are happiest and mentally healthiest

when they are busy. This view regards industry as good and idleness as bad. Having unoccupied time is seen as psychologically and behaviorally risky.

**optimal arousal** One of the conditions postulated to be essential for an individual to interpret an experience as leisure. The level of cognitive stimulation (i.e., optimal arousal) that is maximally motivating and satisfying to the brain.

**peak experiences** A transcendent experience postulated by Maslow that is characterized by high levels of the attributes of leisure experience.

**residual/discretionary models** Early definitions of leisure that defined leisure in terms of the time remaining after one's obligations had been satisfied.

**spillover hypothesis** A theory that individuals seek the same benefits from their leisure and work activities.

**transcendent leisure experiences theory** A theory that when four conditions postulated as necessary to experience leisure are satisfied, (i) an individual will experience the activity as leisure, (ii) the experience will be characterized by high levels of the attributes of leisure experience, and (iii) some psychological needs of the individual will be satisfied.

Psychological investigations of leisure behavior have focused on the causes, attributes, and benefits of leisure experiences. The attributes of leisure experience—the subjective perception of leisure from the perspective of the participant—have received much less attention than the other areas. More attention has been given to the conditions necessary for an individual

to interpret an experience as leisure (i.e., the causes of leisure experience). Freedom of choice, intrinsic motivation, optimal arousal, and psychological involvement have been proposed as essential conditions for individuals to interpret an experience as leisure. The largest body of research concerns the benefits of leisure experience. Most work on the benefits of leisure has focused on the psychological needs satisfied by leisure experiences, but attention has been given to the effects of leisure experiences on physical health, mental health, life satisfaction, and personal psychological growth. Psychologists also have given attention to the application of leisure psychology to counseling and recreation resource management.

## 1. RESEARCH ON LEISURE

Only sporadic studies of leisure were undertaken prior to the 1980s. These studies were conducted almost exclusively by psychologists identified with vocational guidance, the forerunner of the modern-day disciplines of vocational psychology and counseling psychology. The most extensive investigation was undertaken under the leadership of E. G. Williamson, a renowned vocational psychologist and one of the intellectual founders of counseling psychology. The influence of vocational psychology is further evident in the narrow focus of this research on leisure activity choice and on the amount of time devoted to various leisure pursuits. A quasi-moral tone is apparent to the contemporary reader in many of these reports; activities such as drinking and playing cards were regarded with less tolerance than is typical in North American society today.

The residual and discretionary definitions of leisure guided leisure research and practice until the 1980s. Both definitions reflect the lay view of leisure as the discretionary time remaining after obligations have been satisfied. Renowned French scholar Joffre Dumazedier gave popular expression to the residual view with his notion of “three eights” (i.e., daily life consists of 8 hours of work, 8 hours of sleep, and 8 hours of leisure). As late as 1980, leisure psychologists, such as Seppo E. Iso-Ahola in his influential book, *The Social Psychology of Leisure and Recreation*, were still defining leisure as the time left after work.

The discretionary model defined leisure as an activity—apart from the obligations of work, family, and society—to which individuals return at will for relaxation, diversion, social participation, broadening their knowledge, or the exercise of their creative

capacity. This definition improved on the residual definition by focusing attention on leisure activity rather than time, by recognizing that an individual has family and social obligations, and by suggesting that leisure involves the exercise of free will.

Two newer models of the relation between work and leisure have received research scrutiny. The spillover hypothesis postulates that individuals seek the same benefits from their leisure and work activities. In contrast, the compensation hypothesis postulates that work activities provide only some of the benefits that most people desire, so they engage in leisure activities that provide the benefits their work activities do not provide.

The compensation model makes better intuitive sense. Few people have an occupation that satisfies all of their important desires. It is only logical, therefore, that people obtain the most fulfilling occupation possible and then attempt to satisfy their remaining needs through family, social, and leisure activities. In this instance, however, common sense leads to a false conclusion. The research evidence provides much stronger support for the spillover hypothesis: People generally receive similar benefits from leisure and work activities. This suggests that people seek to obtain the benefits they value most in all aspects of their lives. For example, people who value interpersonal interactions seek both an occupation and leisure activities that provide this benefit.

The first psychologists to study leisure behavior encountered two serious shortcomings in the existing leisure theory and research. First, the residual and discretionary models defined leisure as a void (i.e., time left over after accounting for other activities). These models were not useful in identifying the essential features of leisure, the role of leisure in the life of the individual, or the effects of leisure on the individual. Second, the theory and research on work and leisure through the 1960s reflected a Western European, male, middle-class perspective. The work and leisure experiences of women, working-class individuals, and individuals from other races and cultural heritages were largely ignored. Psychologists have addressed these issues in creating a uniquely psychological perspective on leisure.

## 2. EMERGENCE OF A PSYCHOLOGICAL PERSPECTIVE

Sporadic investigations of the interaction between individuals and their leisure activities began to appear in the 1960s. The 1974 publication of John Neulinger's

*The Psychology of Leisure* stimulated the birth of a psychological perspective on leisure and influenced the research and conceptual efforts of psychologists for the next two decades. Neulinger defined leisure as a state of mind and identified perceived freedom and intrinsic motivation as necessary preconditions for leisure to occur. He distinguished among six types of activities based on the extent to which the activity was freely chosen and whether participation was motivated by a desire to receive intrinsic rewards, extrinsic rewards, or a combination of intrinsic and extrinsic rewards. Pure leisure activities occur when individuals freely choose intrinsically satisfying activities; pure job activities occur when individuals respond to behavioral constraints to achieve extrinsic rewards. Although leisure and work were considered to be polar opposites by recreation philosophers, Neulinger argued that leisure was more than the mere absence of work. His definition of leisure and work in terms of the conditions under which they occur enabled scholars to investigate ways in which they are similar.

Initial reactions to Neulinger's model were vigorous. His attempt to redefine leisure as a synonym for "phenomenological experience" was criticized for "confusing communication." His suggestion that only intense experiences could be regarded as leisure ignored the diversity of leisure experiences that occur in everyday life. The ensuing debate focused attention on the nature of leisure experiences and the conditions that are necessary for them to occur. Neulinger's emphasis on the individual's subjective experience and his recognition of the possibility that leisure and work might have similar attributes and benefits highlighted phenomena and issues of interest to psychologists. As a consequence, his model was instrumental in shifting attention away from activity participation patterns and time budget studies.

The theory of transcendent leisure experiences by Howard E. A. Tinsley and Diane J. Tinsley expanded Neulinger's work to include optimal arousal and psychological commitment as additional conditions that are necessary for experiencing leisure. Furthermore, this theory postulated seven attributes of leisure experiences and it suggested that leisure experiences were necessary for physical health, mental health, and personal (psychological) growth.

The transcendent leisure experiences theory highlighted the importance of distinguishing between the causes, attributes, and benefits of leisure experiences. For example, if freedom of choice is a necessary condition for experiencing leisure, individuals who are asked to describe their feelings during a leisure experience might say "I feel free." Since feelings of freedom

can be both a necessary prerequisite for leisure experiencing and an attribute of the experience, the failure to distinguish between the causes, attributes, and benefits of leisure experiences leads to confusion. The following sections describe the attributes of leisure experiences, the causes of leisure experiences (i.e., the conditions necessary for leisure experiences to occur), and the benefits of leisure experiences.

### 3. ATTRIBUTES OF LEISURE EXPERIENCE

Leisure experiences vary from intense, life-transforming experiences to the routine experiences that occur on a daily basis. Nevertheless, psychologists have focused on the most intense forms of leisure experience because the attributes of these experiences are more readily identifiable. Five theorists have postulated attributes of leisure experiences. Abraham Maslow's theory of peak experiences is the earliest and most comprehensive theory. Tinsley and Tinsley's transcendent leisure experiences theory was influenced by Maslow's work, but it deals more directly with leisure behavior. Mihaly Csikszentmihalyi proposed a theory of flow experiences that formed the basis for Chester F. McDowell's theory of leisure consciousness. Csikszentmihalyi and McDowell focus on the most intense leisure experience. Kathleen D. Noble's theory of transcendent religious and meditation experiences is narrower in scope, but it identifies some attributes that are not included in the other four models. In contrast to the others, Maslow and Tinsley and Tinsley believe that the attributes of routine leisure experiences are the same as but less intense than the attributes of transcendent leisure experiences.

Several common themes are discernable in these theories. The greatest consensus exists for the following attributes of leisure experiences:

*Unity:* Maslow was the first to propose that the individual and the experience become integrated. Participants report a sense of unification with the world during leisure experiences.

*Absorption:* Participants focus exclusively on the activity and become unaware of or forgetful of themselves.

*Enriched perception:* Continued participation across time leads to a greater awareness of the nuances, details, patterns, and complexities of the experience.

*Positive outcome:* Leisure experiences provide a desirable way of truth finding that can lead to insights into universal truths.

*Passive awareness:* Participants suspend their will and allow the experience to unfold without attempting to organize or control it.

*Loss of inhibition:* Anxiety, defensiveness, restraint, and self-criticism are set aside; the individual feels confident, unique, expressive, and daring.

*Perceptual distortion:* Participants lose track of time and space; hours can seem like minutes.

*Feelings of freedom:* The individual feels free—unconstrained by rules.

*Intrinsic motivation:* Individuals gain enjoyment from the mere act of participating in the activity (i.e., from their experiencing of the pattern and structure of the activity and their use of the skills required by the activity).

#### 4. PREREQUISITES OF LEISURE EXPERIENCE

Leisure psychologists have proposed that perceived freedom, intrinsic motivation, optimal arousal, and psychological commitment are necessary for an individual to experience leisure. “Leisure” (and “license”) derive from *licere*, a Latin word indicating permission, so it is not surprising that leisure theorists view freedom of choice as essential to experience leisure. Social psychologists refer to freedom of choice as the “illusion of control” because the individual’s subjective perception or interpretation of the situation is the critical determinant of the person’s experience. Individuals may perceive themselves as freely choosing to engage in behaviors that appear to external observers to be the result of external constraints. The illusion of control can vary from one of no control to complete control.

##### 4.1. Control and Reactance

Psychologists from many subdisciplines have noted the importance of freedom of choice or the illusion of control. Developmental psychologists cite the desire to develop personal competence as a strong influence on the behavior of the child. Motivation psychologists view the desires for personal autonomy and competence as two of the most powerful motivations underlying human behavior. According to social psychologists, the need to feel in control leads people to perceive themselves as having control over situations even when they do not. Illusions have genuine consequences; innovative research studies have revealed that the illu-

sion of control is important for maintaining mental health and successful functioning.

Threatening an individual’s sense of freedom generates psychological reactance, an aversive feeling that increases the attractiveness of the threatened freedom, stimulates aggressive feelings, and motivates the person to undertake action to restore the lost or threatened options. Thus, the imposition of constraints has a boomerang effect. The strength of reactance is influenced by the strength of the initial expectation of freedom, the importance of the freedom, the amount of freedom lost, and the effects of the loss of freedom. Leisure psychologists believe that reactance influences people’s willingness to accept a substitute activity in place of a lost activity. An individual’s willingness to substitute one leisure activity for another is inversely related to the amount of external pressure to make the switch.

##### 4.2. Intrinsic Motivation

Intrinsic motivation refers to the desire for rewards that are obtained from the act of participation. For example, an individual who desires social contact would find visiting with close friends to be intrinsically rewarding because the very act involves social contacts. Extrinsic motivation refers to the desire for rewards that do not occur automatically as a consequence of performing the activity. Common extrinsic motivations include external coercion, the desire to avoid negative outcomes, and rewards such as money, grades, and recognition.

###### 4.2.1. Two-Factor Theory of Motivation

The distinction between intrinsic and extrinsic motivation received attention in vocational psychology long before it became of interest to leisure scholars. The two-factor theory of work motivation views this distinction as critical to understanding vocational behaviors (e.g., job performance) and outcomes (e.g., job satisfaction). Most job satisfaction inventories provide measures of both intrinsic rewards (e.g., feelings of accomplishment and a sense of creativity) and external rewards (e.g., recognition and working conditions). Job-enrichment efforts typically involve changes in the work environment that increase the intrinsic rewards of the job.

###### 4.2.2. Autonomy and Competence

Psychologists who study human motivation believe autonomy, competence, and relatedness are the three

primary intrinsic motives. Autonomy refers to the desire for self-determination and personal choice. As such, autonomy is related to freedom of choice and the illusion of control, and the loss of autonomy is a cause of reactance. Developmental psychologists define competence motivation as the desire to feel in control of the situation and to exercise control for its own sake. These definitions of autonomy and competence reveal a conceptual overlap with the illusion of control. Additional research and conceptual development are needed to illuminate these constructs and eliminate their conceptual redundancy.

#### **4.2.3. Relatedness**

Relatedness refers to the desire to feel loved, connected to others, and meaningfully involved with the broader social world. As early as the 1950s, vocational psychologists had identified relatedness as the primary dimension distinguishing among occupations. Early theories postulated that an orientation toward people as opposed to an orientation away from people (or an orientation toward things) was a critical dimension distinguishing among occupations. The strength of a person's preference for working with people is one of the primary factors influencing occupational choices. An extensive body of research supports this view. Some leisure psychologists have also suggested that this is the primary dimension that distinguishes among leisure activities. Research has shown that the desires to affiliate with and nurture others are important dimensions distinguishing among leisure activities.

#### **4.2.4. Overjustification**

Providing extrinsic rewards for behaviors that initially are intrinsically rewarding can destroy intrinsic motivation if the extrinsic rewards become too salient (i.e., overjustification). A classic study conducted in 1973 by Lepper *et al.* demonstrated that preschool children who originally chose to play with colorful felt-tip markers during free play were much less willing to play with the markers after an intervening time in which they had been promised and subsequently given a reward for playing with the markers. Leisure theorists believe this is because the extrinsic reward diminished the child's perception of control. They caution against providing extrinsic rewards to encourage individuals (particularly children) to participate in activities that are intrinsically satisfying.

#### **4.2.5. Extrinsic Regulation**

Extrinsic rewards regulate the behavior of individuals, thereby lessening their illusion of control. There are three forms of extrinsic regulation. Extrinsic motivation, the most coercive form of extrinsic regulation, occurs when behavior is controlled by external factors. Interjected regulation occurs when internal pressures not associated with one's sense of personal identity (e.g., feeling obligated to perform the behavior) influence behavior. Integrated regulation occurs when behavior is motivated by the person's self-concept (i.e., when the failure to perform the behavior would have negative implications for the person's self-concept).

### **4.3. Optimal Arousal**

Optimal arousal is a third condition believed to be essential to experience leisure. Psychologists believe that individual differences in sensation seeking, curiosity, and exploratory behaviors are the result of individual differences in optimal arousal level.

#### **4.3.1. Homeostasis and Optimal Arousal**

Cannon observed in 1932 that all living organisms have a natural tendency to maintain a state of internal balance or equilibrium. He called this tendency homeostasis. The maintenance of homeostasis is largely automatic (e.g., we breathe without having to think about how much oxygen we need). Nissen extended this concept in 1954 to include the function of the brain in perceiving and knowing things. He reasoned that there is an optimal level of cognitive stimulation (i.e., optimal arousal) that is maximally motivating and satisfying to the brain.

#### **4.3.2. Wundt's Law of Hedonic Tone**

In 1874, Wundt proposed a general law of hedonic tone to explain how humans interpret stimuli. He theorized that stimulation has both a pleasant and an unpleasant effect. As the strength of a stimulus increases from barely discernable (i.e., neutral) to weak and then to mild, the strength of its pleasantness component increases much more rapidly than the strength of its unpleasantness component. Therefore, the stimulus is perceived as increasingly more pleasant. However, the pleasantness component of each stimulus reaches its maximum while the unpleasantness

component is still relatively weak. Increases in the intensity of the stimulus beyond this point increase only the strength of the unpleasantness component. Wundt's model explains why increases in the intensity of a stimulus (e.g., increasing the tartness of a lemon meringue pie) result in more pleasant sensations up to a point, but further increases beyond that point actually make the sensation less pleasant.

#### **4.3.3. Dual-Process Model**

William James, Daniel Berlyne, Marvin Zuckerman, and Charles Spielberger all have used Wundt's dual-process model to explain optimal arousal. Berlyne suggested that increases in the strangeness, novelty, complexity, and information provided by a weak stimulus will produce increases in curiosity or exploratory behavior. As the stimulus becomes stronger, however, its negative component begins to cause avoidance or withdrawal.

Zuckerman offered a similar explanation of sensation-seeking behavior. Research evidence shows that persons who seek varied, novel, or complex sensory experiences tend to need greater levels of stimulation (i.e., have a higher optimal arousal level) than people who do not value these qualities. Research also has revealed the existence of an inverse relation between sensation seeking and anxiety.

Spielberger proposed a dual-process model in which curiosity and exploratory behavior result from the simultaneous influence of curiosity and anxiety. Both are psychological states that arouse and motivate behavior; both fluctuate as a function of novelty, incongruity, and ambiguity. However, these states are antagonistic drives. Pleasure is optimal when there are strong feelings of pleasantness, curiosity, or arousal accompanied by mild feelings of anxiety or fear. As the negative side effects of anxiety and fear increase, the overall quality of the experience becomes less pleasant.

The presence of both curiosity and anxiety in approximately equal strengths leads to uncertainty and an approach-avoidance conflict. Approach-avoidance conflicts occur when a single goal has both undesirable (i.e., risk) and attractive (i.e., possibility of gain) features. These conflicts often produce vacillation. The individual moves closer to the goal until the avoidance motive becomes stronger than the approach motive and then moves away from the goal until the approach motive once again becomes stronger. The result is that the person vacillates around the point at which the approach and avoidance motivations are equally weighted.

#### **4.4. Commitment**

A sense of psychological involvement (or commitment) is the fourth condition suggested as a prerequisite for leisure experience. The concept of psychological involvement is useful in distinguishing meaningful participation (e.g., watching one's favorite television show) from inconsequential dalliance (e.g., watching television for 5 minutes while waiting for a companion to arrive). "Commitment" and "involvement" are used somewhat interchangeably in the literature and psychologists have not proposed a formal distinction between them.

In 1947, Muzafer Sherif and Hadley Cantril proposed the concept of psychological involvement to explain the process by which individuals relate their personal sense of identity to a social object (e.g., a sports fan's elation when a favorite team wins). Involvement is a psychological state of arousal, interest, or motivation that has personal relevance to the individual. The strength of involvement influences the extent to which a person will seek out and process information about an activity, expend time and energy pursuing the activity, and continue to pursue the activity across time. Involvement has been investigated most extensively in the area of marketing and advertising behavior.

In leisure theory, involvement (commitment) has been described as a psychological attachment to an activity that guides both long- and short-term behavior. Psychological involvement is determined by a person's satisfaction with the leisure activity, investment in the activity (i.e., the things a person would lose if he or she quit the activity), and perception of the potential benefits available from other activities. A person's level of commitment to a leisure activity predicts his or her activity persistence. Committed individuals have more ability to respond positively to any negative aspects of the activity.

### **5. BENEFITS OF LEISURE EXPERIENCE**

Research has examined a wide variety of potential benefits of leisure, including physiological and physical health benefits, mental health benefits, social benefits, economic benefits, environmental benefits, and benefits to the community. Both individuals and social groups obtain benefits from leisure. Individuals obtain benefits such as increased life satisfaction, but this outcome may also contribute to an improvement in the morale of a family, a classroom, a school, and a community. Thus, social

groups may receive second-order benefits that derive from the first-order benefits individuals obtain from their leisure experiences. Social groups may also receive direct benefits, such as when vandalism is reduced or community pride increased as a result of introducing a new leisure program in the community. Determining the benefits of leisure to social groups is more complicated than determining the benefits received by an individual.

### **5.1. The Meaning of Benefit**

A benefit is a beneficial change or the maintenance of a desirable situation (i.e., a positive outcome). Determining that a positive outcome is a benefit of leisure requires evidence that participation in leisure is responsible for the outcome and that the outcome is an improvement over what would have occurred otherwise. This determination involves a series of subjective judgments. First, psychologists must operationally define leisure. Although much leisure research and theory focuses on the nature of the individual's phenomenological experience, research on leisure benefits has not adopted this approach. Another alternative is to define leisure in terms of a particular setting (e.g., in the mountains) or time (e.g., Sunday afternoons). Research on the benefits of leisure typically has investigated the benefits associated with particular leisure activities (e.g., backpacking or knitting).

Next, the value of the outcome (i.e., positive, neutral, or negative) must be determined. This judgment will vary across individuals and social groups. Some people vandalize property, write graffiti, use drugs, and visit prostitutes during leisure. The individuals involved and some small social groups appear to value the consequences of these activities, but society in general interprets most of the consequences of activities such as these as negative outcomes (i.e., costs). Outcomes such as unwanted pregnancies and contracting a sexually transmitted disease are changed conditions that can occur as a result of leisure activity, but society in general and most individuals would not regard these outcomes as benefits. Leisure psychologists have given virtually no attention to the process by which the value of leisure outcomes is determined.

### **5.2. Methodological Problems**

Finally, it is necessary to determine whether an outcome is a consequence of leisure. A critical difficulty encountered when interpreting research on leisure benefits is that almost all the evidence linking leisure to benefits

was obtained using correlational research designs. For example, a typical research finding is that active retirees are more satisfied than passive retirees. Causal inferences cannot be drawn from correlational data.

There are three difficulties inherent in drawing causal inferences from such data. First, the direction of causation is not clear. Perhaps active retirees are more satisfied because they engage in leisure more often, but it is just as plausible to conclude that people who are satisfied with life in general (i.e., not depressed) are more likely to engage in leisure activities. Second, reciprocal causality may exist. Perhaps participation in leisure activities increases retirees' life satisfaction, which causes them to engage in more leisure activities, which in turn further increases their life satisfaction. Another possibility is that a third variable may be the cause. For example, active and inactive retirees may differ in their physical health. Poor health may prevent retirees from participating in leisure activities and detract from their life satisfaction. It should be clear, therefore, that data obtained using correlational research designs do not yield an unambiguous interpretation.

Despite the limitations of correlational research designs, psychologists rely on them almost exclusively when investigating the benefits of leisure. An experimental research design would require psychologists to deprive some individuals of leisure opportunities while enhancing leisure opportunities for others. This would be unethical because of the strong likelihood that leisure has positive, life-enhancing benefits and the possibility that the lack of leisure could harm the individual. Likewise, it would not be ethical for psychologists to manipulate potential leisure benefits, such as life satisfaction, physical health, and morale, to determine if this influenced leisure behavior. For these reasons, it is not possible for psychologists to conduct controlled experiments to identify the benefits of leisure.

### **5.3. Theories of Benefit Production**

Psychologists are unsure of the mechanisms whereby participation in leisure produces the benefits of leisure. Numerous explanations of some aspect of the leisure benefit production process have been offered, but most of these are little more than sketchy outlines of vague ideas. Among these are psychological hedonism theory (psychological well-being requires many small moments of pleasure, which are obtained through leisure), identity affirmation theory (people's interpretation of their own behavior, including their leisure behavior, shapes their personal identities), and buffer and coping theory

(leisure indirectly influences health and well-being by improving one's ability to cope with stressful life experiences).

### **5.3.1. Idle Hands Theory**

A widely held belief is that people are happiest and mentally healthiest when they are busy. This view is rooted in the Protestant work ethic, which regards industry as good and idleness as bad. Having unoccupied time is seen as psychologically and behaviorally risky. For example, community groups frequently express a concern that the absence of recreation facilities increases the likelihood that adolescents will engage in delinquent behavior. They justify recreation facilities as a means of preventing antisocial behavior.

Governmental agencies also cite the idle hands theory to justify their policies. For example, Daniel L. Dustin's analysis of the recreational pastimes provided at Manzanar, a World War II concentration camp for Japanese American internees, revealed that the camp provided a wide variety of recreational pastimes. Most common were art, followed by sports and games, clubs and organizations, and nature appreciation. These opportunities were plentiful because the war relocation board believed that keeping the internees busy would reduce their resistance and help maintain peace in the concentration camp.

There has been very little research on the idle hands theory. One study of approximately 3000 high school students found that those who were bored with their leisure were more likely to engage in risky behaviors, such as drinking, drug use, bulimia, and attempted suicide. Another study of more than 80 older women found that more active women reported lower levels of guilt, sadness, and emotional distress. Both studies used correlational research designs, however, so unambiguous causal interpretations are not possible for the reasons stated previously.

An even more telling criticism is that idle hands theory does not identify the mechanisms that produce leisure benefits. This view replaces an unknown intervening variable with activity, but it does not explain how activity produces positive consequences nor how low levels of activity produce negative consequences. Further conceptual development is necessary before idle hands theory can be considered to be more than a folk adage stimulated by a belief in the Protestant work ethic. Nevertheless, idle hands theory is widely

endorsed by the lay public and government agencies and tacitly assumed by many psychologists.

### **5.3.2. Theory of Transcendent Leisure Experiences**

Tinsley and Tinsley's theory of transcendent leisure experiences integrates need compensation theory, personal growth theory, identity formation and affirmation theory, and buffering and coping theory within the framework of person-environment fit theory. This theory postulates that when the necessary prerequisites are satisfied, participation in leisure activities enables individuals to have leisure experiences. These experiences are characterized by the attributes described earlier and they satisfy some of the individual's psychological needs. For example, intrinsic motivation is a prerequisite for leisure experience, so some of the participant's needs for autonomy, relatedness, and competence will be satisfied whenever a leisure experience occurs. The illusion of control is an attribute of leisure experience, so leisure experiences provide some satisfaction of the individual's needs to feel competent and in control. Leisure experiences are characterized by enriched perception so it is likely that some gratification of the individual's needs for mastery and intellectual stimulation will be obtained. The loss of inhibition that characterizes leisure experiences should reduce an individual's feelings of anxiety, defensiveness, restraint, and self-criticism.

According to the theory of transcendent leisure experiences, the satisfaction of personal needs through leisure experiences enhances an individual's mental health and contributes to improved physical health. These, in turn, increase overall life satisfaction. This process frees psychological resources that otherwise would be needed to satisfy psychosocial needs and maintain health. The result is that the individual has sufficient psychological resources to undertake the self-exploration and personal risks that are necessary for psychological growth. In the terminology of Maslow's theory, leisure experiences enable the individual to strive toward self-actualization.

A goal of Tinsley and associates was to identify the psychological needs satisfied by the full range of leisure activities. Their exhaustive study of psychological needs led to the development of the Paragraphs about Leisure, a research questionnaire that measures the extent to which 44 psychological needs are satisfied by participation in a leisure activity.

Statistical analysis of the data from more than 4000 research participants revealed 11 dimensions that are important in distinguishing among leisure activities. In order of their importance, these dimensions are as follows:

*Exertion:* Activities high on this dimension emphasize vigorous physical activity; activities low on this dimension offer relaxation and stress reduction. This finding stands in marked contrast to occupational research, which has subsumed this dimension under rubrics such as occupational level and given it scant attention.

*Affiliation:* Activities high on this dimension provide opportunities to be with other people; activities low on this dimension satisfy the need to be alone and enjoy solitude. The importance of this dimension is consistent with motivation theorists' conclusion that relatedness is one of the primary intrinsic motives and with vocational psychologists' conclusion that orientation toward people versus orientation toward things is the primary dimension underlying vocational interests.

*Enhancement:* Activities satisfying the need for enhancement give individuals an opportunity to use their talents and develop their skills; activities low on enhancement provide little opportunity for personal development. This and the following dimension are consistent with motivation theorists' conclusion that competence is an important intrinsic motive.

*Self-expression:* Activities high on self-expression emphasize variety and they afford participants the flexibility to try their own unique approaches and ideas. Activities low on self-expression provide an opportunity to follow a familiar, comfortable routine.

*Nurturance:* Like affiliation, activities that satisfy this need offer a chance to be with and enjoy the company of others, but they also provide opportunities to support, encourage, and help others. Activities low on nurturance satisfy individuals' needs to pursue their own self-interests.

*Compensation:* Activities high on compensation provide an opportunity to make up for things that are missing from one's job or other routine activities; activities low on compensation provide pleasures similar to those obtained from other aspects of one's daily routine. As noted previously, the available research provides stronger support for the spillover hypothesis than for the compensation hypothesis.

*Sensibility:* Activities high on sensibility provide intellectual and artistic stimulation; those low on sensibility provide opportunities to reduce cognitive demands.

*Conscientiousness:* Leisure activities high on conscientiousness emphasize responsible behavior and personal restraint; activities low on conscientiousness allow individuals to escape feelings of obligation and duty and to focus on pleasant aspects of the present situation.

*Status:* Activities high on status provide opportunities to receive attention, influence others, and feel important; those low on status allow individuals to avoid attention and the obligations of leadership.

*Challenge:* Activities high on challenge tend to be goal-directed activities that emphasize personal skill and high levels of performance; activities low on challenge emphasize process more than outcome. Csikszentmihalyi identified challenge as the critical determinant of the flow experience. However, research on the benefits of leisure reveals that leisure activities provide a complex array of psychological benefits and that challenge explains relatively little of the variability among activities.

*Hedonism:* Activities high on hedonism provide immediate pleasure without the need for long-term planning or commitment; activities low on hedonism require a long-term commitment or special effort and typically involve long-term goals.

## 6. APPLICATIONS OF PSYCHOLOGY OF LEISURE

### 6.1. Leisure Counseling

In the 1970s, counseling psychologists and recreation professionals began to advocate the development of a leisure counseling specialty. Papers appeared in the leading journals of both disciplines (e.g., *The Counseling Psychologist* and *Journal of Leisure Research*), and several books appeared that emphasized the need for a leisure counseling specialty, explained the potential benefits of the specialty, and elaborated its conceptual underpinnings.

Advocates of leisure counseling were encouraged by the research evidence that shows a relation between leisure activity participation and enhanced physical and mental health. They reasoned that helping clients select leisure activities that were well suited to meet their needs would result in long-term physical and mental health benefits, and they recommended an integrated approach to career and leisure counseling. Furthermore, they argued that knowledge of the substitutability of leisure pursuits would be useful in helping clients respond to the changes that occur as a result

of factors such as physical injury, unemployment, or the aging process.

Robert T. Overs and associates at the Curative Workshop of Milwaukee, Wisconsin, an agency specializing in rehabilitation counseling, developed an innovative leisure counseling program. Their program was designed to address the needs of clients with disabilities and older persons. They developed a series of questionnaires to be used in assessing the needs of their clients, and they included a research component to evaluate the benefits of their interventions. Their books and monographs represent the most detailed and informative description of an effort to implement a leisure counseling program.

In addition to Overs and associates, several practitioners in the United States began to specialize in leisure counseling. Despite the encouraging results reported by Overs and others, however, mainstream counseling psychology and the recreation profession have given leisure counseling scant attention.

## **6.2. Recreation Opportunity Spectrum**

Beverly L. Driver and associates developed the idea of the recreation opportunities spectrum as a means of summarizing information about the leisure benefits available in an outdoor recreation area. For example, a national park might have many discrete areas. One area might provide opportunities for boating, fishing, hiking, and camping. Another area might provide an opportunity for rock climbing and rappelling, and a third area might be set aside for picnicking and interpretative lectures from park rangers. Each area of the park offers a different sample of leisure benefits because different activities occur in each area.

The recreation opportunities spectrum provides an inventory of the potential benefits provided by an outdoor recreation area. This information is intended for the use of management personnel who oversee public areas such as national parks and wilderness areas. It allows administrators to frame decisions about the addition, modification, or discontinuation of facilities and programs in terms of their effects on the array of psychological benefits provided by the area.

## **6.3. Amenity Resource Valuation**

An issue that often confronts management personnel is how to balance the competing claims of stakeholders

to an area. For example, an oil company that wanted to drill for oil in a wilderness area would support its request by promising to add millions of dollars to the economy. When called upon for comment, leisure psychologists and recreation professionals would express concerns that drilling for oil would degrade the quality of the psychological experience for backpackers and other recreational users. In the past, when situations such as this arose the concerns of leisure psychologists and recreation professionals were dismissed summarily. In response to these difficulties, leisure economists and leisure psychologists developed a series of methods for placing a dollar value on leisure experiences.

Three general classes of methods have been developed for valuing leisure amenities. Travel cost methods, the simplest of the approaches, attempt to place a dollar value on recreational experiences by estimating the cost per visit to the site. These methods assume that the user's travel expenditures were made solely for the purpose of using the recreation site. Hedonic pricing methods are based on the notion that the price one would pay for a recreation experience on the open market provides an estimate of the value of an experience obtained in a public area. For example, the cost of renting a canoe from a commercial vendor at a lake in the countryside provides an estimate of the value of canoeing in the boundary waters canoe area. Contingent value methods rely on data obtained using research questionnaires that ask participants how much they are willing to pay for various recreation experiences. Participants' responses are transformed into monetary values that indicate the relative preferences of the respondents for alternative leisure experiences.

## **7. SUMMARY**

In many ways, psychological investigations of leisure have paralleled the study of occupations. This is true not only when viewing the developmental progression of the two disciplines but also when examining the research on the benefits of leisure and work experiences. For example, many of the benefit dimensions identified in research on leisure are similar to those that have been identified in studies of occupational interests, thereby providing a form of convergent validity. However, there is evidence that the relative importance of the benefit dimensions differs. Exertion, a benefit that has received little attention in vocational

psychology, appears to be a more significant dimension than previously recognized.

Important differences are also apparent in the focus of leisure and vocational psychologists. Leisure psychologists have delved into the nature of participants' phenomenological experiences to a greater extent than have vocational psychologists. Although much of the focus has been on extreme (i.e., transcendent) experiences, more is known about the causes and attributes of leisure experiences than is known about work experiences. In contrast, vocational psychology is far more advanced than leisure psychology with regard to the development of practical applications to benefit individuals and social groups.

Almost everyone experiences leisure in some form, and almost everyone enjoys and values leisure. Therefore, leisure represents an important and understudied aspect of life. It seems likely that leisure will be of central importance in the newfound emphasis of the profession on developing a positive psychology of the healthy personality. Leisure behavior provides an amazingly fertile applied laboratory within which to study phenomena of interest to a broad array of sub-disciplines in psychology, including cognitive, social, motivation, human factors, vocational, and counseling psychology.

### See Also the Following Articles

Competence at Work ■ Intrinsic Motivation and Self-Determination

### Further Reading

- Crawford, D. W., Jackson, E. L., & Godbey, G. (1991). A hierarchical model of leisure constraints. *Leisure Sciences, 13*, 309–320.
- Driver, B. L., Brown, P. J., & Peterson, G. L. (Eds.). (1991). *Benefits of leisure*. State College, PA: Venture.
- Iso-Ahola, S. E. (1980). *The social psychology of leisure and recreation*. Dubuque, IA: Brown.
- Jackson, E. L., Crawford, D. W., & Godbey, G. (1993). Negotiation of leisure constraints. *Leisure Sciences, 15*, 1–11.
- Neulinger, J. (1974). *The psychology of leisure*. Springfield, IL: Thomas.
- Peterson, G. L., Driver, B. L., & Gregory, R. (Eds.). (1988). *Amenity resource valuation: Integrating economics with other disciplines*. State College, PA: Venture.
- Tinsley, H. E. A., Hinson, J. A., Tinsley, D. J., & Holt, M. S. (1993). Attributes of leisure and work experiences. *Journal of Counseling Psychology, 40*, 447–455.
- Tinsley, H. E. A., & Tinsley, D. J. (1986). A theory of the attributes, benefits and causes of leisure experience. *Leisure Sciences, 8*, 1–45.



# Lifespan Development and Culture

**Patricia M. Greenfield**  
University of California, Los Angeles, California, USA

**Heidi Keller**  
University of Osnabrueck, Osnabrueck, Germany

**Ashley Maynard**  
University of Hawaii, Manoa, Hawaii, USA

**Lalita Suzuki**  
University of California, Los Angeles, California, USA

1. Conceptual Issues in Cultural Lifespan Psychology
  2. Developmental Pathways in a Multicultural Society: Consequences for Application
  3. Conclusion
- Further Reading

## GLOSSARY

**apprenticeship** A cultural mode of learning where the child (learner) is considered to be a novice who participates in cultural activities as performed by parents (knower) or other experts.

**attachment** The unique emotional bond that develops between an infant and his or her primary caregivers; the internalization of the mode and quality of these special relationships is supposed to influence the development of later relationships.

**collectivism** A value system that maximizes and prioritizes harmony and interdependence of the group, especially the family; other terms that are sometimes used to denote this value framework are sociocentrism and communitarianism.

**cultural conflict** The confrontation of different cultural value systems within an institutional setting (e.g., a school) or a quasi-institutional environment (e.g., a pediatrician's office); in particular, individualist and collectivist value systems

have been described as cultural scripts that often specify mutually exclusive choices for individuals' behavior.

**cultural learning** The processes of verbal interaction, participation in cultural activities, observation, and instruction so as to acquire cultural knowledge.

**individualism** A value system that maximizes and prioritizes individual choice, independence, and technological mastery.

**life stages** Functional historical units during ontogeny that can be characterized by universal developmental tasks such as development of attachments during infancy and becoming a productive worker during adulthood (not everyone is a parent); the optimal solution of the universal tasks depends on the sociocultural environment and, therefore, is context and culture specific.

**parental ethnotheories** Belief systems that specify values with regard to children, parents, and developmental processes; parental ethnotheories are shared among members of cultural communities and set the frame for parental behavior.

**peer relations** The socioemotional bonds among age mates and between younger children and older children; peers are differentially defined in different cultural environments, being mainly unrelated friends in individualist societies and related children (e.g., siblings, cousins) in collectivist societies.

**socialization** The process of growing up in a cultural environment; it stresses the active acquisition of information and knowledge of the developing child in the context of a social relational matrix.

Culture can be perceived as external to the individual in terms of artifacts and institutions; internal to the individual in terms of knowledge, skills, and values; and interactive to the individual in terms of social practices and social interactions. The role of culture in behavior is pervasive and should be considered carefully in any examination of human development over the lifespan. Human development can be understood as the lifelong interactional processes of perceiving and experiencing culture as triggered from the outside, created from the inside, and cocreated with others. Thus, processes of development and processes of culture are inextricably intertwined. Although groups of people share culture and negotiate cultural issues conjointly, at the same time, culture is a subjective individual process.

## **1. CONCEPTUAL ISSUES IN CULTURAL LIFESPAN PSYCHOLOGY**

This section first presents the ecocultural approach as informed by cross-cultural methodology. Then, it presents the sociohistorical approaches that mainly represent the cultural perspective. Finally, it introduces the developmental pathway approach as a new synthesis that incorporates the ecocultural and sociohistorical approaches and takes the whole lifespan into account.

### **1.1. The Ecocultural Approach**

The ecocultural approach emphasizes the role of culture in adapting to the environment and derives from the work of John Berry as well as John and Beatrice Whiting. Berry performed large-scale studies in 21 different sites around the world to analyze how ecocultural information is transformed into styles of perception and cognition. He demonstrated how parameters of the ecological environment (e.g., spatial layout of the landscape) and the mode of subsistence (e.g., farming, foraging) shape the perceptual style of field dependence/independence as the boundedness of perception of patterns to the surrounding visual field.

The Whittings' six-culture study, including research sites from the major parts of the world, was designed on the basis of their ecocultural model linking ecology to modes of subsistence, modes of subsistence to child rearing, child rearing to child personality, and child personality to adult personality. The Whittings' study

examined this developmental trajectory using a set of consistent methods across different cultures.

Recent theoretical formulations of the ecocultural approach have moved research practice away from the cross-cultural transfer of a set of consistent methods to a more cultural approach with methods appropriate for study in each culture. For example, in 1984, Weisner developed a theory in the Whittings' tradition that considers different aspects of the ecocultural model such as subsistence patterns, child-rearing behavior, and assigned roles for men, women, and children—all of which affect behavior. Subsequently, in 1997, Harkness and Super applied the ecocultural approach to their concept of the developmental niche. The developmental niche is defined as the physical and social settings in which children live, the customs of child rearing, and the psychology of caretakers. Their approach emphasizes the importance of parental cultural belief systems as the link between elements of the larger culture to parenting practices and the organization of the daily routine.

Moreover, the ecocultural approach has become synthesized with evolutionary considerations, specifying the course of development conceptualized in the ecocultural model as following an implicit reproductive logic of optimizing an individual's inclusive genetic fitness resulting from his or her own reproductive success as well as from the reproductive success of genetically related individuals.

### **1.2. The Sociohistorical Tradition**

The sociohistorical research tradition in developmental psychology is derived from the work of Vygotsky. It emphasizes that human development is constructed through social interaction, cultural practices, and the internalization of symbolic cultural tools. An example of a cultural tool that is internalized is money; each culture's system of currency influences the development of certain mental strategies for doing mathematics. Cultural practices and tools are developed over long periods of time. Therefore, they have an important historical dimension.

In addition to artifacts taken in isolation, some cultural psychologists, such as Gauvain and Schliemann and colleagues, have explored the cognitive effects of subsistence practices and trades in terms of representation and cognitive operations, respectively. Cultural practices involve everyday experiences and education (both informal and formal). Therefore, practices such as reading and writing, candy selling, and tailoring shape cognitive development. For example, it has been found that carpenters in Brazil have formal operations in the

domain of proportional reasoning that they use for their trade, even though they do not use this level of reasoning when solving some of Piaget and Inhelder's original tasks. These latter tasks use the much less familiar domains of chemistry and physics.

The locus of social interaction is most often between parents and children due to the importance of parents' caring for their young and for transmitting cultural knowledge from generation to generation. However, teacher-child and peer interactions are also important influences on children's development. Some researchers, such as Maynard and Zukow-Goldring, also recently found that sibling interaction is an important socialization force in both technological and nontechnological societies. Social contexts and the interactional patterns within these social dyads and groups influence both cognitive and social development.

Some researchers, such as Bruner, Tomasello and colleagues, and Shweder, have developed some generalized conceptual approaches to cultural psychology. Bruner noted how the structure of language embodies a structuring of human action in terms of agency and intentional actions. The mutual understanding between people of these categorical structures of intentional action is diagnostic of a theory of other minds. It follows that as language develops, so does this mutual understanding of intentional human action. It is the comprehension of other minds that is central to social sharing and, therefore, to culture.

Tomasello and colleagues, in contrast, emphasized the apprenticeship processes, such as imitation and direct teaching, that enable cultural knowledge to be transmitted from generation to generation. Moreover, they emphasized that cultural learning depends on people's empathic understanding of other minds.

Shweder has developed a notion of environmental constraints versus cultural preferences. This distinction integrates an ecocultural approach with the values approach discussed in the next section. Preferences in cultural practices are guided by an individual's cultural value system. However, the nature of the environment provides constraints to which ideal practices must be adapted.

### **1.3. Cultural Pathways through Universal Development: A New Synthesis**

LeVine was the first to argue that developmental goals guide cultural practices of child care and parenting.

He proposed a universal hierarchy of parental goals with survival and health as the foundation, followed by goals relating to economic independence and finally by goals related to the cultural definition of the personality. Systems of developmental goals and values that guide child rearing and socialization for particular parents or groups of parents are called parental ethnotheories of development. Parental ethnotheories can also be thought of as a stage in the lifespan development of cultural values. Two particularly salient models, derived from the industrial and social psychologies of Hofstede, Triandis, and Markus and Kitayama, are individualism and collectivism. These two models are considered to represent different cultural value systems. Each model is adaptive in different ecological environments, thereby incorporating the ecological approach. Each model is actualized through different kinds of social interactions and emphasizes different kinds of social activities and cultural tools, thereby incorporating the sociohistorical approach.

The cultural models are supposed to be linked to the individual's conception of the self in terms of an independent (cf. autonomous, self-contained, and separate) self linked to individualism and an interdependent orientation (cf. allocentrism and relationalism) toward self linked to collectivism. Accordingly, the primary goal of socialization in the independent model is an autonomous, individuated, self-fulfilled person who enters into social relationships and responsibilities by personal choice. The primary goal of socialization in the interdependent model is for the mature person to be embedded in a network of relationships with responsibilities to others. Personal achievements are ideally in the service of a collectivity, most importantly the family. With these cultural goals in mind, many researchers (e.g., Greenfield and Cocking in 1994) have examined the influence of independent and interdependent goals on parenting styles and child development.

The independent pathway can be characterized as follows. Parents and society socialize children to see themselves as predominantly separate from the social context, that is, as bounded individuals who are unitary and stable by definition. The tasks of development are for children to express themselves, their uniqueness, and their own goals. The role of others is to provide reflections for self-evaluation. This model is fostered by, and adapted to, large-scale commercial societies with complex technologies.

The interdependent pathway can be characterized as follows. Parents and society socialize children to

see themselves as connected to the social context and as having flexible and variable selves. The tasks of development are for children to occupy their place in society, to fit in, and to promote the goals of others. The role of others is to provide self-definitions. This model is fostered by, and adapted to, small-scale subsistence societies with simpler technologies.

There are important developmental changes in both of these orientations over the lifespan and over historical time. From a lifespan perspective, the two pathways travel through universal developmental nodes with particular issues that must be negotiated at particular points in each person's lifespan. Each issue presents a particular transformation of the basic value orientations.

For example, a universal node of infancy is attachment, which can be negotiated by maximizing independence (e.g., with a separate crib and baby room as cultural practices) or interdependence (e.g., by cosleeping between mother and baby, by carrying the baby on the caregiver's body). However, Shweder's distinction between preferences and constraints would also call attention to the fact that if a family is living in a small space, ecological constraints might dictate closer sleeping arrangements than their preferred ideal.

Another universal node of development is peer relations. In a culture that prioritizes independence (e.g., mainstream U.S. culture), peer relations focus on separating from the family. In contrast, in cultures that prioritize interdependence, peer relations often focus on relationships with cousins and siblings. Inherent in the pathway approach, with the conception of nodes that organize developmental progress, is the assumption that life trajectories are meaningful, continuous, and coherent. This implies that the early socialization experiences set the stage for the subsequent developmental tasks during the lifespan. After childhood socialization has resulted in an early conception of a more independent or more interdependent self, adolescence constitutes the transition to adulthood. The independent pathway conceives of adolescence as an educational moratorium with the distinct socializing instruction to develop a separate and unique identity. The interdependent pathway conceives of adolescence as a short transition period that is immediately followed by starting one's own family. Accordingly, ages at marriage and birth of the first child differ significantly across these cultural orientations. Adulthood in the interdependent pathway is primarily defined by procreation and parenting. For example, in Cameroonian Nso, a person is considered

an adult only if he or she has children (a badge of interdependence), irrespective of age and expertise. Adulthood in the independent pathway is defined, apart from legal boundaries, mainly through educational attainments and economic independence from the family of origin. Finally, old age differs with regard to the greater integration in the families of offspring and especially the respect and care that is focused on the older generation in the interdependence framework. From the historical perspective, Greenfield in 2000 explored the impact of historically driven ecological change on cultural apprenticeship and cognitive representation. Studying two generations of Zinacantan Mayan families over a period of two decades, she demonstrated that the historical transition from a subsistence ecology to a commercial ecology transformed the apprenticeship from a more interdependent style of cultural learning to a more independent one. At the same time, artisanal products, the fruits of a changing apprenticeship process, moved away from a model of community creativity toward a model of individual self-expression.

Critics of these views often argue that the conception of these two cultural orientations is overly simplistic and does not cover the complexity of human psychology within and across cultures. Although everyone exists as both an individual and a group member, and although this dualism is part of the evolutionary heritage and is a structured part of each social group, each social group also denotes a complex system with its own set of priorities. A system of priorities, by its very nature, includes both kinds of components. It is simply that the independence components dominate in one kind of system, whereas the interdependence components dominate in the other. In addition, individualist systems prefer certain forms of social relationships, whereas collectivist systems prefer others. Logically (although less is known about this), individualist systems prefer certain forms of independence, whereas collectivist systems prefer others. Adding to the complexity, priorities shift through processes of ecocultural change and intercultural contact.

Nonetheless, it is heuristic to envision two idealized systems that form two extremes of ecology and values. Each system is not simply a collection of individuals with their individual traits. Instead, each system constitutes a deep structure paradigm that generates many characteristics in a multitude of domains and developmental periods. Although every culture has individual differences, one must note that these differences play a

larger role and are valued more in individualist cultures. Note too that each orientation has its own mode around which individual differences are organized. The two cultural types are structured systems; they are not simply overlapping distributions on a common scale or two. Finally, cultural traits are, by definition, coconstructed and shared among a defined group of people. Consequently, they are less variable than non-cultural characteristics.

The nature of development as an individual acquisition process precludes the development of uniformity. Thus, even if there is an ideal modal picture of personality in a cultural community, different real pictures of personality constitute the actual spectrum of the population. Nevertheless, there are obvious differences in developmental pathways that can be captured best by these different value orientations. Indeed, the concept of two paradigmatic pathways allows continuities across developmental pathways to be revealed for the first time. In addition, individuals develop genuine profiles from which they experience themselves as consistent and coherent.

## **2. DEVELOPMENTAL PATHWAYS IN A MULTICULTURAL SOCIETY: CONSEQUENCES FOR APPLICATION**

Intergroup contact is a major source of diverse developmental pathways in multicultural societies such as the United States, Canada, and Australia. As conflicting ethnotheories of development come into contact, many different combinations of values and practices arise. Often, the developmental pathway of the dominant majority tends to be taken for granted as normative, or even universal, rather than as cultural. Intergroup contact raises ethnic identity awareness, often bringing the self-identity of racially distinctive groups to the forefront of developmental issues. This generates more complex processes of cultural learning such as those involved in bilingual and bicultural development.

The distinctive patterns of socialization and development manifest in various groups in the United States and other multicultural societies derive, to a great extent, from ancestral values. Therefore, the specific values that ethnic groups maintain must be understood with respect to ancestral homelands. However, the type of power relation between dominant groups and less dominant groups, as well as the

interactions among various groups, also molds cultural development.

Acknowledging cultural influences as a genuine component of development also implies that there is not one (normative) conception of health and psychological well-being. Rather, the definition of healthy development is relative to overarching developmental goals that vary from culture to culture. The following section highlights the implications of incorporating culture into our understanding of developmental problems and health. It introduces two examples. First, it discusses different views on developmental regulation during infancy. Second, it discusses intercultural misunderstandings and their alleviation in the school setting.

### **2.1. Application of Cultural Pathways During Infancy**

As discussed earlier, the early socialization environment reflects cultural values pertaining to highly valued developmental goals within specific contexts. During infancy, developmental contexts and parenting practices that support interdependent developmental goals imply a close union between mother and baby. Cosleeping, nursing on demand, and constant close body contact constitute the invisible foundation of cultural practices with babies. In this context, mothers often meet their infants' needs before such needs are explicitly expressed.

In contrast, in cultures that subscribe to independent developmental goals, developmental contexts and parenting practices during infancy imply separateness and early autonomous functioning. From birth onward, infants are geared to sleep alone and to sleep through the night to establish the circadian rhythm as early as possible. From early infancy, mothers orient their babies to the object world to foster the infants' exploration and to develop attachment to objects that can potentially replace the caregivers (e.g., security blankets, snuggly toys). As a consequence of these different cultural views on closeness and separation, what is regarded as ideal in the interdependent model of the early parent-child relationship (e.g., establishing constant bodily contact between baby and mother) is evaluated as pathological through the cultural lenses of the independent developmental pathway.

For multicultural societies, conflicts between cultural value systems are the inevitable consequence. These conflicts become especially pertinent when both cultural value systems need to interact. For example, the ideology of a society's health care system, including the

expressed views of pediatricians, is usually oriented to the value system of the majority culture.

When the first Turkish families immigrated into Germany during the 1950s, the common practice of Turkish mothers to swaddle their infants for at least a couple of hours per day was strongly discouraged by pediatricians and health care professionals. They claimed that the infants' freedom of movement and their possibilities to explore their environments were largely disrupted by the swaddling; that is, the Germans' independent cultural values did not mirror the interdependent value system of the Turkish immigrants. In this example, the trust of parents in the health care system on which they must rely is deeply affected and constitutes a major risk factor for minority children's development.

Therefore, culturally sensitive developmental counseling and intervention with babies should rely on the assessment of the preferred modes of parenting as related to the cultural values, resources, and constraints that are available to a particular group. Instead of orienting the parents to the cultural ideals of the majority culture (e.g., encouraging working-class Turkish parents to allow their infants more freedom of movement and exploration possibilities), the cultural reality of the parents must be negotiated in terms of meeting the needs of both the infants and the parents.

## **2.2. Applications of Cultural Pathways in Schools**

Another locus of conflict for the interaction of diverse cultural value systems is the school environment, including school-home interaction. Research on Asian American and Latino children in the United States indicates that their families maintain interdependent and collectivist cultural orientations, although the collectivist orientation becomes more restricted in its expression across the generations (e.g., reserved for family celebrations such as weddings). This reduction across the generations results from a contrasting set of priorities in the individualist host culture. These individualist priorities are especially salient in the institution of the school, where they are strongly socialized.

Children and adolescents often experience these divergent value systems as they go between home and school environments in their daily lives. Greenfield, Raeff, and Quiroz analyzed the cultural values of parents, teachers, and children in two schools in Los Angeles. In one study, the sample in one school

consisted of groups of predominantly European American children. As expected, the value systems of parents, teachers, and children were consonant with each other (the expressed values were primarily individualist) and there was no cultural conflict among the three groups.

In the other school, the population consisted entirely of Latino immigrant families from Mexico and Central America. As expected, the value systems of parents, teachers, and children differed markedly from each other. The parents held mainly interdependent or sociocentric values, and the teachers held mainly individualist values. The value systems of the children either were between those of the teachers and the parents or were more similar to those of the teachers. Therefore, the children had to serve incompatible developmental goals at the same time (e.g., being obedient and self-effacing at home while being verbally expressive and self-enhancing at school). This type of conflict can produce emotional distress and alienation from either home or school. Although the teachers in both schools were multiethnic, they were monocultural. Regardless of their ethnic backgrounds, their education and teacher training had made them highly individualist in their teaching practices and educational values.

In a subsequent phase of the research, Greenfield, Rothstein-Fisch, and colleagues developed "Bridging Cultures," a series of teacher training workshops to help teachers understand the two developmental pathways and create a bridge between them for students and their families. In this way, it was hoped that the developmental stress caused by the conflict between the two cultural pathways could be ameliorated. Teachers and researchers found that parents participated more in school activities and that children were less conflicted in their classroom interactions.

Another source of developmental stress may emerge from the peer interactions of children with different value systems. In still another study, Suzuki and colleagues analyzed the reactions of adolescent girls to praise and criticism on school sports teams in Los Angeles. These authors demonstrated that reactions to praise and criticism differed with respect to self-enhancement and self-effacement, as predicted by the ethnic origins of the families. Asian American and Latina players valued self-effacing strategies to a greater extent, whereas African American and European American players valued self-enhancing strategies to a greater extent. Self-effacement serves the collectivist goal of fitting into the group, whereas

self-enhancement serves the individualist goal of standing out from the group. In line with their own value preferences, Asian Americans and Latinas generally evaluated self-enhancement negatively (e.g., as conceit), whereas European Americans and African Americans generally viewed self-effacement negatively (e.g., as hurting the team). In sum, there was a tendency to view as negative violations of one's own cultural model. These negative interpretations of differences may form the roots of cross-cultural misunderstanding and conflict.

During a later phase of the project, an intervention was carried out to try to use the same workshop methodology to help teams and coaches understand the two cultural orientations. The goal was to increase intercultural understanding by introducing the notion of distinct basic value systems. Rather than increasing tolerance and understanding for different points of view, the intervention simply made students more collectivist. Further research is needed to develop means appropriate to the high school level to increase tolerance and understanding of values for the two contrasting pathways of development.

### 3. CONCLUSION

The sociohistorical approach to the role of culture in development emphasizes the importance of social interaction and cultural tools as socializing forces in individual development. The ecocultural approach notes that socialization practices that are typical of a particular culture mold behaviors that are adaptive in particular ecocultural niches. The cultural pathways approach organizes cultural differences into two underlying paradigms that guide universal developmental tasks, such as attachment and peer relations, in distinctly different directions. Each of these pathways is created through cultural processes identified by the sociohistorical approach, and each is adapted to a different type of ecocultural niche. Most important, each pathway provides a cultural lens for viewing and evaluating the social world. Therefore, different cultural frameworks may affect people's relationships with each other, resulting in conflicts in multiethnic societies. Many such conflicts could be avoided if understanding of cultural frames of reference were part of the educational system. In conclusion, the theory of cultural pathways through universal development provides an important theoretical framework for real-world application.

### See Also the Following Articles

Child Development and Culture ■ Family and Culture

### Further Reading

- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (1992). *Cross-cultural psychology*. Cambridge, UK: Cambridge University Press.
- Gauvain, M. (1995). Thinking in niches: Sociocultural influences on cognitive development. *Human Development*, 38, 25–45.
- Greenfield, P. M. (2000). Culture and universals: Integrating social and cognitive development. In L. P. Nucci, G. B. Saxe, & E. Turiel (Eds.), *Culture, thought, and development* (pp. 231–277). Mahwah, NJ: Lawrence Erlbaum.
- Greenfield, P. M., & Cocking, R. R. (1994). *Cross-cultural roots of minority child development*. Hillsdale, NJ: Lawrence Erlbaum.
- Greenfield, P. M., & Suzuki, L. K. (1998). Culture and human development: Implications for parenting, education, pediatrics, and mental health. In I. E. Siegel, & K. A. Renninger (Eds.), *Handbook of child psychology* (5th ed.), Vol. 4: *Child psychology in practice* (pp. 1059–1109). New York: John Wiley.
- Harkness, S., & Super, C. M. (1996). *Parents' cultural belief systems: Their origins, expression, and consequences*. New York: Guilford.
- Keller, H., & Greenfield, P. M. (2000). History and future of development in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, 31, 52–62.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Maynard, A. E. (2002). Cultural teaching: The development of teaching skills in Maya sibling interactions. *Child Development*, 73, 969–982.
- Munroe, R. L., & Munroe, R. H. (1997). A comparative anthropological perspective. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology*, Vol. 1: *Theory and method* (pp. 171–213). Boston: Allyn & Bacon.
- Ochs, E., & Schieffelin, B. B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In R. Shweder, & R. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 276–320). Cambridge, UK: Cambridge University Press.
- Rogoff, B. (1998). Cognition as a collaborative process. In D. Kuhn, & R. S. Siegler (Eds.), *Handbook of child psychology* (5th ed.), Vol. 2: *Cognition, perception, and language* (pp. 679–744). New York: John Wiley.
- Schliemann, A., Carraher, D., & Ceci, S. J. (1997). Everyday cognition. In J. W. Berry, P. R. Dasen, & R. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology*, Vol. 2: *Basic*

- processes and human development* (pp. 177–216). Boston: Allyn & Bacon.
- Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Super, C. M., & Harkness, S. (1997). The cultural structuring of child development. In J. W. Berry, P. R. Dasen, & J. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology*, Vol. 2: *Basic processes and human development* (pp. 1–39). Boston: Allyn & Bacon.
- Weisner, T. S. (1984). Ecocultural niches of middle childhood: A cross-cultural perspective. In W. A. Collins (Ed.), *Development during middle childhood: The years from 6 to 12*. Washington, DC: National Academy Press.
- Zukow-Goldring, P. (2002). Sibling caregiving. In M. Bornstein (Ed.), *Handbook of parenting*, Vol. 3: *Status and social conditions of parenting* (pp. 253–286). Mahwah, NJ: Lawrence Erlbaum.



# Literacy, Improvement of

Robert C. Calfee

University of California, Riverside, California, USA

1. Literacy: Tools for Thinking and Communicating
2. Improving Literacy: Tasks for Applied Psychology
3. Prospects for Improving Literacy  
Further Reading

## GLOSSARY

**academic language** A formal style of language use found in schools, businesses, and other “ceremonial” settings such as courts and hospitals.

**achievement gap** The substantial and continuing difference in test performance between students from lower- and middle-class backgrounds, often indexed by “Free and Reduced Lunch” programs.

**balanced** Programs of reading instruction that aim to combine elements from a variety of more extreme positions.

**curriculum-based measurement** Assessment systems that attempt to directly measure what is being taught on a daily or weekly basis.

**design experiment** A research method that combines elements of case studies and field experiments, including both quantitative and qualitative data and emphasizing collaboration between researchers and practitioners.

**dynamic assessment** Techniques for measuring student learning that allow departures from complete standardization; as with many clinical procedures, the tester uses the informant’s responses to decide on additional questions and probes.

**dyslexia/learning disabilities** Related labels for describing students with reading problems; learning disability is a formal identification used to assign students to special programs supported by federal funding, whereas dyslexia

connotes a medical-like syndrome in which brain dysfunction is presumed to be the source of the problem.

**initiate–respond–evaluate (IRE)** A frequent pattern in classroom discourse during reading lessons whereby the teacher asks a question, a student responds, and the teacher evaluates the response.

**linguistic register** A style of language use or discourse that is dependent on the individual’s background and the demands of a particular situation; in explaining a traffic ticket, a person is likely to use different language when describing the situation to the police officer, to an acquaintance, and to a judge.

**No Child Left Behind** A multifaceted educational program, promulgated in 2001 by the federal government, calling for widespread testing and accountability.

**phonics** A system for describing the relations between letter and sound patterns and a set of instructional techniques for teaching students how to decode and spell words; it is often associated with the “simple view” that reading depends on acquiring phonics skills that will then connect to the student’s oral language.

**student-centered instruction versus direct instruction** The contrast between educational techniques that emphasize student input and engagement in a lesson and techniques that emphasize the teacher’s role in transmitting information.

**whole language** A system for early reading instruction that relies on immersion of students in printed material that will then lead to the natural acquisition of literacy skills without the need for extensive phonics instruction.

Since the emergence of psychology as a science, literacy has been a significant topic theoretically, empirically,

developmentally, and practically. Studies range from word perception to comprehension strategies, from motivational correlates to neurophysiological profiles, and from laboratory studies of college undergraduates to classroom observations of beginning readers. These endeavors often aim to inform basic psychological issues, but possibilities for practical contributions are always close to the surface. “Improving literacy” offers several opportunities for applied psychologists: (a) enhancing performance across the range of tasks in this domain, (b) developing effective methods to support acquisition of skills and knowledge, (c) creating models to undergird instruction and assessment, and (d) incorporating technological inventions that continue to transform the meaning of literacy. In addition, literacy, unlike piloting or driving, is seen as a public “good” that is supported by educational systems and is everyone’s business. Psychologists aiming to understand and improve literacy must also deal with social, cultural, organizational, and policy issues. This article is divided into three sections. The first defines literacy for the current analysis, lays out several problems that call for “improvement,” and sketches the knowledge base bearing on these problems. The second centers on application and examines the psychological contributions to (a) assessment techniques, (b) evaluation of programs and interventions, and (c) individual differences. The final section offers thoughts about the prospects for improving literacy during the next decade given the current knowledge base and the policy environment.

## 1. LITERACY: TOOLS FOR THINKING AND COMMUNICATING

### 1.1. Definition of Literacy

The definition of literacy shapes the problems, claims about what is known about the phenomenon, and strategies for addressing various issues. This section is far more than a philosophical exercise.

Hunt, in a provocative review of DiSessa’s *Changing Minds*, suggested in 2003 that “an individual’s material intelligence is determined by literacy, the ability to use the intellectual tools that the culture provides.” In contrast, reading and writing, viewed as basic skills, may resemble assembly-line routines whereby repetitions with feedback are the primary psychological elements. Hunt’s characterization built on the concept of academic language, that is, an acquired and specialized capacity to use various forms of language for thinking

and communicating. As basic skill, literacy takes shape as associative learning to be acquired in the primary grades, allowing students to move on to serious topics. As academic register, literacy spans the entire course of schooling and beyond, where print serves as a technology for reducing demands on both short- and long-term memory. The Greeks, before refining their alphabet, employed a variety of mnemonic strategies to support and amplify language capacity, all in the service of organizing thoughts and clarifying communication. This article focuses on the years from preschool through adolescence and spans the continuum from basic to academic literacy.

Academic literacy might seem to entail privilege, and it does. The reference is to a formal style of language use that emphasizes explicitness, expository techniques, coherence, and dependability. The style is typical of educational institutions but also appears in business and the professions. Whether writing a term paper, negotiating a contract, or debating a traffic ticket, individuals in modern society are at an advantage when they can call on the strategies and tools associated with this register.

The acquisition of literacy can be viewed through developmental, sociocultural, and educational lenses. Literacy is rooted in language. Virtually all children progress through developmental stages during which they attain increasing competence in the various elements that constitute language—phonology, semantics, syntax, and discourse. These developments reflect children’s sociocultural environment. Much like sponges, children internalize the features of the particular linguistic environments in which they are immersed, leading to variations in linguistic registers or particular styles of use. Children enter schooling with wide variations in their natural registers reflecting their background experiences, but they all possess a full-fledged language system.

A major goal of the educational system is to inculcate students in the academic register. In 1983, Heath offered an exquisite portrait of the encounters of children from varying backgrounds with “school language.” The story described the task of learning about print—letters, books, “reading and writing”—but more fundamentally the theme was about variations in language registers and ways of using language individually and in social settings. This article focuses on the role of schools in supporting the acquisition of literacy within a policy context that struggles with numerous issues, including the definition of goals and values.

## 1.2. Areas Calling for Improvement

What needs to be fixed? At the time of this writing, the answer might seem to be “nearly everything.” The U.S. federal government is in the midst of a major initiative, No Child Left Behind, that holds states, districts, and schools accountable for guaranteeing that by 2010 virtually every student will achieve proficiency in the primary school subjects, most especially reading and writing, with dire consequences if the goal is not met. Leading the list of problems are results from the National Assessment of Educational Progress (NAEP) showing that (a) two-thirds of U.S. students are not proficient in reading and (b) there has been little improvement in literacy achievement levels during the past three decades or so. The claim is that the United States is a nation of illiterates, undermining both economic and social capacity.

The second problem is the achievement gap. The NAEP data show that children from poor families do poorly in school, especially in literacy. The differences are enormous and have not diminished during the past several decades despite substantial investments by the federal government through various entitlement programs.

The third problem, which is more diffuse, centers on individual differences in the capacity to read and write and in facility with language. The labels are school centered: special education, learning disability (more than 80% of students in this category have “reading problems”), and dyslexia. The correlates are demographic: students from underrepresented minorities (for whom English is a second language) and boys more often than girls.

The fourth problem is essentially technological on the surface: What is the best method for promoting the acquisition of literacy? The response clearly depends on the definition of literacy, but one might think that for acquisition of basic skills, it should be possible to find an answer. The First Grade Reading Study, a large, federally sponsored field experiment comparing various reading programs, showed some short-term differences favoring intensive phonics programs. But student learning depended more on teachers within programs than on differences among programs. What began as a technological matter of instructional program design metamorphosed into issues of professional development, classroom practice, and organizational interactions.

The context for all of these problems depends on the interpretation of literacy. This article emphasizes behavioral and cognitive perspectives. But in the background are issues of motivation, social and cultural influences, and political–organizational opportunities and constraints.

## 1.3. The Knowledge Base

During the past several years, the National Academy of Science has conducted several project reports in the areas of language, literacy, and learning to which psychologists have contributed significantly and that provide enormously useful summaries and reviews: Snow and colleagues' 1998 *Preventing Reading Difficulties in Young Children*, the National Reading Panel's 2000 *Teaching Children to Read*, Donovan and Cross's 2002 *Minority Students in Special and Gifted Education*, August and Hakuta's 1997 *Educating Language-Minority Children*, Bransford and colleagues' 2000 *How People Learn*, and Pellegrino and colleagues' 2001 *Knowing What Students Know*. In addition to these resources is the “Rand Report,” that is, Snow's 2002 *Reading for Understanding*. Other fundamental resources appear in several handbook series: the three volumes of the *Handbook of Reading Research*, the *Handbook of Research on Teaching Literacy Through the Communicative and Visual Arts*, and the *Handbook of Research on Teaching the English Language Arts* as well as pertinent chapters in the *Handbook of Educational Psychology* and the four volumes of the *Handbook of Research on Teaching*.

The several thousands of pages in these collections constitute a compendium of extraordinary proportions, spanning nearly a half-century of theory and research in the literacy arena, with critical pointers to primary research documents. Included are references to the activities of several educational research centers sponsored by the federal government to explore the application of basic research to school settings as well as numerous large-scale evaluation projects designed to explore the improvement of literacy.

Numerous themes emerge from this collection, three of which are explored in detail in the following section. Three additional points are mentioned briefly due to the relevance to applied psychology. The first point centers on the role of research psychologists as collaborators with other educators, including curriculum developers, assessment designers, and program evaluators. Not so many years ago, experimental and educational psychologists focused on laboratory studies and quantitative data analysis. During the past few decades, increasing numbers of researchers have “taken to the field,” moving their investigations into classrooms and schools and joining with other educators in the construction and implementation of programs incorporating theoretical concepts. In many instances, theoretical concepts have not arisen from laboratory findings; rather, field settings have provided the context for generating and refining

theoretical concepts, often taking shape as design experiments that combine the rigor of the laboratory with the messiness of the classroom, joining both quantitative and qualitative sources of data.

A second observation centers on the acquisition of literacy, a process spanning the decade from preschool through adolescence and beyond. The literature provides snapshots of this progression but offers few dynamic portrayals of individual trajectories during the shift from novice to expert. The lack of such intensive data is understandable; research funding comes in 3- to 5-year packages and is “results oriented.” The few longitudinal databases available to educators tend to rely on inexpensive and convenient indicators. Much is known about the psychology of skilled reading, that is, the performance of college students in laboratory settings. Perceptual activities in word recognition and sentence comprehension, strategic processes in text comprehension, and “think aloud” protocols during comprehension and composition exemplify domains where substantial work is available regarding expert performance. Less frequent are convincing portraits of the “expert reader/writer” such as those available for mathematicians, aircraft pilots, and chess players. The impediment may be the breadth of the domain encompassed by literacy compared with the other examples. Another set of snapshots comes from the early stages of learning to read, primarily the acquisition of letter–sound skill and knowledge. The bottom line is that research offers perspectives on early and late facets of literacy, but much less is known about the succession of learning in between.

A third point is the challenge of conducting research in school settings, including the task of “scaling up.” In other fields of applied psychology (e.g., human engineering, organizational psychology), tradition connects researchers with practitioners. In education, these liaisons are much looser. The researcher approaches schools for “access to subjects,” which may be “book-ended” by a technical report of the findings presented to the schools at the end of the project. The client relation, in brief, is often rather distant. During the past several years, the federal government has reintroduced the scaling-up concept, encouraging researchers to move a concept or program from experimentation to wide-scale implementation within a fairly short time frame. The researcher-entrepreneur can be found in other domains, including applied psychology, but in education the tasks are extremely complex, clients range from classroom teachers to state superintendents of education, and researchers might lack the skill, knowledge, and incentives to pursue such endeavors.

## **2. IMPROVING LITERACY: TASKS FOR APPLIED PSYCHOLOGY**

Three arenas have particular relevance in the application of psychological theory and research for the improvement of literacy: assessment, intervention, and individual differences. Assessment is fundamental in operationalizing the constructs that define literacy. Intervention covers a broad and complex array of activities, and psychologists are one of several players in this endeavor. Finally, conceptualizing and investigating individual differences has been a specialization since the beginning of modern psychology.

### **2.1. Assessment Concepts and Tools**

“Literacy is whatever literacy tests measure.” This take-off on a well-worn phrase overstates the matter only slightly. Two dimensions frame much of the activity in the assessment field. The first contrast can be characterized by objectives-based versus performance-oriented approaches to capturing information about individuals and groups. The second facet ranges from the classroom-level practice to large-scale tests used for surveys and policy purposes. The two dimensions are not completely independent, but examples can be found in the “four corners” of the matrix.

Across the spectrum, psychological contributions provide the foundation for virtually all assessments. The Standards for Educational and Psychological Testing establish guidelines for policy and practice, laying out fundamental constructs such as validity and reliability, establishing significant practical criteria (e.g., important decisions should rest on multiple sources of evidence), and defining the ethical responsibilities of practitioners.

Within the classroom, assessments can serve several functions such as grading, monitoring progress, and shaping instructional decisions. Assessments may be externally mandated (e.g., the school district requires a writing test), incorporated within instructional materials (e.g., an end-of-unit test), or developed by the teacher. Techniques range from formal activities (e.g., “Close your books and sharpen your pencils; this is a test”) to informal and observational (e.g., the teacher listens to a first-grader’s reading of a sentence).

Across these settings, the objectives–performance contrast offers an important psychological dimension. In objectives-based assessments, learning outcomes are

defined by highly specific behaviors, and instruments are designed to indicate student mastery of each outcome. In 1971, Bloom and colleagues' *Handbook of Formative and Summative Evaluation of Student Learning* was the progenitor of this approach, which is foundational for the basal reading series widely used for reading instruction in the elementary school grades. The tasks include multiple-choice and short-answer activities, "drawing" answers, and so on. The strategy is to specify outcomes as precisely as possible in behavioral terms. In curriculum-based measurement, a recent addition to this approach, fluency is a primary indicator of mastery (e.g., how many words or sentences can a student read aloud in a minute?).

Performance-oriented techniques rely on cognitive (and metacognitive) constructs as the basis for assessment design built around broader domains and with greater emphasis on explanation. The reading-writing portfolio is a typical artifact in this category, relying on judgmental rubrics to quantify complex performances, much as in the Olympics. A related activity emerges in dynamic assessment, where the task varies in response to student responses so as to explore conditions that evoke student success. This end of the spectrum most often serves for writing assessment.

Externally mandated assessments typically take shape as the standardized multiple-choice test, arguably one of the most significant contributions of applied psychology. Reading is the centerpiece of standardized testing in the elementary grades, soundly established in the "vocabulary-comprehension" subtests. These instruments, although not intended for instructional decisions, can be critical for policy and program evaluation. Writing poses a substantial challenge for large-scale assessment due to the difficulty in capturing complex performances in a cost-effective manner. During recent years, however, several projects have explored the possibilities in this field, most notably in the states of Kentucky, Maryland, Vermont, and California. The importance of these projects arises from showing that it is possible to design systems that connect classrooms with large-scale assessments. Unfortunately, all of these projects have been abandoned in response to the No Child Left Behind initiative.

## 2.2. Interventions: Programs and People

This segment is organized around three necessarily selective topics: early reading acquisition, the reading-

writing connection in the later grades, and pedagogical strategies. The first is a long-standing field of applied psychology, the second illustrates the collaboration of psychologists with others, and the third is more generic.

The First-Grade Reading Study mentioned earlier, Project Literacy, and Chall's landmark review all engaged psychologists in research on beginning reading. Most of the centers and laboratories established by the federal government during the 1960s included projects for the improvement of early reading. Much has been learned as a result of these efforts. For instance, it was previously thought that young children with reading problems suffered from visual perceptual deficiencies, but research demonstrates that the primary correlates are phonological. Another example comes from analyses of English spelling, which is often characterized as inherently chaotic. Research demonstrates the coherence of the system when viewed through historical and phonomorphological constructs. A final example is the discovery that, although direct instruction phonics-oriented programs in the primary grades produce higher scores on standardized tests, (a) program effectiveness varies substantially among teachers within programs and (b) the short-term outcomes tend to "poop out" over the grades. Two general observations seem to hold up. First, much has been learned about the early stages of the acquisition of literacy, including substantive ideas about how to deal with the problems laid out at the beginning of this article. Second, a substantial chasm remains in the practical application of these findings within the complexities of the public education system.

The contributions of psychologists to the study of writing are generally less recognized than those to the study of reading, but the former are nonetheless substantial. The integration of reading and writing, and of language and literacy, becomes increasingly significant as students move through the elementary grades into secondary education and beyond and is not so well studied. From a practical perspective, integration is the key to competence in adult literacy tasks, including college work, jobs in business and industry, and the complexities confronting adults in today's society. Schools tend to teach reading comprehension skills in isolation from composition tasks, and researchers tend to follow this model both conceptually and methodologically. The practical reality is the individual who needs to obtain information from a variety of sources, extract (and record) the key ingredients, and transform the results into a package that serves a practical purpose.

Investigations of pedagogy are generic, although reading often provides the context for research on instructional methods. Psychologists have played a major role in these endeavors, with the first three volumes of the *Handbook of Research on Teaching* being edited by psychologists. Two items warrant particular attention from studies of classroom teaching in the elementary grades: (a) the contrast between strategies that rely on teacher-based direct instruction and those that emphasize student involvement and engagement and (b) explorations of understanding the nature of classroom processes and of the role of teacher decision making in these complex activities.

Probably the most clear-cut finding about classroom instruction is that teachers do most of the talking. This is captured in the initiate–respond–evaluate (IRE) cycle, where the teacher asks a question designed to evoke a particular response, a student responds, and the teacher evaluates the result, turning to other students for a correct answer if necessary. In the primary grades, reading instruction also entails a considerable amount of oral reading, where students take turns reading portions of the day’s passage, with IRE events following each student’s performance. In the later elementary grades, students are more likely to read the daily passage silently, followed by discussion, but still following the IRE pattern. These techniques, known collectively as direct instruction and typically associated with an emphasis on objectives-based curriculum design, lend themselves to the preprogrammed materials found in teachers’ manuals.

The contrastive strategy, often labeled student-centered instruction, places the teacher in the role of building an environment that “scaffolds” student involvement in learning tasks. Students often work in small groups on various projects, with the teacher circulating among the groups to monitor and assist. In whole-language programs, which have gained notoriety during the past decade or so but spring from a much longer history, students can select readings from a range of authentic literature (unlike the tailored passages found in basal readers), with occasional whole-class sessions to introduce learning outcomes (phonics objectives but also vocabulary and comprehension) and to allow students to report their responses to readings. The implicit (and sometimes explicit) assumption is that literacy, like language, will emerge as children are immersed in literary artifacts and supported in gaining mastery of the technical elements, including the technology of print. The assumption that the learning of literacy is as natural as the learning of language is not essential to these programs, nor is it particularly well founded.

Research has illuminated the benefits and limitations of the extreme positions, leading to a search for a middle ground, that is, a balanced approach. Although the concept appears to be attractive, the challenge of defining and designing instruction that incorporates widely dissimilar assumptions and techniques is substantial for both program developers and classroom teachers. It requires more than allocating some time to direct instruction and some time to student-centered activities, a mix that might be programmed into instructional materials and manuals. Investigations of expert teachers show that they create a middle ground that is quite different from either extreme. They typically construct a principled plan for the school year that they constantly modify based on continual assessments of student progress.

Studies of classroom processes and teacher decision making over the past half-century did not begin with expert teachers. During the 1960s, the process–product paradigm employed classroom observations and standardized achievement measures to establish correlational patterns that then led to experimental interventions. For example, if the teacher established effective classroom management techniques, more time was available for academic learning, leading to increased student achievement. By the 1980s, attention turned to issues related to how teachers negotiated the complexities of classroom practice, bringing to bear a range of perceptual, cognitive, and decision-making concepts. Psychologists have played a significant role throughout this progression, moving from a predominantly quantitative perspective to one that incorporates qualitative concepts and techniques and that requires significant collaboration with other disciplines. This paradigm is illustrated by studies of the progression of teachers from novice to professional expertise.

### 2.3. Individual Differences

Threading through the previous material has been the notion that people vary, an important theme from the earliest days of applied psychology. For the literacy field, attention to individual differences is important in its own right but also provides guidance in practical issues such as identification and assistance for students with special needs. Again, three topics emerge in the literature.

The first and most obvious correlate linking individuals with literacy achievements, in the United States as well as worldwide, is poverty—the achievement gap. Understanding the mechanisms that underlie this correlation, and the exceptions to the generalization, is

perhaps the greatest challenge confronting the United States during the decade ahead, not only for the nation's well-being but also in its role as world leader. Correlational and observational studies of the problem abound. Missing from the mix are sustained and intensive longitudinal studies that might illuminate the processes by which poverty undermines the capacity of students to realize their full potential and to point toward practical interventions. "Success stories" that stand the test of time are in short supply.

The second arena centers around the collection of constructs referred to as learning disabilities or dyslexias. A complete treatment of this topic would require much more space. "Learning disability" is, practically speaking, a bureaucratic label. Students who meet certain behavioral and achievement criteria are identified for federal funding to provide special assistance. The recent history of these programs is generally problematic. Although early intervention seems most likely to make a difference, students often have to "wait to fail" in order to receive assistance. Evaluations suggest that the programs have limited effectiveness. Most of the clients are boys, behavior is the most frequent referral "flag," and low reading achievement is a correlate in more than 80% of the cases.

The various dyslexias spring from medical models, with frequent references to neurological processes and studies of brain functioning. The latter area has burgeoned during recent years with the development of new technologies such as functional magnetic resonance imaging (fMRI). Practically speaking, the overlap with learning disabilities is quite substantial when it comes to the practical aspects of identification and treatment. Substantial federal investments in research and development in these arenas, primarily under auspices of the National Institute for Child Health and Human Development (NICHD), has supported much of the fundamental research in early literacy reported previously, especially in the neurological and brain-imaging areas.

The third area in this category focuses on linguistic and sociocultural differences among students. As noted earlier, language development mirrors children's linguistic environments. Socioeconomic status is a significant dimension in that children from the middle and upper classes are immersed in the academic register virtually from birth. A related issue centers on the increasing proportion of students for whom English is a second language. The growing Hispanic population presents the greatest impact, especially along the nation's southern border, but immigrants from Asia and Eastern Europe are also part of the picture. The

issues are quite complex, involving socioeconomic status, cultural styles, home and community languages, and correspondence between student and teacher backgrounds and experiences. English language learners, as they are now labeled, perform poorly on standardized tests across the various programs—bilingual, dual immersion, sheltered English—that have been proposed as remedies. Their performance in mathematics achievement is noticeably stronger, suggesting a linguistic basis for the literacy scores.

### 3. PROSPECTS FOR IMPROVING LITERACY

Those reading this article are literacy experts who bring their experiences to bear on the definition of the concept, the nature of the problems, and ideas about how to improve literacy. The research literature by no means speaks with a single voice; indeed, the National Reading Panel Report evoked considerable debate due to the selectivity of the investigations and interpretations. Nonetheless, a few generalizations and projections appear to be warranted.

First, several areas call for advancements in theory, research, and practice, most clearly in the achievement gap but also in the pragmatics of curriculum and instruction represented in the concept of balance as an alternative to the continual pendulum swings.

Second, the knowledge base available for improving literacy achievement is substantial, albeit rather contentious, in scope and variety. By any criterion, however, the practices of literacy instruction are quite distant from this base. Literacy teachers learn how to implement various "methods" for teaching reading and writing but are less likely to manifest expertise in linguistics, child development and learning, rhetorical concepts, and methods of behavioral and social science research. The elementary school teacher typically relies on materials and mandates to shape practice along with the occasional counsel of others in the school when time allows, and that is not often.

Third, the application and evaluation of the results of literacy research will call increasingly for studies on application and evaluation, including analysis of the range of policy environments that affect public education. In fields such as engineering and medicine, practitioners are much more closely connected with the methods and findings of the underlying disciplines than with those of education. Once in the classroom,

the individual teacher has considerable leeway about practice (imagine a similar situation for the doctor in the operating room). Schools are described as “decoupled,” referring to the disconnections among various levels of the system from district to school to classroom. Those engaged in research and development in the literacy arena are often frustrated by this state of affairs. The remedy might be for these individuals, many of whom are psychologists, to include organizational and contextual elements as an integral part of their research and development agendas.

### See Also the Following Articles

Learning ■ Reading, Teaching of ■ Writing, Teaching of

### Further Reading

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- August, D., & Hakuta, K. (1997). *Educating language-minority children*. Washington, DC: National Academy Press.
- Barr, R., Kamil, M., Mosenthal, P., & Pearson, P. D. (Eds.). (1990). *Handbook of research on reading* (2nd ed.). New York: Longman.
- Barton, P. E. (2003). *Parsing the achievement gap: Baselines for tracking progress*. Princeton NJ: Educational Testing Service.
- Berliner, D. C., & Calfee, R. C. (Eds.). (1996). *Handbook of educational psychology*. New York: Macmillan.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (Eds.). (1971). *Handbook of formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, 2(4), 10–141.
- Boring, E. G. (1929). *A history of experimental psychology*. New York: Appleton-Century.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn* (rev. ed.). Washington, DC: National Academy Press.
- Donovan, M. S., & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.
- Flood, J., Jensen, J. M., Lapp, D., & Squire, J. R. (Eds.). (1991). *Handbook of research on teaching the English language arts*. New York: Macmillan.
- Flood, J., Heath, S. B., & Lapp, D. (Eds.). (1997). *Handbook for literacy educators: Research on teaching the communicative and visual arts*. New York: Macmillan.
- Flood, J., Jensen, J. M., Lapp, D., & Squire, J. R. (Eds.). (2002). *Handbook of research on teaching the English language arts* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Foorman, B. R. (Ed.). (2003). *Preventing and remediating reading difficulties: Bringing science to scale*. Baltimore, MD: York Press.
- Gage, N. L. (Ed.). *Handbook of research on teaching*. Chicago: Rand McNally.
- Heath, S. B. (1983). *Ways with words*. New York: Cambridge University Press.
- Hunt, E. (2003). Thinking in the age of the computing machine: Review of A. A. DiSessa, *Changing Minds: Computers, Learning, and Literacy*. *Contemporary Psychology*, 48, 549–551.
- Lidz, C. C., & Elliott, J. G. (Eds.). (2000). *Dynamic assessment: Prevailing models and applications*. Greenwich, CT: JAI.
- Linn, R. L. (Ed.). (1988). *Educational measurement* (3rd ed.). New York: Macmillan.
- National Assessment of Educational Progress. (2004). *The nation's report card: Reading highlights 2003*. Washington, DC: National Center for Education Statistics.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- Ong, W. J. (2002). *Orality and literacy*. New York: Routledge.
- Pearson, P. D. (1984). *Handbook of research in reading*. New York: Longman.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know*. Washington, DC: National Academy Press.
- Phye, G. D. (Ed.). (1997). *Handbook of academic learning*. San Diego: Academic Press.
- Phye, G. D. (Ed.). (1997). *Handbook of classroom assessment*. San Diego: Academic Press.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Shinn, M. R. (Ed.). (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.
- Snow, C. E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Travers, R. M. W. (1973). *Second handbook of research on teaching*. Chicago: Rand McNally.
- Wittrock M. C. (Ed.). (1986). *Third handbook on research on teaching*. New York: Macmillan.



# Marital Therapy

Angèle Fauchier and Gayla Margolin

University of Southern California, Los Angeles, California, USA

1. Introduction
  2. Assessment
  3. Common Considerations across Treatment Approaches
  4. Specific Therapies
  5. Outcome Research
- Further Reading

## GLOSSARY

**behavioral marital therapy** Behaviorally based couple treatment focusing on behavior exchange and training in communication and problem-solving skills.

**brief strategic couple therapy** Family systems-based couple treatment emphasizing a reduction in attempts to solve problems when such attempts may maintain or even worsen problems.

**cognitive-behavioral marital therapy** Couple treatment combining a behavioral marital therapy approach with a focus on partners' cognitions.

**emotion-focused therapy** Couple treatment focusing on resolution of insecure attachment patterns and processing of emotion.

**insight-oriented marital therapy** Couple therapy emphasizing resolution of unconscious maladaptive relationship patterns that arise from previous developmental experiences.

**integrative behavioral couple therapy** Acceptance-based couple treatment combining a behavioral marital therapy approach with a paradoxical reduction in the struggle to change.

**Prevention and Relationship Enhancement Program (PREP)**  
One of several approaches attempting to prevent couple

distress; it is often applied to couples prior to their wedding or early in their marriage.

The term "marital therapy" is something of a misnomer because the same therapies apply for couples who are not married, including those who are cohabiting and those who do not live together. However, some treatments specifically target couples at certain points in the marriage process such as prior to marriage or during the first years of marriage, particularly during the transition to parenthood. Often, treatments during the early stages of marriage focus on prevention. Treatments during subsequent stages of marriage usually aim to ameliorate relationship distress. Because individual psychological disorders, such as depression and anxiety, are major risk factors for relationship distress, several investigators have also evaluated the efficacy of treating individual disorders through couple therapy. When relationship problems are contributing to or maintaining individual disorders, couple therapy can assist in treatment of the individual disorder or can change relationship patterns that influence the individual disorder. The current focus, however, is on couple therapies that address relationship problems. Several couple therapies have been empirically supported, meaning that randomized clinical trials have found couples undergoing the treatments to have superior outcomes to couples not experiencing those treatments. Although several approaches have been shown to benefit distressed couples, important questions remain about whether specific therapies can be matched

to specific couples and whether couple therapy has long-term effectiveness.

## 1. INTRODUCTION

The vast majority of partners report high satisfaction at the beginnings of their marriages. Yet relationship satisfaction tends to decline over the first 10 years of marriage. Nearly half of all U.S. marriages end in divorce, and most couples are likely to have suffered considerable marital distress prior to divorce. Many other couples whose members remain together also experience marital distress. Marital distress is not stable across the life span. Researchers generally report one of two patterns over time: either a steady decline of marital satisfaction or a complex pattern in which satisfaction is high initially, declines during the child-rearing years, and returns to a higher level during the later years. Marital distress is linked to a variety of individual psychological and physical health problems (e.g., depression, substance abuse). Children growing up in families with marital distress, especially high levels of overt marital conflict, are at risk for psychological problems such as aggression and anxiety. Although approximately 30% of marrying couples receive some form of preventive premarital counseling or relationship education, only 10 to 20% of divorcing couples have met with mental health professionals. The most common sources of disagreement in marriage are money, children and other relatives, the division of housework, communication, and sexual intimacy. It should be noted that the specific interventions available to treat individual or couple sexual problems are not presented in this article.

## 2. ASSESSMENT

Before commencing treatment, it is important that the clinician perform a thorough assessment of the couple. Such an assessment can encompass a variety of methods. The clinician virtually always interviews the partners, whether separately or together, about their individual and relationship histories, aspects they would like to change, and areas of strength. Couples can complete standardized questionnaires about various aspects of their relationship and individual functioning. These questionnaires provide a measurement of partners' overall marital satisfaction relative to that of other distressed and nondistressed spouses. The

questionnaires also evaluate the extent and nature of disagreements in specific relationship domains (e.g., household management, sex, religion, personal habits). Some clinicians perform a formal behavioral observation in which the partners spend a short period interacting with each other so that the clinician can observe their communication style. Patterns of conflict escalation and differences in affective expression are important relationship features that can be observed directly. Clinicians who do not conduct a formal observation invariably observe these dimensions of couple interaction throughout the course of the sessions. Finally, some couples are asked to systematically record information reflecting the nature and tone of their everyday interactions on a daily basis between sessions. These data can include tallies of salient relationship behaviors, both positive (e.g., instances of emotional support) and negative (e.g., arguments), and ratings of important relationship dimensions (e.g., feeling understood by the partner).

## 3. COMMON CONSIDERATIONS ACROSS TREATMENT APPROACHES

A defining feature of couple therapy is that both partners are seen together in the therapy sessions. This mode of therapy, known as conjoint therapy, raises unique ethical issues. For example, will the therapist keep individual confidences of the partners or will the therapist share information between the partners? How the therapist manages these issues has important consequences for therapeutic alliances and the development of therapeutic goals. Even if the therapist schedules occasional individual sessions with each partner, an overarching principle of conjoint therapy is to balance the therapeutic relationship with both partners and to guard against benefiting one partner to the detriment of the other. Having both spouses in therapy means that the therapist must plan interventions that take into account the well-being and goals of both individuals, and this is a challenge when the partners' goals are at odds.

A common goal in marital therapy, regardless of the specific theoretical perspective, is to improve partners' communication. This goal generally involves exploring new ways of sending and receiving messages so that communications are clearer, less polarized, and more productive. An important benefit of doing conjoint

therapy is that the therapist can directly observe dimensions of the couple's communication style that contribute to relationship distress. The therapist can observe problematic styles such as withdrawal from emotional or conflict-laden discussions and escalation of hostility. The way in which partners communicate in the therapy sessions reveals recurring relationship patterns and provides opportunities to explore how each partner thinks about and responds to those patterns.

Although most marital therapies are designed to relieve distress and to help partners become more satisfied with their relationship, therapy also can be used to help partners decide whether or not to end their relationship. The therapist must balance the extent to which he or she promotes improvement of the partners' relationship versus acknowledges irreconcilable differences between the partners.

## 4. SPECIFIC THERAPIES

### 4.1. Behavioral Marital Therapy

Behavioral marital therapy, as originally presented by Jacobson and Margolin, grew out of a social learning perspective of marital distress that focuses on the ways in which partners shape each other's behaviors and on couples' skills deficits. Behavioral marital therapy has two major interventions: behavioral exchange and training in communication and problem-solving skills. Behavioral exchange, through a variety of specific strategies, enables partners to engage in a greater number of pleasing and positive interactions. The therapist assists the partners in planning and carrying out mutually enjoyable activities as well as in showing their affection and caring in small everyday actions. An important part of this intervention is getting partners to switch their focus of attention from negative relationship events to positive ones by becoming more mindful of the exchange of pleasing interactions. The second major intervention is training in communication and problem-solving skills. Communication is one of the most common areas of concern and intervention; partners who communicate well are most likely to maintain high levels of satisfaction. The therapist instructs the partners in listening skills as well as in skills for brainstorming solutions to specific, carefully identified problems. The skills training is designed to help partners discuss important problems in a non-blaming fashion and to develop a specific course of action that reflects the objectives and opinions of

both spouses. Concrete plans are made to try the solution on a time-limited basis, to reevaluate the plan at a specific point in time, and (if necessary) to revise and try a modified solution to the problem.

Behavioral marital therapy, with its focus on reducing negative relationship interactions and increasing positive interactions, generally emphasizes the present over the past. One goal of the clinician is to make the partners self-sufficient in solving their own future problems once therapy has ended. Partners who are more attuned to the positive dimensions of their relationship and who have had some success with resolving major conflicts are then more prepared when they are confronted with the future problems that are inevitable in any relationship.

### 4.2. Cognitive–Behavioral Marital Therapy

Cognitive–behavioral marital therapy, as described by Baucom and Epstein, supplements the elements of behavioral marital therapy by additionally focusing on partners' cognitions about one another and the marriage. Each person brings his or her own interpretation to every moment of every interaction. Sometimes, these interpretations are faulty or can lead to relationship distress as well as individual distress. For example, a wife's disappointment in her husband depends not only on the husband's behavior but also on the wife's expectations of the husband. If the wife's expectations of her husband are unreasonably high, she will find him to be disappointing more often. Cognitive–behavioral therapy might address this problem by focusing on a combination of the wife's expectations and the husband's behavior.

Cognitive–behavioral marital therapy is based on the premise that understanding and changing relationship distress must take account of ways in which partners interpret and evaluate their relationship and one another. In addition to focusing on unrealistic ideals of marriage, this mode of therapy also considers discrepancies in the values and standards that spouses bring to marriage. For example, partners often disagree on how much time should be spent together versus in independent activities, whether it is acceptable to maintain friendships with members of the opposite sex after marriage, how special events (e.g., birthdays, anniversaries) are to be celebrated, and who makes decisions about spending money. In cognitive–behavioral marital therapy, as in behavioral marital

therapy, an important goal is helping spouses to pay attention to the positive aspects of their relationship and not just to notice what is wrong with their relationship. Although this form of therapy also considers partners' behaviors and emotions, its unique contribution is the focus on partners' beliefs, that is, where they came from and what are the advantages and disadvantages of maintaining those beliefs.

#### **4.3. Integrative Behavioral Couple Therapy**

Integrative behavioral couple therapy was developed by Jacobson and Christensen to address marital problems that represented serious irreconcilable differences between marital partners and that were not amenable to the change strategies of traditional behavior marital therapy. In integrative behavioral couple therapy, the overriding goal is to increase each partner's acceptance and understanding of the behaviors they would like to see changed in the other partner. Once partners better understand and accept one another and develop a more collaborative approach to their problems, they are more willing to make desired changes. Paradoxically, as each partner "lets go" of the struggle to change the other partner, some changes that could not be achieved before can become possible when both partners accept the situation. Alternatively, for other problems, the therapist's validation of each partner's point of view leads to greater acceptance and greater intimacy, albeit not to specific behavioral change.

An important component of integrative behavioral couple therapy is the formulation of the couple's problems. A couple's problems are understood in terms of thematic issues such as struggles over one partner's desire for closeness and the other's desire for more distance and struggles over which partner has control over a particular area of the relationship. The way in which a couple attempts to solve these differences often polarizes the partners to extreme positions and makes the problems worse. During a feedback session, the therapist shares the formulation about how the partners have become divided over certain issues, with the goal of increasing tolerance of their differences. An important role of the therapist throughout this treatment is to validate each partner's perspective and to highlight how each partner feels hurt and sad, rather than just angry and disappointed, by what has transpired. As initially described in 1981 in Wile's collaborative marital therapy approach, eliciting spouses'

vulnerable emotions facilitates closeness and acceptance in their relationship.

#### **4.4. Brief Strategic Couple Therapy**

Shoham and Rohrbaugh modified family systems approaches to develop brief strategic couple therapy. The primary focus of this model is the identification and disruption of "ironic processes." Ironic processes are the negative patterns that develop when partners' efforts to solve a problem actually maintain or even worsen the situation. The goal of therapy is to identify and then interrupt this cycle by reducing the partners' attempts to solve the problem. Thus, this approach requires spouses to do less of their currently used solution or even to do the opposite of what they had been doing. An important part of therapy is accepting the partners' definitions of the problem but also reframing and reshaping their understanding of the problem so that they will interact differently around the problem situation. Less of the solution should lead to less of the problem, and this in turn will require even less of the solution, and so on.

Although the focus on interrupting couples' own solutions to their problems is similar to integrative behavioral marital therapy, other features of this approach set it apart from previously described models. In this model, therapy is directed to the specific complaints that clients present. No attempts are made to assess or identify problem areas beyond what the partners themselves have specified. This therapy can be done with one spouse so long as that person is motivated toward change. This approach does not require that the partners understand or have insight into why their problematic patterns have developed. This approach is particularly applicable to those partners for whom skills-based approaches reaffirm and replicate the ineffective solutions that the partners already have tried, for example, structuring more interaction between partners when one spouse already has been criticized about withdrawing from the relationship.

#### **4.5. Emotion-Focused Therapy**

As its title suggests, emotion-focused therapy, developed by Johnson and Greenberg, focuses on the emotional experiences of spouses and how these experiences define and maintain interaction patterns. Emotion-focused therapy integrates gestalt/experiential and systems approaches with attachment theory. As an experiential approach, emotion-focused therapy fosters corrective

emotional experiences in therapy sessions. As a systems approach, this model attempts to interrupt repeating negative cycles of interaction. According to this model of therapy, rigid and insecure attachment styles make it difficult for partners to be emotionally open and responsive in marriage, and this in turn creates marital distress. Thus, treatment goals include using the therapy situation to evoke new interactional experiences that create a more secure bond between the partners through reprocessing and restructuring of each partner's inner experiences. When partners develop a sense of security and a better understanding of attachment needs, they can communicate more clearly due to increased empathy and decreased defensiveness.

As contrasted with therapy models that focus on skill building or on relationship beliefs, the emotion-focused therapist helps partners to identify, experience, and express clearer and more poignant emotional needs. These deeper emotional experiences generally reveal vulnerable rather than hostile emotions and lead to a need for connection with the partner. An underlying assumption of this model is that when partners are emotionally engaged with one another, conflict can deescalate and new patterns of interaction can emerge.

#### **4.6. Insight-Oriented Marital Therapy**

Insight-oriented marital therapy, developed by Snyder and more recently called affective reconstruction, is similar to emotion-focused therapy in its focus on unidentified and unexpressed affect as the source of relationship distress. However, this approach asserts that maladaptive relationship patterns arise from previous developmental experiences and exist primarily on an unconscious level. Correction of the current marital difficulties is accomplished through the use of clarification and interpretation to uncover each partner's underlying dynamics. A goal of insight-oriented marital therapy is to reconstruct affective experiences and relationship schemas through exploration of each partner's own relationship history. This treatment is flexible and inclusive and often incorporates traditional behavioral techniques of communication training and behavioral exchange.

#### **4.7. Prevention**

Prevention-based approaches, such as Markman and colleagues' Prevention and Relationship Enhancement

Program (PREP), attempt to improve the quality of the relationship prior to the emergence of problems. Couples can enroll in prevention programs before marriage, early in their marriages, or in long-term committed relationships. Based on research showing that the quality of communication is related to marital distress, skills training in constructive communication, problem solving, and conflict management is a major component of programs such as PREP. Other foci of prevention programs include exploration of each partner's relationship expectations, reinforcement of the partners' commitment, and enhancement of the positive aspects of the relationship. Prevention approaches generally are conducted in groups of couples with a strong educational component through lectures, demonstrations, the practice of new skills through communication exercises, and feedback from the group facilitators. For some couples, the goal of prevention programs is to reduce the likelihood of later distress. For others, the goal is to reverse a downward trend in satisfaction that already has begun.

### **5. OUTCOME RESEARCH**

Most research to evaluate whether couple therapy is effective has compared couples receiving therapy with a control group of couples receiving no therapy. These comparisons have consistently concluded that couple therapy increases satisfaction and decreases distress more than does no treatment. The possibility that another outcome, such as divorce, might be the best outcome for a given couple is generally not addressed in outcome research. The meta-analysis in 1993 by Shadish and colleagues found an effect size of approximately .60, indicating that the average therapy couple has a better outcome than do 70% of controls. Although reviews generally indicate that more than half of couples receiving couple therapy demonstrate reliable improvement, a smaller proportion of couples show improvements by both partners or show movement from distressed to nondistressed status. Long-term follow-up studies have found that most couples maintain gains over the first year but that some couples deteriorate several years following therapy.

Multiple studies have indicated efficacy for the specific types of couple therapy reviewed in this article—behavioral, cognitive-behavioral, integrative behavioral, brief strategic, emotion focused, and insight oriented—as well as for prevention approaches such as PREP. As yet, there is no consistent evidence

that any one of the treatments is better than any other. Although each of these marital therapy treatments seeks to disrupt recurring dysfunctional interaction patterns, the treatments differ in their emphases on behaviors, cognitions, or emotions. Future efforts to match type of treatment to type of presenting problem or to characteristics of the partners might afford higher rates of treatment efficacy.

### See Also the Following Articles

Emotion and Culture ■ Psychological Rehabilitation Therapies

### Further Reading

- Baucom, D. H., & Epstein, N. B. (1990). *Cognitive-behavioral marital therapy*. New York: Brunner/Mazel.
- Baucom, D. H., Shoham, V., Mueser, K. T., Daiuto, A. D., & Stickle, T. R. (1998). Empirically supported couple and family interventions for marital distress and adult mental health problems. *Journal of Consulting and Clinical Psychology, 66*, 53–88.
- Christensen, A., & Heavey, C. L. (1999). Interventions for couples. *Annual Review of Psychology, 50*, 165–190.
- Epstein, N. B., & Baucom, D. H. (2002). *Enhanced cognitive-behavioral therapy for couples: A contextual approach*. Washington, DC: American Psychological Association.
- Jacobson, N. S., & Christensen, A. (1996). *Acceptance and change in couple therapy: A therapist's guide to transforming relationships*. New York: Norton.
- Jacobson, N. S., & Margolin, G. (1979). *Marital therapy: Strategies based on social learning and behavior exchange principles*. New York: Brunner/Mazel.
- Johnson, S. M., & Greenberg, L. S. (1995). The emotionally focused approach to problems in adult attachment. In N. S. Jacobson, & A. S. Gurman (Eds.), *Clinical handbook of couple therapy* (2nd ed., pp. 121–141). New York: Guilford.
- Margolin, G. (1987). Marital therapy: A cognitive-behavioral-affective approach. In N. S. Jacobson (Ed.), *Psychotherapists in clinical practice: Cognitive and behavioral perspectives* (pp. 232–285). New York: Guilford.
- Shadish, W. R., Montgomery, L. M., Wilson, P., Wilson, M. R., Bright, I., & Okwumabua, T. (1993). Effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology, 61*, 992–1002.
- Shoham, V., & Rohrbaugh, M. J. (2002). Brief strategic couple therapy. In A. S. Gurman, & N. S. Jacobson (Eds.), *Clinical handbook of couple therapy* (3rd ed., pp. 5–25). New York: Guilford.
- Snyder, D. K., & Schneider, W. J. (2002). Affective reconstruction: A pluralistic, developmental approach. In A. S. Gurman, & N. S. Jacobson (Eds.), *Clinical handbook of couple therapy* (3rd ed., pp. 151–179). New York: Guilford.
- Wile, D. B. (1981). *Couples therapy: A nontraditional approach*. New York: John Wiley.



# Mathematics, Teaching of

Steven R. Williams

Brigham Young University, Provo, Utah, USA

1. Introduction
  2. Current Views of Mathematical Understanding
  3. Applications of Cognitive Psychology to Teaching Mathematics
  4. The Discursive Turn in Mathematics Education
  5. Applications of Cultural Psychology to Teaching Mathematics
  6. Conclusion
- Further Reading

## GLOSSARY

**constructivism** An epistemological stance claiming that humans actively construct knowledge in ways that are viable and useful for getting along in the world but that knowledge does not necessarily match an external reality. One educational implication of this stance is that knowledge cannot be transferred from person to person, so that teaching becomes a process of providing experiences that will promote and guide knowledge construction.

**cultural psychology** The study of the role that culture and society play in the mental life of human beings.

**understanding** Learning with understanding refers to the acquisition of a richly connected network of knowledge, including facts, procedures, concepts, and habits of mind that allow for retrieval and application of knowledge in appropriate situations.

The teaching of mathematics refers to actions specifically intended to promote the learning of mathematics, usually in formal classroom settings.

The specific actions taken depend on views of what is fundamental to mathematics itself, what prior knowledge students have, and the environment in which they can best learn mathematical habits of mind.

## 1. INTRODUCTION

The problem of how mathematics can best be taught and learned has been a matter of concern throughout the past century. Facility with mathematics is seen as a fundamental skill and is increasingly important in modern society. At a time when international studies draw attention to performance gaps, the question of improving students' mathematical understanding has great political and social currency. Applying psychological principles in order to solve this problem seems natural. Indeed, modern psychology has been applied to the learning and teaching of mathematics at least since the 1920s, when E. L. Thorndike published *The Psychology of Arithmetic*. Since that time, various psychological theories, including behaviorism, Piaget's genetic epistemology, and current connectionist and cognitive theories, have been applied with varying success to explain and facilitate mathematical learning.

At the same time, such application is not straightforward. The prototypical question, "How might a teacher best help a student to learn X?" that administrators, policymakers, and even parents would naturally ask presupposes common interpretations of what it means to learn X and in what sense an approach might be "best."

That these common interpretations might not be in place is partly a result of shifting views of what it means to know mathematics. It is thus important to know how this issue is currently conceived and how it affects the application of psychology to learning mathematics.

## **2. CURRENT VIEWS OF MATHEMATICAL UNDERSTANDING**

In many ways, the foundational issue in mathematics education is the nature of mathematics itself. The decisions we make about teaching mathematics and the kinds of understanding we value in students both depend on what we believe to be most fundamental about mathematics. Mathematics has traditionally been seen as a collection of facts and procedures to be mastered and later applied to appropriate situations. Certainly, learning mathematics does involve mastery of some basic information—multiplication facts, procedures for solving equations, and so forth. Many early applications of psychology to mathematics learning viewed mathematics in just this way. However, it is increasingly recognized that mathematics is more than a collection of facts and procedures. Knowing mathematics involves having the conceptual understanding and habits of mind that enable one to use mathematics powerfully in daily life. Conceptual understanding of mathematics involves understanding how the facts and procedures fit together and when they apply. It involves knowing why procedures work and how to use them to solve problems. In short, it is connected knowledge that provides an organized framework for understanding. Mathematical habits of mind include the ability to reason logically with mathematics ideas, to justify procedures and answers, and to explain mathematical ideas to others—in short, to engage in the practices of mathematics. The acquisition of all these skills, concepts, and habits of mind is what is meant by learning mathematics with understanding. When this becomes our instructional goal, it significantly affects the psychology of teaching mathematics.

## **3. APPLICATIONS OF COGNITIVE PSYCHOLOGY TO TEACHING MATHEMATICS**

Early attempts to apply psychological theory to mathematics teaching were dominated by associationist and behaviorist models of learning. However, the so-called disaster studies of the early 1970s demonstrated

that children could attain a degree of computational proficiency without any clear idea of what the computations meant, how they should be applied, or how they were related to other knowledge. Such studies pointed to the inadequacy of “black box” models of human cognition for educational purposes. This realization prompted the more careful study of how children thought (as opposed to whether their answers were right or wrong) and a focus on children’s conceptual as well as procedural knowledge. The tools of such study became clinical interviews and the careful observation of children doing mathematics—tools still fundamental to the study of mathematical learning.

As part of this focus on student understanding, and against the behaviorist models, there arose the theoretical stance of constructivism, which has had a profound effect on how mathematics learning is viewed. Grounded in the works of Piaget and explicated by von Glasersfeld and colleagues, constructivism holds that knowledge is actively built up by learners, not in ways that mirror reality but in ways that have viability or fit—ways that help learners make sense of and work in their world. Thus, the understandings built are those that both work for learners and fit with their previous understandings; however, these understandings do not necessarily match an external reality, nor do they necessarily match those of a teacher. One implication is that knowledge cannot be transferred from teacher to student. The teacher can only encourage the construction of viable knowledge by arranging learning situations that support that construction. Teachers who accept constructivist principles focus less on providing students facts and explanations and more on designing the learning environments that help students confront the viability of their knowledge.

Emerging from this background are two critical factors affecting the learning and teaching of mathematics: prior knowledge and mathematical activity. Tasks given to students must connect with students’ prior knowledge in order to be approachable at all. Moreover, they must challenge that understanding so there is a need to construct richer and more useful understanding. Thus, tasks must give rise to reflection on what is known and allow for restructuring of knowledge.

### **3.1. Studies of Early Number Learning**

An example of how these ideas bore fruit is provided by the study of how children develop early arithmetic understanding. In the early 1980s, researchers began to document the strategies children used to solve

addition and subtraction word problems as well as the typical developmental paths from simple to more advanced strategies. These studies demonstrated that children's strategies were strongly affected by the action in the word problems (whether two collections of objects were joined together or whether one collection was separated into two, for example) as well as the role of the unknown quantity in the story (e.g., whether the unknown quantity was the result of joining two collections or was one of the two collections being joined).

These studies showed that although children's understanding of arithmetic differed quite dramatically from adults', their knowledge was robust and very useful. Without formal instruction, children were able to develop modeling and counting strategies that allowed for the solution of a wide number of problems. Later studies have repeatedly shown that students come to school with a wide range of mathematical concepts and informal problem-solving strategies. Similarly, studies of older children's mathematics use outside of school have demonstrated that they can develop quite sophisticated and successful contextual strategies for doing mathematics. Research also demonstrates that unless these informal strategies are taken into account when planning instruction, students' abilities to solve problems actually diminish over time. Rather than developing meaningful and useful problem-solving strategies, students come to see mathematics as a collection of arbitrary rules that they have great difficulty applying to word problems.

In contrast, instruction that builds on informal knowledge and provides tasks that push students to reflect on that knowledge has been shown to be effective. For example, studies have shown that elementary teachers who are given information about students' typical strategies, together with tools for assessing students' thinking, can use it to guide their instructional decisions with beneficial results. Thus, a fundamental tool for teachers is the ability to assess their students' understanding of mathematics. The purpose of their assessment is not merely to discern if their students have achieved mastery but also to know the informal knowledge on which students' further understanding can build. Such teaching, heavily dependent on understanding individual students' concepts and strategies, has become an important focus for mathematics educators, with the result that capturing students' mathematical understandings has become a critical issue for both researchers and teachers.

The studies of children's solutions to word problems revealed that the knowledge required to solve word problems did not comprise a set of isolated skills but was indeed a connected network of understandings. These include mathematical ideas associated with

number, counting, and order, as well as understanding of problem situations, story structure, and how strategies and problem types were connected. The research also showed that as children developed more advanced solution strategies, their knowledge became even more richly connected and flexible. Thus, "learning with understanding" results in richly connected knowledge structures. Such rich knowledge structures have become the goal of instruction. Because of this, the problem of capturing and describing them becomes critical.

In an attempt to capture students' knowledge, a number of psychological constructs and methods have been employed by mathematics education researchers. These began, perhaps, with Piaget's early structural approach and the approaches of Bruner and Dienes, all of which had a profound effect on early thinking in the field. In the 1980s, information processing psychology became popular as a way to find, describe, and categorize the schemata used by students in the production of mathematical knowledge. Researchers attempted to develop production systems that modeled students' actual performance on mathematical tasks as a means of putting forward rigorous hypotheses about these schemata. Recently, researchers in mathematical thinking have been turning to other areas of cognitive science in order to capture both knowledge structures and the means by which they are created and updated. English and Halford emphasize the importance of analogical processes to learning. Another promising approach is the application of theories of embodied cognition to mathematics learning.

### 3.2. Summary

The tools of cognitive psychology are critical to the capturing of student's knowledge, which in turn is important to understand both where students are and where we would like them to be. Unfortunately, there has been less success with finding applications of cognitive psychology that help us understand how our students move from current understanding to richer understanding, or of the situations we must create to facilitate this move. It is partly to address this concern that recent work in mathematics education has turned to a focus on social and cultural aspects of learning.

## 4. THE DISCURSIVE TURN IN MATHEMATICS EDUCATION

Although the focus on individual students' knowledge structures and cognitive processes has met with some

success, it has also been subject to some criticism. In general, cognitive psychology has done a better job of capturing static “snapshots” of students’ understanding than in describing the processes of change and learning. Additionally, the cognitive approach often does not capture the full complexity of the classroom, where, after all, much learning and certainly most teaching are thought to occur.

In part because of these criticisms, and in part because of changes in the view of knowledge and learning in general, recent scholarship has focused much more on the social and cultural context of learning mathematics. In some cases, this has taken the form of attending to the ways that social interaction among students and teachers affects the ways that individual understanding is developed in students. A more radical approach calls into question the way we look at mathematical knowledge. From this perspective, understanding mathematics is seen as the ability to participate in particular discourse communities, or communities of practice. This viewpoint has been motivated in part by the recognition that at the highest professional levels of mathematics use, doing mathematics is a social phenomenon with well-established ways to collaborate, share mathematical ideas, and negotiate what counts as mathematically valid. Other influences include the discovery of cultural psychology, as represented in the work of Lev Vygotsky, Michael Cole, and others. All of this has tended to shift attention from mathematics as a purely cognitive activity to include a focus on social practice and cultural acquisition, particularly as played out in classrooms.

In this emerging tradition, the focus shifts from students’ individual understanding of mathematical ideas to careful observation of how teachers and students together construct meanings in a social and cultural context, and how these social meanings are appropriated by the students as cognitive tools. The research tools then become participant observation, ethnographic interviews, and similar techniques aimed at describing the social life of the classroom.

## **5. APPLICATIONS OF CULTURAL PSYCHOLOGY TO TEACHING MATHEMATICS**

A good deal of research has focused on learning mathematics in classroom settings, paying particular attention to the sociocultural aspects of learning. In this

context, the teachers’ responsibilities expand beyond attending to individual students’ cognition and providing appropriate mathematical tasks to include the establishment of norms and expectations for social behavior in the classroom. A number of studies have begun to converge on similar descriptions of effective classrooms.

In such classrooms, teachers do provide appropriate tasks that take into account students’ prior understanding. However, they also seek to establish expectations that students will (i) seek to learn with understanding, rather than memorizing rules without meaning; (ii) try to understand other students’ ideas and solution methods; (iii) explain and justify their thinking to others; (iv) view mistakes as opportunities for learning; and (v) evaluate the usefulness of mathematical ideas based on reason and argument, not solely on the authority of the text or teacher. Such expectations create an environment in which students can safely engage in the mathematical practices of reasoning, justifying, testing conjectures, and communicating and evaluating ideas. In such a classroom, the teacher serves as a representative of the larger mathematical discourse community and can therefore help make needed connections, facilitate the use of established terminology, and provide needed counterexamples or questions as appropriate. In this way, students engage in a sort of “legitimate peripheral participation” in mathematical activity as a way of gaining greater access to the full community of adult users of mathematics. This participation in turn helps students to build the habits of mind that are part of knowing mathematics with true understanding.

## **6. CONCLUSION**

The tools of psychology applied to the problem of teaching mathematics have allowed researchers to gain access to students’ existing knowledge, which in turn allows teachers to build on that knowledge through the selection of appropriate tasks. In addition, careful observations of classroom life, made from the perspective of cultural psychology, have allowed teachers to create classrooms in which students engage in mathematical activity as a means of learning what it means to do mathematics.

### **See Also the Following Articles**

Cultural Psychology ■ Learning ■ Teaching Effectiveness

**Further Reading**

- Carpenter, T. P., Franke, M. L., Jacobs, V., & Fennema, E. (1998). A longitudinal study of intervention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 29, 3–20.
- English, L. D., & Halford, G. S. (1995). *Mathematics education: Models and processes*. Mahwah, NJ: Erlbaum.
- Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. *Journal of Children's Mathematical Behavior*, 1(2), 7–26.
- Fennema, E., & Romberg, T. A. (Eds.). (1999). *Mathematics classrooms that promote understanding*. Mahwah, NJ: Erlbaum.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.
- Kilpatrick, J. (1992). A history of research in mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 3–38). New York: Macmillan.
- Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.
- von Glasersfeld, E. (Ed.). (1991). *Radical constructivism in mathematics education*. Dordrecht, The Netherlands: Kluwer.
- Wood, T., Nelson, B. S., & Warfield, J. (2001). *Beyond classical pedagogy: Teaching elementary school mathematics*. Mahwah, NJ: Erlbaum.



# Measurement and Counseling

Richard M. Lee and Brooke L. Dean

University of Minnesota, Minneapolis, Minnesota, USA

1. Introduction
  2. Brief History of Psychological Measurement
  3. Relevance of Measurement to Counseling Psychology
  4. Counseling Process and Outcome
  5. Concluding Remarks
- Further Reading

## GLOSSARY

**assessment** The application of measurement to achieve a “snapshot” of a particular variable (e.g., behavior, emotion, ability, interest); counseling psychologists are interested in typical performance on the dimension or variable.

**measurement** The technique of assigning a dimensional or quantitative representation to a psychological construct; it is usually an attempt to place a numerical value on a person’s behaviors, thoughts, emotions, or physiological reactions so as to study a construct.

**psychometrics** The field of study devoted to psychological measurement; psychometricians investigate the properties of measures used in behavioral science research.

**reliability** The consistency or dependability of a measuring technique.

**validity** The extent to which a measurement procedure actually measures the variables intended to be manipulated or measured.

As a cornerstone of counseling psychology, the quantitative method of measurement allows for assessing

human behaviors, thoughts, feelings, and attitudes to aid in further developing an individual’s potential for personal growth and understanding.

## 1. INTRODUCTION

Since the age of antiquity, people have sought to differentiate individuals according to abilities so as to select qualified individuals for specific tasks and duties. The earliest record of this type of testing dates back to ancient China (2200 BC) when oral exams were administered to people seeking entrance into the civil service. History is replete with other examples of formal testing on individuals and groups, but these efforts often failed to make reliable and valid predictions. It was not until the 19th century that a breakthrough occurred when psychologists developed quantitative methods to reliably assess individual differences in mental abilities. This quantitative method of assessment, as well as other historical influences that are discussed later, contributed to the emergence of counseling psychology as a specialty field of psychology. In counseling psychology, professionals are trained to assess the interests and abilities of individuals using a variety of methods and to provide reasonable recommendations regarding school, work, and personal life. This article briefly reviews the history of measurement and its relevance to counseling psychology.

## 2. BRIEF HISTORY OF PSYCHOLOGICAL MEASUREMENT

The modern-day measurement of psychological attributes dates back to the late 1800s and early 1900s when some of the early founders of psychology, such as Galton, Pearson, Cattell, Binet, Spearman, Thurstone, and Thorndike, attempted to quantify and measure individual differences in mental ability (what later became known as human intelligence). During this heyday period, numerous statistical procedures that now serve as the foundation for psychological measurement were developed. Two of the most important statistical advances were the correlation method and factor analysis. A correlation refers to the strength of the relationship between a pair of variables, with the understanding that knowledge about one variable may carry information about the other variable (e.g., grade point average [GPA] and Scholastic Aptitude Test [SAT]). Factor analysis is a method of analyzing a set of correlations to discover which variables in the set are correlated with each other but are largely independent of the other variables. Thus, factors represent the underlying processes that hold together each subset of correlations.

Using these new statistical procedures, psychologists were able to develop more reliable and valid tests of mental ability or intelligence. These tests were composed of many subtests, and each subtest consisted of multiple items or questions that were intended to measure a specific ability or attribute. The correlation method and factor analysis allowed psychologists to determine whether the items in one subtest (e.g., vocabulary items) were correlated with each other to a greater extent than they were with items from another subtest (e.g., math items). Psychologists also were able to determine which of these tests and subtests correlated with observable behaviors (e.g., school grades and income level).

In short, the pursuit to accurately measure mental ability or intelligence led to the development of the field of psychological measurement, with its emphasis on the construction of reliable and valid scales and instruments, standardization of test administration, and use of probability statistics to aid in the interpretation of test results. Perhaps more important, the advent of psychological measurement as a scientific method allowed psychologists to empirically investigate individual differences in a wide range of attributes and behaviors (not just intelligence) as well as the

relationship among these variables. Today, the field of psychological measurement can be divided into psychometrics (i.e., the theory and techniques of measurement) and psychological assessment (i.e., the application of measurement).

The first widespread use of psychological assessment occurred during World War I. Psychologists developed a group intelligence test that was used by the U.S. Army to select and train military personnel for various occupations. This type of assessment began an influx of mass testing and personnel selection—two psychological innovations that continue to play a significant role in society today. Shortly thereafter, psychologists developed measures for a wide range of psychological attributes, including people's attitudes, values, interests, abilities, and personality traits. Among the three most widely known today are the Wechsler Intelligence Scales for children, adolescents, and adults (which measure cognitive functioning); the Minnesota Multiphasic Personality Inventory (which measures personality and psychopathology); and the Strong Interest Inventory (which measures occupational interests).

As society became more concerned about the educational and occupational needs of its citizens, psychologists adapted these measurement advances to everyday practice. The end of World War II is a well-known example of how the fields of measurement and counseling became united. After the war, thousands of veterans needed assistance in their adjustment back to civilian life. They were struggling psychologically with the traumas of combat while seeking higher education opportunities and trying to select appropriate occupations and careers. The Veterans Administration hired counseling psychologists to help returning war veterans overcome personal and vocational adjustment difficulties. Counseling psychologists responded by providing guidance and counseling that was informed by the psychological assessment of individual differences in interests and abilities.

## 3. RELEVANCE OF MEASUREMENT TO COUNSELING PSYCHOLOGY

Counseling psychology has its earliest historical roots in the vocational guidance movement at the turn of the 20th century. During this period, the rapid industrialization of the United States was accompanied by a diversity of new occupations and careers available to young adults. Vocational guidance helped people

to make appropriate educational and occupational choices based on their abilities and interests. To assist in this guidance process, counseling psychologists turned to recent developments in psychological measurement and the broader study of individual differences to assess these abilities and interests accurately and to ensure a proper fit between people and their environment. The previous example of counseling psychologists working with returning war veterans is an apt case in point and is one of the earliest examples of the use of measurement in counseling psychology.

Measurement and counseling psychology have, in many ways, developed a mutually beneficial relationship over the years. Perhaps the best example of this reciprocity is the development of person–environment fit theory. The person–environment fit theory is derived from the original “trait and factor” work of Frank Parsons, the founder of the vocational guidance movement during the early 1900s. Parsons construed vocational guidance as a three-step process of learning about one’s interests and abilities (i.e., trait), understanding the requirements of a given occupation (i.e., factor), and finding a reasonable solution to one’s vocational options. E. G. Williamson and John Holland also are credited with advancing the idea that a match or fit between a person’s personality and the work environment is the best predictor of occupational satisfaction and success. Although this idea might not seem revolutionary by today’s standards, it was groundbreaking during the mid-1900s and set the stage for a plethora of psychological research on the measurement of the person and the environment and on the interaction of person and environment to predict behavior and adjustment.

Furthermore, person–environment fit theory has influenced other developments in measurement and counseling psychology. For example, the concept and measurement of aptitude can be credited to counseling psychologists who sought to understand the readiness and potential of individuals and not simply their levels of intelligence. This emphasis on human potential remains a strong influence in counseling psychology, which as a profession focuses on individual differences in normal development and ways in which to encourage growth and overcome barriers in one’s environment. Counseling psychologists also have been at the forefront of multicultural issues in counseling and development. For example, counseling psychologists have contributed to our understanding of test bias in measurement—caused by poorly constructed test items and improper interpretation of test results—and

its societal impact on the occupational and educational opportunities and barriers for women and members of minority groups. The ways in which racial/ethnic minorities adapt to a majority culture environment is another major emphasis in counseling psychology. Counseling psychologists have developed numerous theories and measures of racial/ethnic identity development, and have established standards for multicultural competencies in counseling and research, to assess the impact of minority–majority status on behavior and adjustment.

As suggested, numerous psychological instruments were developed and used worldwide in vocational guidance and counseling, including the Strong Vocational Interest Blank (commonly known as the Strong Interest Inventory) in 1943, the American Council on Education Psychology Examination (forerunner to the SAT) in 1943, and the General Aptitude Test Battery (which actually dates to pre-World War II) in 1968. These tests have been revised over the years; however, these tests, along with more recently developed ones, continue to provide individuals with an empirical basis for making informed choices about which occupations and careers to pursue. Counseling psychologists have developed and used psychological instruments for many purposes beyond vocational and educational needs. Some current examples include the measurement of identity development, self-concepts, leisure interests, personal attitudes and biases, cultural values and expectations, and multicultural counseling competencies. In all of these instances, counseling psychology uses measurement to understand the ways in which the person and the environment interact and thereby affect normal development and adjustment.

#### **4. COUNSELING PROCESS AND OUTCOME**

By using measurement to assess individual differences in psychological attributes and then applying this knowledge to guidance and counseling, psychologists developed what is now known as differential individual diagnosis. Donald Paterson and others who worked at the University of Minnesota’s Employment Stabilization Research Institute during the 1930s are often credited with developing differential individual diagnosis, which involved the use of psychological assessment to form a diagnosis of a client’s employment

potential or presenting problem. By the 1970s, this technique was gradually refined from a “shotgun” approach to a “rifle” approach. In the shotgun approach, psychologists administered a standard battery of tests to every client who was referred for an assessment regardless of the presenting problem. In the rifle approach, psychologists selectively administered tests that would provide the specific information necessary to make an appropriate diagnosis and treatment plan. Today, differential individual diagnosis is a cornerstone technique of all counseling approaches.

A nagging question for psychologists has been whether differential individual diagnosis is very accurate. That is, do these diagnoses actually predict behaviors and outcomes? In 1954, Meehl wrote his seminal book, *Clinical Versus Statistical Prediction*, which compared the accuracy of actuarial tables with clinical judgments. To the surprise of many, Meehl found that actuarial (or statistical) method based on psychological test scores was far better than were clinical judgments (or diagnoses) made by psychologists in predicting behaviors and outcomes.

Meehl’s finding caused quite a controversy in psychology because it seriously questioned the validity of a psychologist’s expertness in making diagnoses. Also, this finding spurred many changes in the ways in which psychologists are trained to use assessment data and make diagnoses. Nowadays, counseling psychology training programs embrace what is referred to as the scientist–practitioner model. This model emphasizes a balanced training in how to conduct counseling and how to conduct empirical research that informs and is informed by the counseling experience. Moreover, the model ensures that psychologists, whether scientists or practitioners, understand how to best formulate, test, and revise hypotheses (or diagnoses) using data generated from psychological assessments and other relevant sources of information (e.g., clinical interviews, medical histories).

Over the years, other questions regarding the effectiveness of counseling have been asked and addressed with the help of refinements in measurement. In his 1966 book, *Revolution in Counseling*, Krumboltz was among the first to ask what is now a popular question in psychology research: Which techniques are most effective with what kinds of clients when used by what kinds of counselors? Counseling researchers have responded to this multifaceted question by focusing mostly on how to best measure and study the process and outcomes of counseling. Researchers have investigated, among many other variables, the

ways in which counselors use social influence and persuasion to change client behavior, the importance of a close working relationship between a counselor and a client to effect behavioral changes, and (more recently) the relevance of minority and cultural factors in the counseling experience. In all of these instances, the research has required the construction and validation of scales and instruments that assess client and counselor attributes as well as the interaction between client and counselor.

## 5. CONCLUDING REMARKS

The role of measurement in counseling psychology helps to distinguish counseling from other helping professions such as psychiatry, social work, and clinical psychology. In contrast to these professions, which tend to operate from a medical model of providing remedial treatment to individuals with more serious psychopathology, counseling psychology uses psychological assessment to identify people’s strengths and assets and to evaluate their adjustment to the environment. Armed with these assessment data, counselors are able to provide education, prevention-based programs, and goal-oriented counseling to help individuals prepare and cope with situational problems that arise in everyday living (e.g., school-to-work transition, unemployment, marital and family conflict).

## See Also the Following Articles

Attitude Measurement ■ Intelligence Assessment  
 ■ International Classification of Diseases (WHO)  
 ■ Measurement and Counseling ■ Personality Assessment  
 ■ Psychological Assessment, Standards and Guidelines for  
 ■ Psychometric Tests

## Further Reading

- Dawis, R. V. (1992). The individual differences tradition in counseling psychology. *Journal of Counseling Psychology*, 39, 7–19.
- Gelso, C., & Fretz, B. (2001). *Counseling psychology* (2nd ed.). Belmont, CA: Thomson Wadsworth.
- Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, 6, 35–45.
- Krumboltz, J. D. (1966). *Revolution in counseling: Implications for behavioral science*. Boston: Houghton Mifflin.
- Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

- Sue, D. W., Arredondo, P., & McDavis, R. J. (1992). Multicultural counseling competencies and standards: A call to the profession. *Journal of Counseling and Development, 70*, 477-486.
- Walsh, W. B., & Betz, N. E. (1995). *Tests and assessment* (3rd ed.). Englewood, NJ: Prentice Hall.
- Whiteley, J. M. (1984). Counseling psychology: A historical perspective. *The Counseling Psychologist, 12*, 7-93.



# Mental Measurement and Culture

Fons J. R. van de Vijver

Tilburg University, Tilburg, The Netherlands

1. Culture and Mental Measurement
2. Bias
3. Equivalence
4. Cultural Factors in the Assessment of Personality and Intelligence  
Further Reading

## GLOSSARY

**absolutism** The viewpoint that holds that cognitive processes are the same across the globe and that emphasizes the context-independent nature of psychological processes.

**bias** A generic name for potential sources of error in cross-cultural measurement; it refers to the presence of nuisance factors in cross-cultural measurement.

**construct bias** Bias that occurs when the construct measured is not identical across groups.

**equivalence** Comparability of test scores obtained in various cultural groups; it involves the question of whether scores obtained in various cultures can be compared in a meaningful way.

**item bias** A generic term for all disturbances at the item level.

**measurement unit equivalence** The type of equivalence that instruments show if their scales have the same units of measurement but a different origin (e.g., the Celsius and Kelvin scales in temperature measurement).

**method bias** A label for all sources of bias emanating from the method and procedure of a study, including factors such as sample incomparability, instrument differences, tester and interviewer effects, and the mode of administration.

**relativism** The viewpoint that holds that intelligence comes into existence in a specific cultural context and that psychological processes and culture are so closely linked that any attempt to compare these across cultures is futile.

**scalar (or full score) equivalence** The only type of equivalence that allows for the conclusion that average scores obtained in two cultures are different or equal; it assumes that identical interval or ratio scales are applicable across cultural groups.

**structural equivalence** When an instrument administered in various cultural groups measures the same construct in all of these groups.

**universalism** The viewpoint according to which basic cognitive processes are universal, although manifestations of these processes may well vary across cultures.

Comparing individuals from various cultures on psychological constructs, such as depression and intelligence, is more involved than comparing individuals from a single cultural group on the same constructs. This article describes the methodological background of such comparisons. It defines two key concepts: bias and equivalence. Bias is a generic term for potential sources of error in cross-cultural measurement due to the presence of nuisance factors. Equivalence refers to the comparability of test scores obtained in various cultural groups, addressing the question of whether scores obtained in different cultures can be compared in a meaningful way. The article discusses taxonomies of bias and equivalence and uses the assessment of personality and intelligence as an illustration.

## 1. CULTURE AND MENTAL MEASUREMENT

Suppose that we are interested in cross-cultural differences in the nature and levels of depression across cultures. Following Berry and colleagues, three approaches can be distinguished: absolutist, relativist, and universalist. A proponent of an absolutist (or etic) position would start from a validated instrument (usually developed in a Western culture) and would administer the instrument in other cultures. Depression is assumed to be a pan-cultural phenomenon, and instruments that were tried and tested in one culture presumably also work in other cultures. Differences in average scores of various countries on the scale are likely to be interpreted as pointing to country differences in depression. The only challenge to this interpretation would be the occurrence of response sets. For example, if there are strict norms in a culture not to disclose information about one's personal feelings to strangers, a depression inventory asking for reports about these feelings may give a poor rendering of the psychological status of the participant or client.

A proponent of a relativist (or emic) view is much less likely to take cross-cultural score differences at face value, as an absolutist does. Depression, like all psychological phenomena, is always expressed in a specific cultural way. Each culture has its own way of dealing with it, and we cannot expect invariance of symptoms and the accuracy of questionnaires developed in another culture. If we want to understand (and measure) depression, we should try to understand the culture and assume the culture's perspective on depression. Instruments are not assumed to travel well across cultures due to the close link between culture and psychological phenomena.

The intermediate view, called universalist, tries to combine both approaches by emphasizing that basic processes are universal but without ruling out the existence of culture-specific expressions of a particular phenomenon. It is an essential property of the universalist position that the (in)variance of behaviors and attitudes associated with an underlying construct (e.g., symptoms of depression) is not assumed but rather is open to empirical examination; cross-cultural studies should determine to what extent depression is pan-cultural, and questions used in Western cultures are relevant for the measurement of depression elsewhere. A universalist attempts to keep an open mind to all possible sources of cross-cultural differences and to verify or falsify specific interpretations of these differences.

It is proposed that an exclusive focus on either etic or emic measurement will not solve measurement issues in cross-cultural psychology. Instead, both are needed because they address different aspects of the same phenomenon; an etic measure focuses on the pan-cultural aspects, whereas an emic measure focuses on the culture-specific aspects. A universalist approach may offer the most fruitful approach to cross-cultural differences because it is the only approach that is sufficiently flexible to deal with the multiple ways in which cultural factors can affect mental measurement.

A good starting point for appreciating the universalist approach is provided by an overview of bias, a generic name for potential sources of error in cross-cultural measurement. The next two sections of the article discuss bias and equivalence (comparability of test scores); an overview of the concepts is provided in [Table I](#). The subsequent section gives a brief overview of the main findings in the cross-cultural studies of personality and intelligence. Conclusions are drawn in the final section.

## 2. BIAS

Bias refers to the presence of nuisance factors in cross-cultural measurement. If scores are biased, their psychological meaning is dependent on the culture in which they have been obtained. Differences between cultural groups in assessment outcome are due, at least to some extent, to cultural or measurement artifacts. For example, differences in scores on self-esteem obtained by Korean and American participants may be influenced by a norm of modesty, which is stronger in Korea than in the United States, and may be contaminated by social desirability.

### 2.1. Construct Bias

Construct bias occurs when the construct measured is not identical across groups; examples of sources of bias can be found in [Table II](#). Construct bias precludes the cross-cultural measurement of a construct with the same measure. An example can be found in work on filial piety, defined as the psychological characteristics associated with being perceived as a good son or daughter. The Chinese conception, according to which children are supposed to assume the role of caretaker of elderly parents, is broader than the Western concept, which is more restricted to showing love and respect. An

**TABLE I**  
**Definitions of Types of Bias and Equivalence**

<i>Concept</i>	<i>Definition</i>
<i>Bias</i>	Nuisance factors that threaten the comparability of scores
Construct bias	Construct measured is not identical across groups
Method bias	Nuisance factors result from factors such as sample incomparability (sample bias), instrument characteristics (instrument bias), and tester effects and communication problems (administration bias)
Item bias	Nuisance factors are at the item level
<i>Equivalence</i>	Comparability of test scores across cultures
Construct inequivalence	Comparisons lack a shared attribute ("comparing apples and oranges")
Structural or functional equivalence	Instrument measures the same construct in all of the groups studied
Measurement unit equivalence <sup>a</sup>	Measurement scales have the same units of measurement but different origins (e.g., the Celsius and Kelvin scales in temperature measurement)
Scalar (or full score) equivalence <sup>a</sup>	Scores are fully comparable across cultures (same origin and measurement unit in all cultures); this does not imply that means are identical across cultures

<sup>a</sup>Assumes interval- or ratio-level measurement in all cultures studied.

inventory of filial piety based on the Chinese conceptualization will cover aspects unrelated to the concept among Western individuals, whereas a Western-based inventory will leave important Chinese aspects uncovered. Construct bias can also be caused by differential appropriateness of the behaviors associated with the construct in the different cultures. Depression is often taken to consist of somatic aspects (e.g., loss of appetite, sleeplessness) and psychological aspects (e.g., loss of motivation, pessimism). Cross-cultural studies show that the psychological symptoms usually show up as a single cluster, whereas the somatic symptoms do not

always form a single cluster or do not form a cluster that is related to the somatic symptoms.

## 2.2. Method Bias

Method bias is a label for all sources of bias emanating from the method and procedure of a study, including factors such as sample incomparability, instrument differences, tester and interviewer effects, and mode of administration. There are three types of method bias: sample bias, administration bias, and instrument bias. The first type, sample bias, refers to confounding sample differences. Research on the cognitive differences between literate and illiterate individuals, for example, has been plagued by sample bias because a comparison of literates and illiterates is nearly always a comparison between schooled and unschooled persons. So, a study of the influence of literacy almost inevitably becomes a study of the influence of schooling. A few studies have been carried out in cultures where literacy in a local script is transmitted in other than school settings. It is interesting to observe that these studies (which successfully avoided sample bias) found smaller differences between literates and illiterates than did the older studies in which schooled and unschooled individuals were compared. Sample bias tends to increase with the cultural distance between the samples.

Administration bias, the second source of method bias, can be caused by differences in the procedures or mode used to administer an instrument. For example, when interviews are held in participants' homes, physical conditions (e.g., ambient noise, presence of others) are difficult to control. Participants and clients are more prepared to answer sensitive questions in (anonymous) self-completion questionnaires than in the "shared" discourse of an interview. Another source of administration bias is ambiguity in the questionnaire instructions and/or guidelines or a differential application of these instructions (e.g., answers to open questions are considered to be ambiguous and require follow-up questions). The effect of test administrator or interviewer presence on measurement outcomes has been studied empirically. Larger effects have been found in interview studies of attitudes than in standardized testing of cognitive abilities. Deference to the interviewer has been reported; participants were more likely to display positive attitudes to a particular cultural group when they were interviewed by someone from that group. A final source of administration bias is constituted by communication problems between the

TABLE II  
Sources of Bias in Cross-Cultural Assessment

Type of bias	Source of bias
Construct bias	<ul style="list-style-type: none"> <li>• Only partial overlap in the definitions of the construct across cultures</li> <li>• Differential appropriateness of the behaviors associated with the construct (e.g., skills do not belong to the repertoire of one of the cultural groups)</li> <li>• Poor sampling of all relevant behaviors (e.g., short instruments)</li> <li>• Incomplete coverage of all relevant aspects/facets of the construct (e.g., not all relevant domains are sampled)</li> </ul>
Method bias	<ul style="list-style-type: none"> <li>• Incomparability of samples (e.g., caused by differences in education or motivation)<sup>a</sup></li> <li>• Differences in environmental administration conditions, either physical (e.g., recording devices) or social (e.g., class size)<sup>b</sup></li> <li>• Ambiguous instructions for respondents and/or guidelines for administrators<sup>b</sup></li> <li>• Differential expertise of administrators<sup>b</sup></li> <li>• Tester/Interviewer/Observer effects (e.g., halo effects)<sup>b</sup></li> <li>• Communication problems between respondent and tester/interviewer (including interpreter problems and taboo topics)<sup>b</sup></li> <li>• Differential familiarity with stimulus material<sup>c</sup></li> <li>• Differential familiarity with response procedures<sup>c</sup></li> <li>• Differential response styles (e.g., social desirability, extremity scoring, acquiescence)<sup>f</sup></li> </ul>
Item bias	<ul style="list-style-type: none"> <li>• Poor item translation and/or ambiguous items</li> <li>• Nuisance factors (e.g., item may invoke additional traits or abilities)</li> <li>• Cultural specifics (e.g., incidental differences in connotative meaning and/or appropriateness of the item content)</li> </ul>

Source. Adapted from van de Vijver and Tanzer (1997).

<sup>a</sup>Sample bias.

<sup>b</sup>Administration bias.

<sup>c</sup>Instrument bias.

tester and the participant or client. Language problems may be a potent source of bias when, as is not uncommon in cross-cultural studies, a test or an interview is administered in the second or third language of one of the parties involved.

The third source of method bias is instrument bias, which involves general features of an instrument that lead to unintended cross-cultural differences. Individuals from various cultures do not deal in the same way with Likert-type rating scales such as 5- or 7-point scales for assessing agreement. For example, Hispanics have been found to show more extreme scores than have European Americans on a 5-point scale but not on a 10-point scale.

In addition to the construct and the method, items can be the source of bias. Item bias, also known as differential item functioning, is a generic term for all disturbances at the item level. According to a definition that is widely used in psychometrics, an item is biased if respondents who have the same standing on the underlying construct (e.g., equally intelligent), but who

come from different cultures, do not have the same mean score on the item. For instance, if a geography test administered to pupils in Poland and Japan contains the item "What is the capital of Poland?" Polish pupils can be expected to show higher scores on this item than can Japanese students, even when pupils with the same total test score are compared. The item is biased because it favors one cultural group across all test score levels. Several psychometric techniques are available to identify item bias. The most common sources of item bias are poor item translation, ambiguities in the original item, low familiarity or low appropriateness of the item content in certain cultures, and the influence of cultural specifics such as nuisance factors or connotations associated with the item wording. For example, in a translation of the American word "aggression," it is difficult or even impossible to maintain the combined meaning of violence (e.g., "an aggressive predator") and enterprising energy (e.g., "an aggressive salesperson") of the original.

### 3. EQUIVALENCE

Equivalence refers to the comparability of test scores obtained in various cultural groups. It involves the question of whether scores obtained in different cultures can be compared in a meaningful way. Equivalence and bias are related. If scores are unbiased (free from nuisance factors), they are equivalent and can be compared across cultures.

Four different types of equivalence can be distinguished: construct inequivalence, structural equivalence, measurement unit equivalence, and scalar equivalence. Construct inequivalence amounts to “comparing apples and oranges.” Many examples of inequivalent constructs can be found in the clinical literature on culture-bound syndromes. For example, “Latah,” a syndrome that can be found only among some Asian groups, consists of a sudden fright, resulting in uncontrollable imitative behaviors (e.g., verbal repetition of obscenities).

An instrument administered in various cultural groups shows structural equivalence if it measures the same construct in all of these groups. Statistical techniques, notably factor analysis, are employed to examine structural equivalence. If an instrument yields the same factors in various cultural groups, there is strong evidence that the instrument measures the same underlying construct(s). Structural equivalence does not presuppose the use of identical instruments across cultures. For example, a depression measure may be based on partly or entirely different indicators in each cultural group and still show structural equivalence. Structural equivalence has been addressed for various cognitive tests and personality measures.

The third type of equivalence is measurement unit equivalence. Instruments show this type of equivalence if their scales have the same units of measurement but different origins (e.g., the Celsius and Kelvin scales in temperature measurement). This type of equivalence assumes interval- or ratio-level scores (with the same measurement units in each culture). It applies when a source of bias with a fairly uniform influence on the items of an instrument affects test scores of various cultural groups in a differential way. Social desirability and stimulus familiarity may exert this influence. If these factors have a differential influence on the scores obtained in various cultural groups, observed score differences are hard to interpret due to the confounding of valid score differences and measurement artifacts. Empirical research shows that the role of response sets (in personality and attitude measurement) and stimulus

familiarity (in cognitive testing) cannot be neglected. For example, there are country differences in social desirability; more specifically, because social desirability is inversely related to national wealth and educational level, individuals who come from richer countries or who are better educated tend to show lower scores on social desirability.

Only in the case of scalar (or full score) equivalence can direct comparisons be made using statistical tests such as the *t* test and analysis of variance (ANOVA). This is the only type of equivalence that allows for the conclusion that average scores obtained in two cultures are different or equal. Scalar equivalence assumes that identical interval or ratio scales are applicable across cultural groups.

Structural, measurement unit, and scalar equivalence are hierarchically ordered. The third type presupposes the second type, which presupposes the first type. As a consequence, higher levels of equivalence are more difficult to establish. It is easier to verify that an instrument measures the same construct in various cultural groups (structural equivalence) than to identify numerical comparability across cultures (scalar equivalence).

### 4. CULTURAL FACTORS IN THE ASSESSMENT OF PERSONALITY AND INTELLIGENCE

There are various examples of emic models of personality. These do not provide a pan-cultural view of personality but rather try to describe personality in a specific cultural context as accurately as possible. For example, Sow developed a model of African personality in which personality consists of various layers. Important elements in the model are spirituality and the role of ancestors and the extended family. Doi described a Japanese model of personality in which “amae” plays an important role. Amae refers to “indulgent dependency” and the passive love and dependence that find their origin in the love of mother and child and that form the template of relationships later in life. Because these studies are not comparative, bias and equivalence might not seem to be an issue. However, from a slightly more distant perspective, these studies address the bias in Western models because they indirectly show where Western models of personality fall short. It is suggestive that in both models (as well as in other non-Western models of personality), social and relational aspects of personality play a more central role than is common in Western models of personality.

A few Western-based personality inventories have been administered in many countries. A notable one is the Eysenck Personality Inventory. According to Eysenck, personality consists of three aspects: neuroticism, psychoticism, and extraversion. The inventory, used to assess these factors, shows these three factors with remarkable consistency. The factors underlying the instrument tend to be stable, indicating that the structural equivalence of the instrument is good. Another currently popular personality instrument is based on the Five-Factor Model of personality. Extensive research has shown that the five factors in this model that are assumed to underlie personality (Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness) are remarkably consistent across cultures. However, studies of the measurement unit equivalence of the scales tend to produce more ambiguous results. Statistical tests (e.g., confirmatory factor analysis) do not provide clear support for measurement unit equivalence, but it is unclear to what extent this problem is due to the high sensitivity of the statistical tests employed to test the equivalence or to real differences in the measurement unit across cultures.

It can be concluded that cross-cultural research on personality has shown that the Western models can claim a broader generality than their cultural contexts of origin. Structural equivalence is strongly supported. It should be noted that studies have been restricted to schooled populations, so it is still unclear to what extent these results also apply to illiterates. It is still unclear to what extent non-Western theories and models could add other universal factors. Recent studies suggest that the Five-Factor Model of personality may be incomplete. Researchers from Hong Kong found a sixth factor, Interpersonal Relatedness, in China and replicated its existence in a Fijian sample. If new factors emerge and it turns out that Western models are incomplete, it is likely that the new factors will be located in the social and relational domains.

The domain with probably the most widely documented cross-cultural research is intelligence. Standard (Western) intelligence tests have been administered in a wide variety of cultures. In many multicultural societies, these tests have been administered to majority group members and usually at least one of the larger ethnic minority groups. Instruments such as Raven's Progressive Matrices and the Wechsler intelligence scales for adults and children have been found to reveal a remarkably stable factor structure across cultures, supporting the structural equivalence of intelligence as measured by these tests.

Statistical tests of measurement unit and full score equivalence of intelligence test scores tend to produce more ambiguous results. It is often difficult to decide whether equivalence in a given case is scalar equivalence or measurement equivalence. For example, the debate about ethnic differences in cognitive test performance can be largely seen as focusing on the level of equivalence of cross-cultural score comparisons and on the presence of biasing factors (notably method bias). For example, Herrnstein and Murray argued that when appropriate intelligence tests are used, cross-cultural differences in test performance reflect valid intergroup differences and show full scalar equivalence. The implication is that group differences in the intelligence test scores reflect differences in intellectual abilities, whereas group differences reflect mainly (or exclusively) method bias in the reasoning of their critics. Many of these critics argue that common intelligence tests show problems such as differential familiarity and that this method bias will allow only measurement unit equivalence.

Despite the disagreement about the higher levels of equivalence of cognitive test scores, cross-cultural studies show that it is very unlikely that intellectual processes are fundamentally dissimilar in various cultures. The findings are not supportive of theories that hold that different cultures produce their own kinds of intelligence; instead, they are more supportive of theories according to which cultures can shape the same set of human cognitive processes in a differential way. However, recent cross-cultural work points to another possible one-sidedness of Western tests. These tests tend to focus on two aspects of intelligence: reasoning and memory. Implicit theories of intelligence have been investigated in a number of countries by asking participants which characteristics they associate with intelligence. Social aspects are nearly always found to be important, yet social intelligence is not part of all intelligence tests. It is interesting to note that the social domain has also been mentioned in relation to personality research as an area where expansion of Western models might be needed to accommodate non-Western models.

Thus, cultural factors affect mental measurement in various ways. Some of these factors, such as a poor item translation, may be artifactual. Other factors, however, are more basic and may refer to problems with the whole instrument or even with the concept measured by the instrument. It is clear that methodological and substantive considerations cannot be separated when dealing with cultural factors in mental measurement. Knowledge of both mental measurement and the culture

being dealt with is essential in establishing quality in cross-cultural mental measurement.

### See Also the Following Articles

Attitude Measurement ■ Conformity across Cultures ■ Cross-Cultural Psychology, Overview ■ Intelligence Assessment ■ Measurement and Counseling ■ Personality Assessment

### Further Reading

- Barrett, P. T., Petrides, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: An examination of the factor similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences*, 25, 805–819.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Bond, M. H. (Ed.). (1996). *Handbook of Chinese psychology*. Hong Kong: Oxford University Press.
- Camilli, G., & Shepard, L. N. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Cheung, F. M., Leung, K., Zhang, J., Sun, H., Gan, Y., Song, W., & Xie, D. (2001). Indigenous Chinese personality constructs: Is the Five-Factor Model complete? *Journal of Cross-Cultural Psychology*, 32, 407–433.
- Church, A. T. (2001). Personality measurement in cross-cultural perspective. *Journal of Personality*, 69, 979–1006.
- Church, A. T., & Lonner, W. J. (1998). The cross-cultural perspective in the study of personality: Rationale and current research. *Journal of Cross-Cultural Psychology*, 29, 32–62.
- Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.
- Dana, R. H. (Ed.). (2000). *Handbook of cross-cultural and multicultural personality assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Jahoda, G. (1982). *Psychology and anthropology: A psychological perspective*. New York: Academic Press.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Johnson, T. P., & Van de Vijver, F. J. R. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195–204). New York: John Wiley.
- Matsumoto, D. (Ed.). (2001). *Handbook of culture and psychology*. Oxford, UK: Oxford University Press.
- Prezworski, A., & Teune, H. (1970). *The logic of comparative social inquiry*. New York: John Wiley.
- Sinha, D. (1997). Indigenizing psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology*, Vol. 1: *Theory and method* (2nd ed., pp. 129–169). Boston: Allyn & Bacon.
- Suzuki, L. A., Ponterotto, J. G., & Meller, P. J. (Eds.). (2001). *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (2nd ed.). San Francisco: Jossey-Bass.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- Van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, 47, 263–280.
- Yamaguchi, S. (2001). Culture and control orientations. In D. Matsumoto (Ed.), *Handbook of culture and psychology* (pp. 223–243). Oxford, UK: Oxford University Press.



# Mentoring

Joyce E. A. Russell

University of Maryland, College Park, Maryland, USA

1. Definition of Mentoring
  2. Functions Provided by Mentors
  3. Phases in Mentoring Relationships
  4. Outcomes of Mentoring
  5. Challenges with Mentoring Relationships for the Organization
  6. Suggestions for Improving Mentoring Relationships in the Organization
  7. New Directions for Mentoring Research
  8. Conclusion
- Further Reading

*role modeling* The protégé's interest in modeling himself or herself after the mentor.

*separation* When the mentoring relationship changes after the cultivation period; typically, the protégé becomes more independent and is less needing of the mentor's career assistance.

*sponsorship* When the mentor nominates the protégé for promotions and/or important lateral moves either verbally or simply by association with the mentor.

## GLOSSARY

*acceptance and confirmation* The feeling that both parties (mentor and protégé) have toward each other whereby they feel comfortable in expressing their views.

*challenging tasks* The demanding or stimulating work assigned by the mentor, who gives feedback to help the protégé grow on the job.

*coaching* The career advice provided by the mentor to the protégé; the mentor suggests strategies and tactics for accomplishing work objectives and reaching career goals.

*cultivation* When career and psychosocial functions are enhanced during the 2 to 5 years after the mentoring is initiated.

*initiation* When the mentoring relationship begins to develop during the first year; both the mentor and protégé see it as mentoring.

*redefinition* The phase after the separation phase when the mentoring relationship may end or change qualities (e.g., the protégé becomes more like a peer).

Mentoring refers to a developmental relationship that exists between two persons, typically a more experienced senior person and a less experienced junior person. The mentoring relationship can also exist among peers or others within and outside an organization. Most mentoring relationships develop according to four phases: initiation, cultivation, separation, and redefinition. Mentors serve two primary functions: fostering career development of their protégés and providing psychosocial support. Numerous benefits have been reported from mentoring relationships for both mentors and protégés as well as for the organization.

## 1. DEFINITION OF MENTORING

Traditionally, mentoring has been defined as a developmental relationship that occurs between a more experienced senior employee and a less experienced junior employee. This involves a single dyadic relationship that often exists within one organization. However,

given the changing nature of organizations today (e.g., use of more project teams, flatter organizational structures), mentoring relationships are also formed among peers within and outside the organization. Mentoring generally consists of a close relationship between two people, with the primary goal of helping one of them (i.e., the protégé) develop in his or her career. This enables the protégé to address career challenges and take advantage of opportunities to grow on the job. In addition, the relationship is often a mutually enhancing one given that the mentor also receives some benefits (e.g., protégé support, work done by the protégé).

## 2. FUNCTIONS PROVIDED BY MENTORS

It is commonly accepted that mentors serve two basic types of functions to protégés: career development and psychosocial support. These are described in the following subsections.

### 2.1. Career Development

Career development refers to those aspects of the relationship that enhance the protégé's career advancement or growth. The mentor is able to help the protégé due to the mentor's experience, organizational rank, and/or influence in the organization. As a result of these functions, the protégé is often able to learn the "ropes" of the organization, gain exposure, and obtain promotions more quickly. In fostering career development, the mentor may serve any of the following roles:

- *Sponsorship*. The mentor nominates the protégé for promotions and/or important lateral moves either explicitly or simply by association with the mentor. The mentor opens doors and makes connections for the protégé that will advance his or her career.
- *Exposure and visibility*. The mentor gives the protégé opportunities to interact with senior managers or other important people through either written proposals or oral presentations. By initiating relationships with other key people, the protégé becomes visible to them and gains access to future opportunities.
- *Coaching*. The mentor provides career advice to the protégé and suggests strategies and tactics for accomplishing work objectives and reaching career goals. The mentor might provide feedback on assignments and offers ideas to improve work. In addition, the mentor helps the protégé to know who the key players are in the organization

and describes strategies for working with them successfully. Essentially, the mentor teaches the protégé the ropes for how to be successful in the organization.

- *Protection*. The mentor guards the protégé from potentially damaging conflicts or problems with others. The goal is to protect the protégé's reputation from harm. Thus, the mentor might take the blame for mistakes that are beyond the protégé's control or might act as a buffer. The mentor might limit the protégé's exposure to senior people if the mentor believes that the protégé is not experienced enough to meet with them.
- *Challenging tasks*. This usually refers to mentoring provided by a manager to a subordinate. The manager assigns demanding or stimulating work and gives feedback to help the protégé grow on the job. The mentor could also place the protégé on various project teams to help expand the latter's knowledge and skills.

### 2.2. Psychosocial Support

Psychosocial support refers to those aspects of the relationship that enhance the protégé's competence, identity, self-worth, and effectiveness in a professional role. The close interpersonal relationship that exists between the mentor and the protégé encourages the development of trust and intimacy and enables the various roles to be fulfilled. These include the following:

- *Counseling*. This refers to the mentor serving as a sounding board to the protégé so that the latter can share with someone who will listen to his or her concerns and doubts about self, work, and family. The mentor provides empathetic listening and serves as a confidant or as someone who provides advice and encouragement.
- *Acceptance and confirmation*. This refers to the feeling that both parties (mentor and protégé) have toward each other that allows them to feel comfortable in expressing their views. The feeling of acceptance by the mentor enables the protégé to feel that he or she can try new things and speak candidly. The mentor shares positive feedback that he or she has heard about the protégé. The mentor sends encouraging messages (e.g., e-mail, phone) to the protégé to help him or her deal with pending challenges.
- *Friendship*. The mentor and protégé spend time together in a nonwork setting or in leisure activities or more relaxed settings (e.g., meals, sports, outings).
- *Role modeling*. This refers to the protégé's interest in modeling himself or herself after the mentor. In this case, the protégé has great respect and admiration for the

mentor and identifies with him or her. The protégé tries to emulate the mentor's relationships with others or the ways in which the mentor works on tasks. The mentor explains what he or she is doing and why. This is so common among mentors and protégés that some researchers consider it a third (separate) function in addition to career development and psychosocial support.

### 3. PHASES IN MENTORING RELATIONSHIPS

Mentoring relationships typically develop according to four phases: initiation, cultivation, separation, and redefinition.

1. *Initiation.* During the first year, the relationship begins to develop and the mentor and protégé both see it as becoming a mentoring relationship. The mentor provides career development (e.g., coaching, visibility) and some psychosocial support (e.g., role modeling). Typically, both individuals sense a feeling of excitement and challenge as they begin to work together. The protégé feels that someone is assisting him or her, and the mentor feels that he or she is taking on someone to groom, develop, and coach.

2. *Cultivation.* During the next 2 to 5 years, the career and psychosocial functions are enhanced. The mentor and protégé both benefit from the relationship and develop a stronger connection. However, not all mentoring pairs will develop in the exact same way. It depends on the needs and interests of the mentor and protégé. In general, both parties view this period in a positive light.

3. *Separation.* In general, after the cultivation period, the relationship changes. Typically, the protégé becomes more independent and is less dependent on the mentor for career assistance. If the protégé has achieved the same organizational level as the mentor, this limits the opportunities that the mentor can provide. In addition, if the mentor experiences job obsolescence or plateauing, this limits the type of career support that he or she can provide to the protégé. Similarly, if the individuals are no longer working at the same geographical location or in similar industries, they might drift apart. During the separation phase, there may be some feelings of sadness or confusion as the relationship ends or changes.

4. *Redefinition.* After the separation phase, the relationship may end or change qualities. For example, the protégé and mentor may become more like peers, and

they may redefine their relationship as being more of a friendship. There may be some ambivalence and discomfort as they alter their relationship. At the same time, the protégé may feel a great sense of self-confidence and autonomy by "making it" to the next level.

## 4. OUTCOMES OF MENTORING

Research has indicated numerous positive benefits associated with mentoring for the protégé, the mentor, and the organization. These are described in the following subsections.

### 4.1. Benefits for the Protégé

The mentoring relationship enhances the protégé's personal and professional development, including subjective and objective indicators of career success. Through enhanced career opportunities (e.g., by sponsorship, by exposure and visibility), the protégé acquires critical learning experience that helps him or her to be able to demonstrate competence on the job. The protégé also gains an appreciation for the informal networks and political processes in the organization. This has typically led to greater career progress, faster promotion rates, and greater career satisfaction. It has been well documented that, when comparing individuals who had mentors with those who did not, protégés report significantly less work-family conflict, greater job satisfaction, greater salary growth, lower turnover intentions, and greater perceived career success. They also have greater socialization into the organization by gaining valuable information about the firm's practices.

### 4.2. Benefits for the Mentor

The mentor also reaps benefits from the mentoring relationship. By sponsoring the protégé for important positions or promotions in the organization, the mentor also gains. If the protégé is successful, others view the mentor as having made the right choice, and this enhances the mentor's credibility in terms of finding and developing talent in the organization. In addition, if the mentor's views are adopted by the protégé, this ensures that the mentor's perspectives will be carried forward in the organization for the future. By providing challenging work to the protégé and delegating to him or her, the mentor gains valuable technical assistance and support that enables him or her to be able to work

on other projects. In addition, the mentor has the loyalty of the protégé and is often kept informed by him or her about new work developments that facilitate the mentor's job performance. The psychosocial support that the mentor provides to the protégé also makes the mentor feel important and valued. If the protégé looks up to the mentor and tries to emulate him or her, this makes the mentor feel respected and appreciated in the workplace. If the mentor is asked to serve as a confidant or "counselor," this may make the mentor feel helpful and enhance his or her feelings of accomplishment. The mentor's satisfaction may also increase due to helping a less experienced colleague.

### 4.3. Benefits for the Organization

The organization also benefits from the mentoring relationship. In fact, of the rated "100 best companies to work for" in the United States each year, the majority indicate that they have established formal mentoring programs. One benefit to the organization is that protégés are better socialized into the corporate culture because mentors spend time teaching them the ropes. This means that protégés learn about the organization's values and performance metrics more quickly. In addition, mentoring is also related to lower protégé turnover intentions. This higher retention can save the firm considerable money and time that would have been spent recruiting new employees.

## 5. CHALLENGES WITH MENTORING RELATIONSHIPS FOR THE ORGANIZATION

There are a number of challenges associated with having a formal or informal mentoring program in the organization. Some of these are described in the following subsections.

### 5.1. Role Modeling Identification

As noted earlier, one important component of mentoring relationships is the degree to which the protégé identifies with the mentor. In some cases, this may be relatively easy due to similar backgrounds or life experiences. In other cases, it may be more problematic. For example, research has indicated some potential problems with the mentoring relationship where the mentor and protégé differ in their gender or race. In both cases, there may be

difficulties with the protégé being able to identify with the mentor and vice versa. For example, a female protégé who has children and is trying to balance work and family may have some trouble relating to an older male mentor who has a wife who takes care of all of the family issues. Likewise, the female protégé's experiences in being given stereotypical tasks may be difficult for her male mentor to understand. Thus, a female protégé may get more support from a female mentor who has experienced similar life and career issues. Interestingly, male protégés often are also better able to identify with role models of the same gender. Consequently, the outcomes that protégés may reap from mentoring relationships may differ depending on whether they are in cross-gender or same-gender mentoring relationships. Similarly, an African American protégé who is placed on numerous committees at work to make sure that there is "diversity of views" may have difficulty in expressing his or her concerns to a White mentor who has never had to deal with tokenism issues. Generational differences in values and concerns may also make it more difficult for protégés to relate to mentors in the organization. For example, "Generation X" protégés who are more concerned with leisure and family pursuits may have more difficulty in identifying with the serious work ethic of their "baby-boomer" mentors.

### 5.2. Perceptions by Others of the Mentoring Relationship

In some cases, there are no problems with the mentoring relationship, yet others' perceptions may cause problems. For example, with a cross-gender relationship, others may make false assumptions about the nature of the relationship (i.e., assuming it is personal/sexual). This might force the mentor and protégé to curb their social activities (e.g., having lunch, playing golf) and, thus, limit the extent to which psychosocial support or friendships can be nurtured. Consequently, they may limit their interactions to a professional or work setting only. In addition, a cross-gender relationship might cause rumors to fester about why the protégé was given various opportunities. Peer resentment and public scrutiny might force the protégé and mentor to deal with new issues and dilemmas.

### 5.3. Establishing Formal Versus Informal Relationships

Formal mentoring relationships are those established by the organization. Typically, mentors and protégés

are matched up based on some criteria such as similarity of interests and experiences. In contrast, informal mentoring relationships evolve spontaneously and without a formal mandate. Either the mentor or the protégé seeks the other one out or they both initiate the relationship. These informal relationships are considered to be voluntary, whereas formal mentoring is considered to be prescribed by the organization. In general, research has indicated that informal mentoring results in more benefits (e.g., greater support), possibly due to the closer bonds that are formed in the relationships. Yet having formal assigned relationships ensures that everyone receives some type of mentoring. Thus, the challenge for the organization is to design effective formal relationships while still facilitating the development of informal relationships.

#### **5.4. Risks Associated with Mentoring**

There may be risks associated with mentoring someone in the organization. For example, a mentor who sponsors or provides exposure and visibility to a protégé who performs poorly in front of senior managers might make the mentor look bad (i.e., the mentor's judgment in picking this person may be called into question). In addition, a mentor who is overly protective of a protégé by holding him or her back from potential opportunities can limit the career advancement of the protégé. Research has indicated that this is more likely to happen in relationships where the mentor and protégé are of a different gender or race. A senior White male mentor might be worried about showing off a female or African American protégé because he assumes that the protégé will be watched very closely and that if he or she does not do well, it will reflect poorly on the mentor. This may keep the protégé from developing critical learning experiences to grow on the job. When establishing mentoring relationships in the organization, it is important to train mentors and protégés on what their roles should be in these relationships and on what obstacles they may encounter and how to deal with them.

#### **5.5. Dysfunctional Mentoring Relationships**

Often, the assumption is that all mentoring relationships are positive or of high quality during all phases of development. Recent research has shown that this is not the case. Sometimes, mentoring relationships can

be dysfunctional or only marginally effective. For example, cases have been reported where protégés receive assistance that actually derails their careers. In other cases, there are instances of deception, sabotage, harassment, bullying, jealousy, gamesmanship, and/or poor performance by either mentors or protégés. More research is needed to further define the types of behaviors that are considered dysfunctional and their implications for protégés, mentors, and the organization. In addition, strategies to address these problems and enhance the quality of the mentoring relationship need to be defined.

### **6. SUGGESTIONS FOR IMPROVING MENTORING RELATIONSHIPS IN THE ORGANIZATION**

#### **6.1. Make Sure That Everyone Has Access to Mentoring**

Research indicates that not everyone has equal access to mentoring, yet given its importance for the success of protégés, the organization should ensure that everyone can have mentors. Creating opportunities for social activities and cross-functional work teams can facilitate the development of informal mentoring relationships. In addition, the formation of formal mentoring relationships can ensure that everyone is assigned to a mentor. The use of mentoring circles to bring multiple senior people together with several junior or mid-level people can enable a diverse group of individuals to work together.

#### **6.2. Encourage Individuals to Establish Multiple Mentoring Relationships**

Protégés should form relationships with several mentors and not rely solely on one mentor to fulfill all career and psychosocial functions. Sometimes, aligning with only one mentor can prove to be troublesome if protégés are required to follow only that one person's ideas and advice. This gives protégés only one perspective on how to be successful. It would be more advantageous if they formed alliances with several mentors at various points in time. It has been shown that professionals with multiple sources of mentors have higher levels of career success and feelings of career satisfaction than do those with only a single mentor or with no mentor at all. In fact, research on successful Black

executives indicates that they draw on multiple sources for mentoring rather than rely on a single mentor. It is also important for mentors to form relationships with several protégés. This enables them to gain support and technical assistance from multiple sources and also reduces charges of favoritism.

### **6.3. Establish a Culture That Encourages Mentoring Relationships**

Mentoring requires time to develop productive relationships. An organization that rewards individuals for developing each other (based on the performance management and incentive systems) will enable mentoring to flourish. However, a firm that focuses only on “bottom line” results and has no incentives for employees to spend time assisting others will discourage mentors from taking on protégés. Likewise, establishing formal social opportunities (e.g., lunches, celebrations) and work activities (e.g., teams) will enable the formation and growth of informal developmental relationships, some of which could lead to mentoring relationships. The organization should also support individuals’ efforts to establish relationships outside of the firm. This is particularly important given the trend today for employees to work at a variety of organizations rather than just one firm during their lifetimes.

## **7. NEW DIRECTIONS FOR MENTORING RESEARCH**

There is considerable research in the field of mentoring today. Some of the current topics for investigation were described previously, and other ideas for future research are listed in the following subsections.

### **7.1. Clarify the Mentoring Construct**

Traditionally, mentoring referred to a single dyadic relationship between a senior person and a junior person. This made mentoring relatively easy to study because it was clearly defined in this manner. Recently, many new ideas have been generated about what mentoring refers to within the broader array of developmental relationships and networks that people form. As such, the degree to which mentoring differs from coaching, support groups, advising, and counseling needs to be better explicated.

### **7.2. Examine Alternative Forms of Mentoring**

Although it has always been suggested that mentoring can take a variety of forms, during the past few years, researchers have encouraged more investigation into these different types of relationships. Future work should examine the nature of peer relationships as well as other developmental networks. It seems clear that people use a variety of relationships to gain both career and psychosocial assistance, yet more research is needed to understand how they develop these relationships and what benefit these relationships have for protégés and mentors over time.

### **7.3. Explore Mentoring for Diverse Types of Employees**

Today’s rapid changes in technology and organizational structures, as well as the emergence and decline of companies, open up new challenges for workers trying to get ahead or stay marketable. Research is needed to study mentoring among some of these workers.

#### **7.3.1. Cross-Generations**

Given differences in the values and work ethic among employees of various generations, it seems important to view what cross-generational mentors and protégés are looking for in these relationships. Do mentor-protégé pairs from different generations have unique obstacles and difficulties? If so, how can these be addressed?

#### **7.3.2. Free Agent Workers**

These refer to independent employees whose goal is to be marketable and flexible enough to move where the opportunities present themselves. Research is needed to determine the type of mentoring they need and seek. Compared with traditional workers, are they more inclined to seek mentoring from a variety of sources and networks?

#### **7.3.3. Entrepreneurs**

With the greater numbers of new business start-ups, especially among women, it will be increasingly more important to examine the types of assistance and mentoring that are needed to help them be successful.

### 7.3.4. Global Workforce

As firms continue to “go global,” it is important to examine the type of career counseling, mentoring, and training that employees need to be successful in international assignments. For example, research has indicated that mentoring for women managers in international assignments is critical for their career success. What type of mentoring is needed, and how does it differ depending on the global assignment?

## 8. CONCLUSION

Mentoring involves a developmental relationship between two people that has benefits for the mentor, the protégé, and the organization. The career and psychosocial support that are given to the protégé enhance the personal and professional development of all parties involved. Although generally positive in nature, these relationships occasionally encounter difficulties and/or obstacles. With effective support and training provided from the organization, some of these issues can be better managed. Future research is needed to examine the underlying dynamics and processes involved in the mentoring relationship.

## See Also the Following Articles

Career Counseling ■ Counseling Interview ■ Developmental Counseling

## Further Reading

- Arthur, M. B., & Rousseau, D. M. (Eds.). (1996). *The boundaryless career: A new employment principle for a new organizational era*. New York: Oxford University Press.
- Feldman, D. C. (Ed.). (2002). *Work careers: A developmental perspective*. San Francisco: Jossey-Bass.
- Hall, D. T. (2002). *Careers in and out of organizations*. Thousand Oaks, CA: Sage.
- Higgins, M. C., & Kram, K. E. (2001). Reconceptualizing mentoring at work: A developmental network perspective. *Academy of Management Review*, 26(2), 264–288.
- Kram, K. E. (1985). *Mentoring at work: Developmental relationships in organizational life*. Glenview, IL: Scott, Foresman.
- Ragins, B. R. (1997). Diversified mentoring relationships in organizations: A power perspective. *Academy of Management Review*, 22, 482–521.
- Russell, J. E. A., & Adams, D. M. (1997). The changing nature of mentoring in organizations: An introduction to the special issue on mentoring in organizations. *Journal of Vocational Behavior*, 51(1), 1–14.



# Models for Transportation

John D. Lee

University of Iowa, Iowa City, Iowa, USA

1. Introduction
  2. Model Purpose
  3. Model Scope
  4. Model Type
  5. Emerging Models of Driver Performance
  6. Driver Performance Model in the Interactive Highway Safety Design Model
  7. In-Vehicle Information Systems Demand Model
  8. Rear-End Collision Warning Assessment Model
  9. Cognitive Architectures for Modeling Driver Performance
  10. Conclusions
- Further Reading

## GLOSSARY

**cognitive architecture** Provides a broad theory of cognitive capabilities and limits; when operationalized in Adaptive Control of Thought (ACT-R) or Soar, it provides a high-level programming language for simulating human cognition that includes many of the cognitive limits and constraints that guide human performance.

**concurrent real-time model** A model that generates output in parallel with driver behavior; this type of model may make it possible to infer driver state and intention when it is not possible to measure these variables directly.

**driving performance** The description of the driver and vehicle trajectory relative to important performance limits such as lane boundaries and speed limits; typical driving performance metrics include lane deviation, steering wheel reversals, speed variation, and brake reaction time.

**driving safety** The description of the likelihood of an event that compromises driver safety; typical driving safety metrics include crashes per million miles driven, injuries

per million miles driven, and fatalities per million miles driven.

**process model** A model that predicts not only the outcome but also the response process that generates the outcome; this type of model is based on theoretical principles that go beyond the data used to develop the model.

**product model** A model that produces an output for given inputs but does not describe the causal links between the inputs to the model and the output; this type of model is typically based on interpolation of experimental data using linear regression, lookup tables, and neural nets.

Driver performance models can help to develop safe in-vehicle information systems by helping to coordinate experiments, evaluate systems, and modulate information flow. This article describes the dimensions of model purpose, scope, and transparency and then uses them to identify how several specific models may help to develop in-vehicle technology. As in-vehicle technology becomes increasingly complex and adaptive, driver performance models may need to address the effect of driver attitudes and preferences on behavior. This brief overview of four modeling approaches demonstrates some of the trade-offs in developing models of driver performance.

## 1. INTRODUCTION

This article focuses on models of driver performance that might address the challenges of integrating technology into the driving environment. Models of driver

performance can support evaluation of in-vehicle technology in several ways. First, models make it possible to examine design options before the systems are actually implemented. In this way, design concepts can be explored without the expense of developing prototype systems. Second, models can support extrapolation beyond experimental data. Experiments typically identify the relative merit of two to eight design options, whereas designers might need to understand the merit of hundreds of design options. Models can be used to estimate the merit of the design options that are not included in the experiments. Also, although catastrophic failures such as crashes are very rare and are unlikely to occur in an experiment, it may be important to understand how design parameters affect the frequency of those events. Models can help to extrapolate from measures of driving performance to measures of driving safety. For example, a model might help to translate measures of performance (e.g., lane deviations) to a measure of safety (e.g., accident rates per million miles driven). Beyond evaluating systems, models of driver performance can also play a useful role in the in-vehicle information systems themselves. Real-time models of driver capabilities can be used to modulate the information flow to the driver. In this way, a driver performance model can help to manage the flow of information so that it is consistent with the perceptual and cognitive limits of the driver.

These benefits of driver performance models are balanced by the challenges of developing such models. Fortunately, substantial research has focused on developing models of driver performance. For some applications, these models can be adapted and used without the cost of developing new models. Unfortunately, there is a great diversity of driver performance models, and understanding whether a particular model meets the needs of a particular evaluation is often unclear. This article describes the dimensions of model purpose, scope, and type and then uses these dimensions to describe how several specific models might be modified to address the demands of evaluating in-vehicle technology.

## 2. MODEL PURPOSE

Every model is created for a purpose. The purpose for which a model is developed has a major influence on the range of applications for which the model can generate useful results. The wide range of purposes of driver performance models demonstrates the challenge of finding a model capable of meeting the needs of

a particular application. Driver models have been developed to investigate vehicle dynamics, suspension, and handling characteristics. They have also been developed to support the design of roadway geometry. Recently, models have also been developed to address driver response to in-vehicle technology such as warning systems and the visual and cognitive demands on in-vehicle information systems. Another class of models has been developed to address the safety implications of personality variables and their influence on driving behavior. For example, researchers have created statistical models that link risk-taking tendencies to driving safety. Others have created statistical models that predict accident involvement based on cognitive and perceptual deficits such as the driver's effective field of view.

## 3. MODEL SCOPE

Driving is a complex multifaceted task that has been modeled in many ways. For example, traffic models have been developed to describe the combined behaviors of many drivers. These models of traffic behavior play a critical role in understanding the effectiveness of in-vehicle technology aimed at traffic management and congestion reduction. Some models of traffic behavior describe driver behavior only at the aggregate level. These models often represent macro-level behavior by using principles of fluid dynamics to describe traffic flow. Other models of traffic behavior rely on micro-level description of the decision-making and response patterns of individual drivers that make up the traffic.

Beyond models of traffic flow, many models have been developed to describe the behavior of individual drivers. These models range from those that describe the entire range of driving activities to those that describe a small subset of activities (e.g., steering behavior for lane keeping) or a small subset of driving scenarios (e.g., response to rear-end collision situations). Michon defined three levels of driver activity: vehicle control (e.g., speed control, lane position), tactical maneuvering (e.g., speed choice, turn choice), and strategic decisions (e.g., choice of route, decision to initiate a trip). These three levels define qualitatively different types of behavior that might be included in a model. Models that address strategic decisions, such as route choice and trip timing, are often represented as statistical choice models. These model results provide important information for both marketing and design of in-vehicle information systems. Most models do not

combine a representation of strategic behavior with a representation of vehicle control or tactical maneuvering.

#### 4. MODEL TYPE

In addition to distinguishing among models based on the scope of activities that they address, models can be distinguished based on whether they describe the driver's response process or just the outcome of that process. Models that describe only the outcome or product associated with a driving technology tend to be static models that pair an outcome with a set of inputs. More sophisticated static product models can combine many variables to predict driver performance in a wide range of situations; however, product models do not describe the causal links between the inputs to the model and the output. In contrast, a process model not only predicts the outcome but also specifies the response process that generates the outcome. In other words, process models are based on theoretical principles, whereas product models are based on interpolation, lookup tables, and curve fitting. Process models tend to be dynamic in that they take time-varying inputs and generate time-varying outputs. One benefit of process models is that they can predict driver responses over a range of conditions not included in the data used to create the models. Another benefit is that process models tend to make assumptions explicit. Rather than drawing on data that embody hidden assumptions, process models make assumptions easy to see because they are explicitly coded into the algorithms. Although dynamic process models have many advantages, they are not always feasible. In many cases, theory may be lacking, making it impossible to construct a process model. In addition, process models are often more complex, and the underlying mathematics may be less understandable. In contrast, product models have the advantage of providing simple solutions to complex problems. The cost of these simple solutions is the uncertainty regarding their accuracy and underlying assumptions.

#### 5. EMERGING MODELS OF DRIVER PERFORMANCE

Several recent modeling efforts demonstrate the potential of driver models to address important questions regarding driver performance and vehicle design. The

following discussion of these models does not present a complete survey of driver models but instead represents a sample of the more promising developments. In addition, these models illustrate the distinctions of model purpose, scope, and type that are important for developing a modeling strategy.

Table I summarizes the three models, showing three distinctly different approaches to modeling driver performance. The first model, the Driver Performance Model (DPM), is an ambitious attempt to model the dynamics of driver behavior. This process model produces a time history of vehicle position and speed as well as a time history of driver state. The second model, Design Evaluation and Model of Attention Demand (DEMAND), is a comprehensive product model that identifies a level of demand associated with a task performed with an in-vehicle information system such as entering a destination for a route navigation system using a touch screen. The third model, the Attention-based Rear-end Collision Avoidance Model (ARCAM), is a much less ambitious process model that generates a time history of a driver's response to the very specific scenario of emergency braking to avoid an imminent rear-end collision. The following discussion also briefly addresses the potential of general cognitive architectures to act as a foundation for a driver performance model.

#### 6. DRIVER PERFORMANCE MODEL IN THE INTERACTIVE HIGHWAY SAFETY DESIGN MODEL

DPM will help designers to understand the demands that a particular curve or other aspect of roadway geometry places on the driver. DPM is a very ambitious model that intends to capture both vehicle control (e.g., speed and path control) and tactical maneuvering aspects of driving (e.g., speed and path choice). This broad scope is bounded by a focus on driver response of road geometry of two-lane rural roads. DPM is a detailed process model. It generates predictions of driver behavior based on algorithms that are tied to elements of human cognition that are relevant to driving. These algorithms are based on a theoretical framework that relates cognitive resources, such as attention, perception, and motor control, to the driving demands associated with moving down a virtual roadway. As a process model, DPM produces a response to state

**TABLE 1**  
**Characteristics of Three Representative Driver Models**

	<i>Purpose</i>	<i>Scope</i>	<i>Type</i>
DPM	Support geometric design of roadways	The control and tactical aspects associated with speed and path maintenance on a two-lane rural road	Process model based on a nested closed-loop feedback control
DEMANd	Support interface design of in-vehicle information systems	The control aspects of interacting with in-vehicle technology Considers driving only implicitly as it is degraded by competition for common attention resources	Product model based on lookup tables and linear regression
ARCAM	Support selection of appropriate algorithm parameters for a rear-end collision warning system	The control and tactical aspects of driver response to imminent collision situations	Process model based on a field theory of driving and a combination of open-loop and closed-loop control
Cognitive architecture	Model high-level cognition in a variety of settings	Aspects of cognition that correspond to control and tactical aspects of driving	Process model based on empirical data collection and experimental validation

changes that are updated at each time step of the simulation. This results in time history of the driver state and vehicle position on the road. Because the developers created DPM using a general framework with a strong theoretical basis, it is feasible to extend DPM to go well beyond its current purpose. Adapting it to address distraction associated with in-vehicle devices seems feasible.

## 7. IN-VEHICLE INFORMATION SYSTEMS DEMAND MODEL

The emergence of a wide array of in-vehicle technology has promised increased safety, mobility, convenience, and efficiency. The purpose of DEMANd is to help engineers evaluate design options before creating a prototype in-vehicle information system. More specifically, this system enables engineers to compare two competing options, evaluate a modification to an existing system, and evaluate a task or design feature against safety criteria.

The focus of DEMANd is on in-vehicle tasks, and driving performance is modeled indirectly; that is, driving performance decrements are estimated based

on the demands of tasks associated with in-vehicle information systems. DEMANd takes the tasks and subtasks associated with an interaction with an in-vehicle device and produces estimates of the attentional demands of those tasks. DEMANd summarizes the consequence of these attentional demands in terms of expected decrements in driving performance. Unlike DPM, DEMANd does not generate a time history of driver actions and driving performance; instead it combines data across in-vehicle tasks to estimate a general indicator of the safety consequence. Because the model does not generate output for the entire time history of the interaction among the driver, the system, and the roadway, it might not reflect how peaks in attentional demand could overwhelm the driver.

DEMANd is a static product model of driver performance. It uses lookup tables and linear regression equations to relate particular design alternatives to declines in driving safety. Unlike DPM, DEMANd does not generate a time history of driver response to events; instead it adds the estimated cognitive demands of a set of tasks and subtasks. This approach does not attempt to describe the dynamic process that the driver goes through in interpreting and responding to an in-vehicle information system.

### 8. REAR-END COLLISION WARNING ASSESSMENT MODEL

ARCAM was developed to aid designers in evaluating candidate collision warning strategies. This model predicts driver response to warnings produced by various collision warning algorithms in different driving contexts (e.g., initial velocity, following distance, deceleration of the lead vehicle, initial velocity of the lead vehicle). With these predictions, designers can compare various algorithms of parameter choices for a wide range of roadway conditions. Currently, ARCAM predicts behavior of drivers only in imminent rear-end collision situations.

Figure 1 shows the major components of ARCAM. A central aspect of ARCAM is the focus of driver attention. A driver's attention to the roadway depends on the uncertainty about the state of the roadway. This uncertainty increases until it reaches an unacceptable level, at which time the driver returns attention to the road. The uncertainty grows as a function of information density of the road, vehicle velocity, rate of for-

getting, and time. The following equation describes the driver's uncertainty:

$$\text{Standard Deviation of Memory Position} = \alpha + \lambda \times t^{1.5}, \quad (1)$$

where  $t$  is the time since the last sample,  $\alpha$  is the initial uncertainty associated with the driver's estimate of collision potential, and  $\lambda$  is the rate of uncertainty growth.

ARCAM is a detailed process model. It generates predictions of driver behavior based on algorithms that are tied to elements of human cognition that are relevant to drivers' response to imminent collision situations. These algorithms are based on a field theory of perception and a model of attention allocation that seems to capture the most critical cognitive processes associated with collision avoidance. As a process model, ARCAM produces a response to state changes that is updated at each time step of the simulation. This results in a time history of the driver state and vehicle position on the

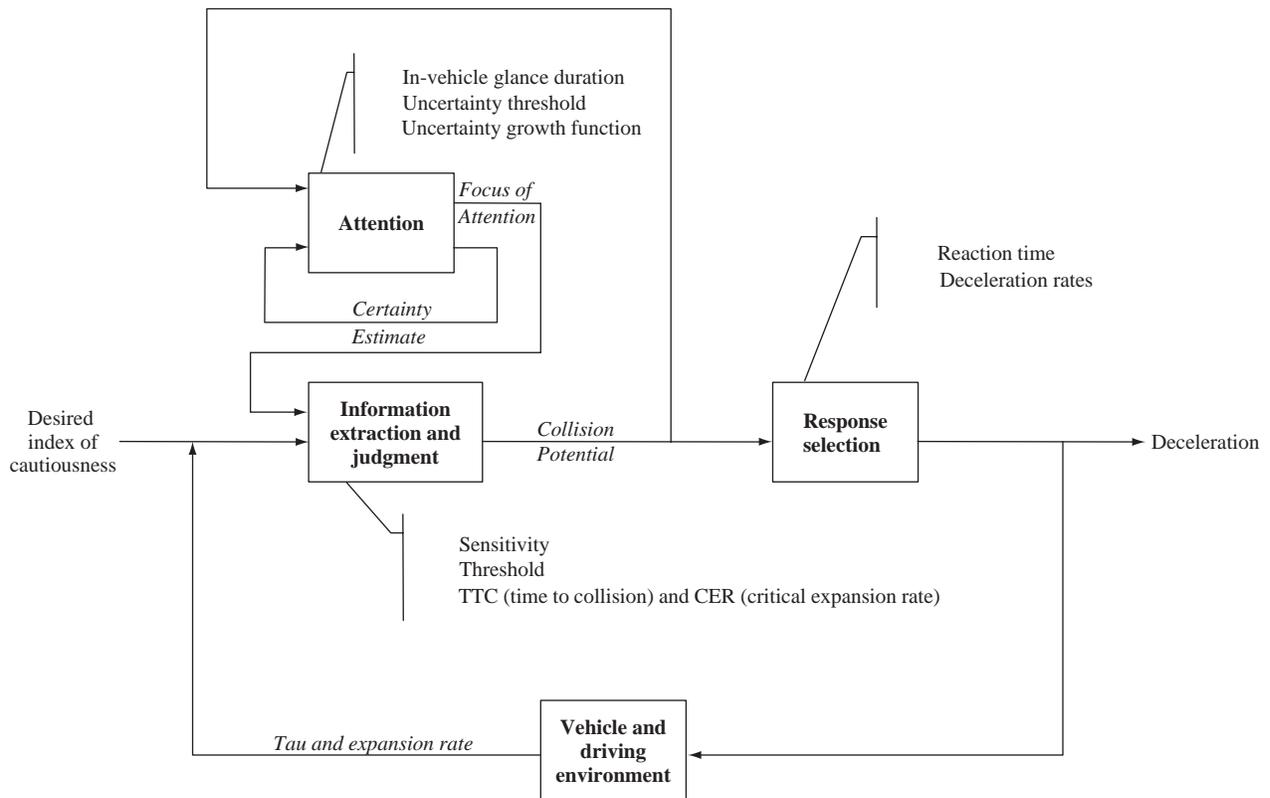


FIGURE 1 Major components of ARCAM. Reprinted from Brown and colleagues (2000).

road. This trajectory provides a basis for understanding how cognitive processes influence driver performance.

## 9. COGNITIVE ARCHITECTURES FOR MODELING DRIVER PERFORMANCE

In addition to models developed specifically to capture the demands of driving, several modeling tools or cognitive architectures could be used to model driving performance. These cognitive architectures include Adaptive Control of Thought (ACT-R), Soar, and Human Operator Simulation (HOS). A major benefit of these approaches is that they embody many of the cognitive limits and constraints of people. For example, ACT-R was developed to account for many elements of high-level cognition. ACT-R captures many aspects of how people learn and respond to problems. Using ACT-R or another cognitive architecture as a starting place for developing a driver model allows model developers to leverage many years of research as well as a validated approach to describing human behavior. Developing a model with ACT-R, Soar, or HOS can be much faster and provide results with greater validity than can an approach that develops the model without the benefit of an underlying cognitive architecture.

Although cognitive architectures offer important benefits, they also pose two substantial challenges. First, the level of detail assumed by many architectures might be more specific and detailed than many analyses need. Second, the cognitive architecture may make fundamental assumptions regarding human behavior that might conflict with how people drive. For example, ACT-R would describe driving in terms of discrete activities of symbol manipulation. Although this might be appropriate for some driving activities, many aspects of driving might best be described as continuous closed-loop control. Third, the tasks that many cognitive architectures have been designed to model are not necessarily those that comprise driving. Cognitive architectures act as a high-level programming language that can greatly enhance the efficiency of model development because they embody many of the cognitive constraints that govern driver performance. However, they have been developed to explain human performance in very different domains (e.g., human-computer interaction) and, thus, may make fundamental assumptions that conflict with the human performance capabilities and time-sharing strategies used for driving.

## 10. CONCLUSIONS

This brief overview of four modeling approaches demonstrates some of the trade-offs in developing models of driver performance. Comprehensive process models of driver performance, such as DPM, can be complicated to operate and resource intensive to create. However, they provide a powerful way in which to understand how design options affect driver cognition and performance. Outcome models, such as DEMAnD, can link large amounts of empirical data, using lookup tables and regression equations, to design parameters. They provide a powerful tool for designers, but they make many assumptions that might not be easily understood, and they might not be usable beyond the original purposes of the models. Focused process models, such as ARCAM, provide the benefits of the more comprehensive process models and are less resource intensive to build. Unfortunately, they might have such limited scope that they are useful only for only a very specialized application such as the evaluation of a rear-end collision warning algorithm. Cognitive architectures promise the generality of a comprehensive process model, such as DPM, without the demands of creating a complete modeling framework. The limitation of cognitive architectures such as ACT-R is that they have generally been developed for applications unrelated to driving, and the ability to adapt them to the driving domain has not been tested.

### See Also the Following Articles

Accidents in Transportation ■ Aviation ■ Driving Safety ■ Traffic Safety Assessment ■ Transportation Systems, Overview ■ Travel Behavior and the Environment

### Further Reading

- Anderson, J. R. (1990). *Adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.
- Ball, K., Owsley, C., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1993). Visual-attention problems as a predictor of vehicle crashes in older drivers. *Investigative Ophthalmology & Visual Science*, 34, 3110–3123.
- Brown, T. L., & Lee, J. D. (2000, January). *Driver modeling in the design of rear-end collision warning systems*. Paper presented at the annual meeting of the Transportation Research Board, Washington, DC.
- Brown, T. L., Lee, J. D., & McGehee, D. V. (2000). An attention-based model of driver performance in rear-end collision situations. *Transportation Research Record*, 1724, 14–20.

- Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson, & C. Lebiere (Eds.), *Atomic components of thought* (pp. 143–168). Mahwah, NJ: Lawrence Erlbaum.
- Cameron, G. D. B., & Duncan, G. I. D. (1996). PARAMICS: Parallel microscopic simulation of road traffic. *Journal of Supercomputing*, 10(1), 25–53.
- Chowdhury, D., Santen, L., & Schadschneider, A. (1999). Vehicular traffic: A system of interacting particles driven far from equilibrium. *Current Science*, 77, 411–419.
- Goldbach, M., Eidmann, A., & Kittel, A. (2000). Simulation of multilane freeway traffic with detailed rules deduced from microscopic driving behavior. *Physical Review E*, 61(2), 1239–1246.
- Hankey, J. M., Dingus, T. A., Hanowski, R. J., Wierwille, W. W., & Andrews, C. (2000). *In-vehicle information systems behavior model and decision support: Final report* (FHWA-RD-00-135). McLean, VA: Turner–Fairbank Highway Research Center.
- Jonah, B. A., & Dawson, N. E. (1987). Youth and risk: Age differences in risky driving, risk perception, and risk utility. *Alcohol, Drugs, and Driving*, 3(3–4), 13–29.
- Kantowitz, B. H., Levison, W. H., Hughes, W., Taori, S., Palmer, J., Dingus, T. A., Hanowski, R., Lee, J. D., Mears, B., & Williges, R. (1997). *Development of prototype driver models for highway design: Task A final Project I work plan*. Seattle, WA: Battelle Human Factors Transportation Center.
- Klar, A., & Wegener, R. (1997). Enskog-like kinetic models for vehicular traffic. *Journal of Statistical Physics*, 87(1–2), 91–114.
- Klar, A., & Wegener, R. (1999). A hierarchy of models for multilane vehicular traffic I: Modeling. *Siam Journal on Applied Mathematics*, 59, 983–1001.
- Lane, N. E., Strieb, M. I., Glenn, F. A., & Wherry, R. J. (1981). The Human Operator Simulator: An overview. In J. Moraal, & K. F. Kraus (Eds.), *Manned systems design methods, equipment, and applications* (pp. 121–151). New York: Plenum.
- Levison, W. H. (1994). A simulation model for driver's use of in-vehicle information systems. *Transportation Research Record*, 1403, 7–13.
- Mannering, F. L., Abu-Eisheh, S. A., & Arnadottir, A. T. (1990). Dynamic traffic equilibrium with discrete/continuous econometric models. *Transportation Science*, 24, 105–116.
- Mannering, F., Kim, S. G., Ng, L., & Barfield, W. (1995). Travelers' preferences for in-vehicle information systems: An exploratory analysis. *Transportation Research C: Emerging Technologies*, 3, 339–351.
- Michon, J. A. (1989). Explanatory pitfalls and rule-based driver models. *Accident Analysis and Prevention*, 21, 341–353.
- Modjtahedzadeh, A., & Hess, R. A. (1993). A model of driver steering control behavior for use in assessing vehicle handling qualities. *Journal of Dynamic Systems Measurement and Control*, 115, 456–464.
- Moray, N. (1981). The role of attention in the detection of errors and the diagnosis of errors in man–machine systems. In J. Rasmussen, & W. B. Rouse (Eds.), *Human detection and diagnosis of system failures*. New York: Plenum.
- Morrall, J., Miller, E., Smith, G. A., Feuerstein, J., & Yazdan, F. (1995). Planning and design of passing lanes using simulation model. *Journal of Transportation Engineering*, 121(1), 50–62.
- Nelson, P., & Raney, B. (1999). Objectives and benchmarks for kinetic theories of vehicular traffic. *Transportation Science*, 33, 298–314.
- Newell, A. (1992). Precis of unified theories of cognition. *Behavioral and Brain Sciences*, 15, 425–492.
- Reagan, J. A. (1994, Summer). The Interactive Highway Safety Design Model: Designing for safety by analyzing road geometrics. *Public Roads*, pp. 37–43.
- Schuller, J., Schaefer, F., Neukum, A., & Krueger, H. P. (1999). A driver model as development tool for vehicle handling design. *Vehicle System Dynamics*, 33, 110–121.
- Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record*, 195, 15–33.
- Sharp, R. S., Casanova, D., & Symonds, P. (2000). A mathematical model for driver steering control, with design, tuning, and performance results. *Vehicle System Dynamics*, 33, 289–326.



# Motivational Taxonomies

Jeffrey B. Vancouver

Ohio University, Athens, Ohio, USA

1. Introduction
  2. Definitions
  3. Content Theories
  4. Process Theories
  5. Implications for Work Settings
- Further Reading

## GLOSSARY

- contingency** Pairing or relationship between two events, behaviors, or thoughts; usually, the pairing or relationship is probabilistic (i.e., it does not always happen).
- intrinsic motivation** Behavior performed for its own sake and not for material or social rewards.
- justice** Perceptions of the fairness of the allocation of outcomes (i.e., distributive justice) or policies and procedures used to make the allocation decisions (i.e., procedural justice).

Motivating oneself or others in a work setting requires understanding many aspects of human thought and behavior. Several theories have emerged to facilitate understanding what people want and how they learn and process information to get what they want.

## 1. INTRODUCTION

Motivation has long been a topic of interest to applied psychologists. In work settings, motivational processes

are thought to relate to decisions of participation, such as joining or leaving an organization, and to performing well in an organization. Motivation is also implied in decisions and actions related to broader organizational concerns such as unionization, constructive behaviors (e.g., speaking well of one's employer), and destructive behaviors (e.g., workplace violence). An understanding of motivational constructs and processes can help employers to predict who will be motivated to join and work hard, why some are not so motivated, and what can be done to increase participation and performance generally and for specific individuals. From an employee's perspective, an understanding of motivation might help the individual to make participation decisions that lead to satisfying work experiences and techniques for motivating oneself to perform desired tasks promptly and well. To address these issues, the field of work motivation has generated a list of theories that speak to specific issues. This article describes some historically important and currently prominent theories in the field.

## 2. DEFINITIONS

Before reviewing the theories, it is useful to define the two concepts comprising the title of this article: motivation and taxonomy. Motivation is difficult to define. It is easier to understand in terms of what it is not: ability, knowledge, and skills. For example, a well-written essay requires knowledge of the topic, the ability to write, and the willingness to write. Only the willingness element is

motivation. More directly, motivation is often defined as that which determines the form, direction, intensity, and duration of one's actions. However, one must recognize that ability or other influences (e.g., situational constraints) might at least partially determine these results. Thus, separating motivational causes from nonmotivational ones represents a continual dilemma for motivational researchers.

Taxonomy represents a classification scheme in which elements are assigned to no more than one class and all of the elements are assigned; that is, it is mutually exclusive and exhaustive. Taxonomies are also often hierarchical; that is, within a class, subclasses are often identified. Here, content may be used as one higher order class and process as another class. Content theories deal with a particular property or set of properties thought to influence motivation. They address the question of what people want or need. Process theories focus on how individuals handle information or content. They address the question of how people learn to obtain or process information to obtain what they want and need. Process theories are content free; that is, they usually assume that people are hedonistic (i.e., seek pleasure and avoid pain), but they make no statements about what people inherently find to be pleasurable and painful. Content theories usually have some description of processes, but they also include descriptions (often taxonomies) of what might give people pleasure and pain.

### 3. CONTENT THEORIES

#### 3.1. Needs Theories

One well-known taxonomy of human desires (or needs) was presented by Maslow, a clinical psychologist, based on a lifetime of interaction with patients. This taxonomy was arranged in a hierarchy designed to represent the order of precedence (Fig. 1). The bottom level (physiological needs) describes what individuals need to survive. Higher level needs describe what individuals need to thrive. A process element to Maslow's theory was that lower level needs had to be met to some extent before higher level needs became motivating. Maslow's hierarchy was important for work motivation because it included in the higher order needs more than just the material needs. Indeed, the highest need (self-actualization) describes the desire of individuals to be all that they can be. It represents a desire to strive, in both personal and work lives, to achieve new heights, to create, and to produce to the best of one's ability. This sentiment resonated with

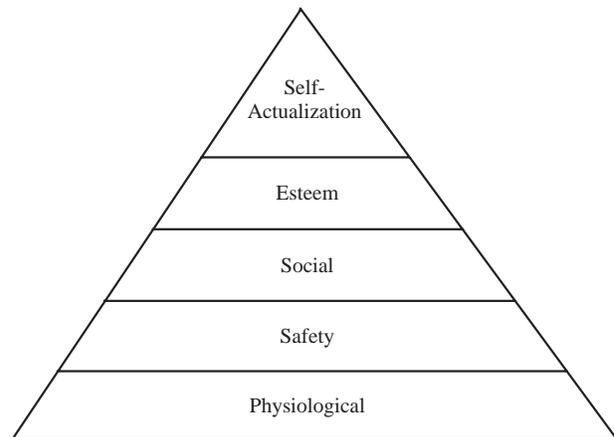


FIGURE 1 Maslow's hierarchy of needs.

managers and executives because they believed that these kinds of desires motivated their behavior.

Unfortunately for Maslow's theory, the majority of empirical research did not support his hierarchy. Some suggested that the hierarchy be reduced to two or three levels with some processes added. Yet even these changes were not reconcilable with empirical and logical counterarguments (e.g., an individual can be eating a sandwich to meet a physiological need while writing his or her self-actualizing masterpiece). These issues led most motivational researchers to abandon the specific framework but to scavenge useful pieces from the hierarchy, including the general idea that people desire more than what money can buy.

#### 3.2. Job Design

One area of motivational research that benefited from the needs approach was job design. If applied psychologists could understand what was intrinsically motivating about jobs, they might be able to redesign them so that workers were more motivated and more satisfied. Interviews with workers seemed to reveal that although lack of material rewards and/or poor working conditions (i.e., extrinsic elements) could make them unhappy, it was the opportunity to do a meaningful job well that was truly motivating and satisfying.

So, what makes a job meaningful? The most prominent theory in this area, called the job characteristics model, claims that it is the degree to which the job involves a variety of skills used to create a whole important thing (or service). In addition, the job should allow the individual (or team) to decide what, when, and how

to do the thing (or provide the service). Finally, the job should provide feedback regarding progress directly to the job holder. To the extent that these factors are present, an individual doing the job should be motivated because it allows the individual to demonstrate competence and an ability to affect his or her world. The theory also recognizes that if these needs in the individual are not strong, these job dimensions are not likely to be important and might even aggravate the individual.

Because job design theories are specific about what should be done, they are easier to test than are vague needs taxonomies. In this case, the empirical literature tends to be much more supportive of the recommendations within the theory, particularly in terms of job satisfaction. In terms of helping an organization's bottom line (i.e., profit and productivity), the findings are less clear. One issue is that recommendations from motivation-oriented job design techniques and engineering-oriented job design techniques often conflict. Engineering-oriented job designers tend to focus on efficiency, error reduction, and reliability. This approach emerged from Ford's assembly line formula that seemed to greatly improve the efficiency of production through work simplification, even as it substantially reduced the humanity of work. Some now suggest balancing all of the relevant criteria (e.g., reliability, intrinsic motivation) when designing and evaluating jobs.

### 3.3. Equity and Justice

A different class of theories focuses on what an individual does not want, that is, to be treated unjustly. Simply observing elementary school-aged children distributing a reward among themselves reveals the importance of this desire for fairness (and the difficulty in achieving it). The desire does not disappear with adulthood. Organizational members seem acutely aware of the fairness of distributions of outcomes and the policies used to make these distributions. The original theory focused solely on the distributions of outcomes (i.e., distributive justice). More recent work has focused on the fairness of distribution policies (i.e., procedural justice).

Actually, the original justice theory, called equity theory, was focused on a process, that is, the comparison between the outcomes that an individual receives and the outcomes that others receive. Moreover, these outcomes were appraised in relation to the inputs used to derive them. Hence, the theory predicted that if an individual received \$5 for a job and a friend received \$10, the former would perceive this as fair if the friend worked twice as long or twice as hard. On the other

hand, if the individual perceived that both he or she and the friend put equal work into the job, that person would see the pay as unfair and be motivated to rectify the situation. The individual might do that by working less hard in the future, finding some other form of compensation (e.g., pilfering office supplies), denigrating the friend's work to the boss, or coming to believe that he or she does not really put in the same quality of work as the friend. If none of these produced a sense of justice, the person might quit that job. Research has found that all of these things can happen. Meanwhile, the friend is likely to believe that his or her efforts are somehow worth more than the first person's efforts, relieving any sense of overpayment inequity. Researchers hoped that individuals would be motivated to work harder to overcome their sense of overpayment inequity, but the evidence seems to indicate this is rarely the case.

More recent research has focused on the perceived policies and procedures for making distribution decisions. Voice in the process, whether at the policy development stage, the policy implementation stage, or a follow-up stage (e.g., ability to appeal decisions), seems to be a central issue. In addition, thoughtful explanations for negative outcomes (e.g., layoffs) seem to reduce justice-restoring behaviors (e.g., sabotage). The procedural justice research seems to be summed up in the concept that one should treat people with respect.

## 4. PROCESS THEORIES

Process refers to how people meet (or strive for) their wants and needs. Three kinds of process theories have gained prominence in organizational psychology: learning theories, decision theories, and goal theories. Learning theories describe how people learn to achieve their desires. Decision theories describe how people process information such that they make choices leading them to meet their desires, including how they choose goals. Goal theories describe the properties of goals that most influence behavior and performance.

### 4.1. Learning Theories

Most learning theories claim that people make associations between responses and the consequences that follow those responses. If the consequences are what people want, people are more likely to make those responses in the future. Thus, if making a big sale earns the respect of bosses and colleagues, an individual is more likely to engage in the activities that led to the big sale in the

future. In addition, when stimuli (e.g., events) are frequently paired, people learn to associate the stimuli, allowing them to predict their environment to some extent. Thus, one might associate certain attributes of customers with sales. Observations of customers with such attributes indicate opportunities not otherwise present. The result might be the engagement of certain behaviors associated with sales directed at customers with certain attributes.

In organizational settings, the most common application of learning theories is behavior modification programs based on reinforcement theory. They involve determining the desired behaviors of employees and then setting up contingencies with rewards when the behaviors are observed. Well-designed studies of these programs have generally supported them, but several aspects of the approach make it onerous for organizations. First, it requires that one or a few specific behaviors are what all employees should be doing. For many jobs, this might not be the case. Second, it requires someone (or something) to look for the behaviors by the employees. Such extensive monitoring can be expensive and aversive to employees. Third, because the theory is content free (i.e., it does not state what people want), it requires figuring out what is rewarding for each employee. This generally is not a problem because most people want pay, which is a very flexible means to other ends. Unfortunately, employee pay is an organizational cost that organizations would like to minimize. Hence, some effort is made at finding less expensive rewards (e.g., praise). Fourth, theorists in this area advocate using the programs to encourage (i.e., reinforce) desired behaviors as opposed to discouraging unwanted behaviors. This can limit the approach's applicability. Finally, some motivation theorists argue that behavior modification programs undermine intrinsic motivation because they do not allow individuals to infer competence from their behavior.

In response, learning theorists point out that learning is constantly occurring in organizations. Organizations could simply hope that their work environments encourage the behaviors desired of employees, or they could attempt to construct environments that encourage the desired behaviors (i.e., behavior modification programs). Moreover, the intrinsic motivation issue does not seem to be a problem because individuals use the rewards to infer competence. Nonetheless, because of the difficulties associated with identifying an easily monitored set of specific behaviors, many researchers have sought to identify important properties of the work environment that might encourage or discourage the

desired behaviors. For example, the beliefs and desires of coworkers have been found to be significant sources of influence on learning and behavior.

## 4.2. Decision Theories

Decision theories have several advantages over other theories of motivation from the perspective of motivational researchers. First, as the key dependent variable, they can use self-report measures of choice rather than behavior, which can be hopelessly confounded with ability. Second, because decision theories assume the cognitive processing of information, researchers can also obtain reports from workers about the contingencies and associations they perceive in their environments. These theories focus on what individuals do with the associations learned as opposed to how they are formed. Hence, researchers can be less concerned with what the environments look like and can focus on how those environments are interpreted by the workers.

Several decision theories exist, but within organizational psychology, the expectancy-value models have received the most attention. Expectancy-value models reduce workers' interpretations of their environments to three key constructs. First, they are concerned with the contingencies that workers see between specific behaviors and performance; that is, individuals' belief in the probability that behaviors will lead to performance (although these theories consider more than performance). Second, expectancy-value theories claim that the associations that individuals form between performance and other outcomes are important. Finally, expectancy-value theorists want to know what anticipated value workers place on the outcomes associated with performance. Given this information, the theories predict that individuals will choose to engage in behaviors that will lead to performance (or choices) if the outcomes associated with performance (or choices) are positively valued. The amount of effort or persistence directed toward the behaviors will weaken to the extent that individuals believe the behaviors might not achieve performance or believe the performance might lead to negative or nonvalued outcomes.

The predictions from decision theories and learning theories are very similar. They both advocate the setting up of contingencies between performance and valued rewards. However, they are more likely to encourage training or the removal of performance barriers if behavior-to-performance expectancies are low instead of setting up specific behavior-based contingencies. This moves the monitoring to the level of performance and away from specific behaviors.

Research on decision theories is generally supportive. People do seem to use their beliefs regarding contingencies and value to make choices regarding what jobs to take and other work-related behaviors. Moreover, self-report methods of eliciting these beliefs can be useful as organizations seek to understand why performance levels are not up to standard or what might improve them. However, for more mundane issues, such as determining how much effort to exert toward performance, the theories are less impressive. Indeed, exceptions seem to be the rule, and decision researchers are attempting to explain why people deviate from doing what seems to be in their best interest as defined by an expectancy–value theory.

### 4.3. Goal Theories

One decision that seems to be critical to understanding, predicting, and influencing performance is the choice of a performance goal. Most current research focuses on goal theories, which continue the cognitive orientation found in the decision theories. That is, goals are defined as desired states internally represented within the mind. Properties of this representation, including its level (e.g., produce a little or a lot), its level of specificity (e.g., do one's best or do 20), and one's commitment to it seem to substantially affect goal-striving behavior. Specifically, difficult (i.e., high-level), well-specified goals with strong commitment will lead to higher performance than will easy vague goals with low commitment. Although the research has been somewhat ambiguous regarding the commitment concept, the level and specificity constructs are well supported within boundary conditions. For example, feedback regarding performance progress is required for goals to have an effect. Moreover, learning and complex tasks might be adversely affected by difficult performance goals, but learning goals or contexts where the goals encourage the development of appropriate strategies for performing complex tasks can enhance these outcomes. Also, it seems clear that through the integration of decision theories, beliefs about contingencies with valued outcomes can enhance commitment and the acceptance of difficult goals.

### 4.4. Integrative Efforts

Although controversy surrounds the value or potential of integrating the various motivational theories, several efforts have been made in that direction. One early attempt combined expectancy–value notions with a particular content, that is, need for achievement. According

to this theory, achieving difficult tasks (where expectancy was low) is more rewarding but may also lead to failure. Depending on each individual's desire for achievement, fear of failure, and expectancies, these factors predict choice and effort. Unfortunately, the evidence was not always supportive, and the theory has lost favor. However, the content (i.e., need for achievement) remains a candidate for being a measure of general work motivation.

More recently, social cognitive theory has emerged as an integrating theory of the process models of motivation. Social cognitive theory grew out of social learning theories that included concepts of learning from experience and from watching others as well as associations derived from people's history of experiences. Social cognitive theory, like expectancy–value theories, added the importance of how individuals perceive and interpret events and contingencies. For instance, individuals might use their feelings of physiological arousal to help them judge their ability to perform behaviors. According to the theory, a person's belief in his or her capacity to engage in a specific behavior, called self-efficacy, is key to understanding how the individual allocates resources to the task and how he or she reacts to frustrations and successes. For instance, individuals who are high in self-efficacy are likely to raise their goals when they achieve existing goals. Hence, goal processes are also included in social cognitive theory.

## 5. IMPLICATIONS FOR WORK SETTINGS

Several implications of specific theories have been described in this article. This section addresses some specific applied questions and what the theories do and do not have to say about them.

### 5.1. Who Should Be Hired?

One job of applied psychologists is to find individual differences that predict (i.e., correlate with) job performance. If measures of these predictors can be given during the application phase, organizations can use them to select individuals who will perform well. This strategy has proven to be fairly successful when it comes to measuring and predicting ability, knowledge, and skill. However, predicting who will be motivated has proven to be more difficult than predicting ability. This makes sense. Content theories (e.g., need for

achievement), which are most likely to provide explicit measures, have often suffered due to measurement problems. Even if the measures were reliable and valid, highly motivated individuals might direct their effort to nonwork endeavors (e.g., hobbies). Moreover, jobs often leave little leeway regarding what should be done or how to do it. For these reasons, research has tended to focus on more specific measures of motivation (e.g., work values) that indicate a fit with the organization where the job allows some autonomy or ambiguity.

### 5.2. Why Do Individuals Accept Job Offers and Stay?

Assuming that an organization likes what it sees in an applicant, the applicant has the next choice with regard to the organization (i.e., whether to join). Later, after being hired, the employee may decide to quit. Decision theories have much to say about these processes, and they have been used extensively to predict these outcomes. Particularly with regard to the decision to join an organization, factors such as pay and benefits play an important role. In regard to quitting, these factors and the working environment have an impact. Of particular importance for these decisions are perceptions of equity and fair treatment.

### 5.3. Why Do Individuals Work Hard (or Not)?

The motivational process theories argue for making pay and other rewards contingent on performance and providing specific difficult goals. Although the second recommendation is generally well received, the first one has created controversy. It seems to help motivate some employees, and some even like the system. However, in other situations, the measures of performance needed for contingency plans to work are considered suspect. If the measures (a) do not capture the behavior or performances that lead to organizational productivity or (b) are perceived to be a function of influences outside of an individual's control (e.g., the politics and idiosyncracies of supervisors), the measures are unlikely to work well and can lead to unwanted effects. Yet if individuals were informally surveyed about their performance appraisal systems, one would likely find few employees who believe that the systems capture all that they are supposed to capture and that they capture only what they are supposed to capture. Adding insult to injury is that in

jobs where performance is difficult to measure, it is also difficult to set fair performance goals. For instance, consider a high school teacher. How should performance for this job be measured? What goals could be set that do not rely on input from the students that is beyond the teacher's control?

Indeed, can the possibilities expressed in Maslow's view of people be realized? Can organizations and jobs be designed to be places where workers can derive meaning and satisfaction? Can policies and procedures that are perceived as just be developed? Can a fair wage be provided for work done without having to monitor the work done? These are challenging questions for which the motivational theories provide some reasons for optimism as well as some reasons for pessimism. This leads to the final, and perhaps most important, question: Can a better understanding of human motivation help to answer these questions? Among work motivation researchers, the answer is uniformly "yes."

### See Also the Following Articles

Intrinsic Motivation and Self-Determination ■ Job Analysis, Design, and Evaluation

### Further Reading

- Ambrose, M. L., & Kulik, C. T. (1999). Old friends, new faces: Motivation research in the 1990s. *Journal of Management*, 25, 231–292.
- Campbell, J. P., & Pritchard, R. D. (1976). Motivation theory in industrial and organizational psychology. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 63–130). Chicago: Rand McNally.
- Donovan, J. J. (2002). Work motivation. In N. Anderson, D. Ones, K. Sinangil, & C. Viswesvaran (Eds.), *The handbook of industrial, work, and organizational psychology*. Thousand Oaks, CA: Sage.
- Kanfer, R. (1990). Motivation theory in I/O psychology. In M. D. Dunnette, & L. M. Hough (Eds.), *The handbook of industrial/organizational psychology* (Vol. 1, pp. 75–170). Palo Alto, CA: Consulting Psychologists Press.
- Landy, F. J., & Becker, W. S. (1987). Motivation theory reconsidered. *Research in Organizational Behavior*, 9, 1–38.
- Mitchell, T. R. (1997). Matching motivational strategies with organizational contexts. *Research in Organizational Behavior*, 19, 57–149.
- Pinder, C. C. (1998). *Work motivation in organizational behavior*. Upper Saddle River, NJ: Prentice Hall.



# Motivation and Culture

**Dennis M. McInerney**

University of Western Sydney, Macarthur,  
Sydney, Australia

**Martin L. Maehr**

University of Michigan, Ann Arbor,  
Michigan, USA

**Martin Dowson**

Institute of Christian Tertiary Education,  
Round Corner, Australia

1. Cross-Cultural Approaches to Achievement and Achievement Motivation
  2. Toward a Basis for Investigating Achievement and Achievement Motivation across Cultures
  3. Applying Social Cognitive Conceptualizations of Motivation and Achievement
  4. Methodological Improvements in Applied Settings
  5. Conclusion
- Further Reading

*personal investment theory* A theory concerned with how persons choose to invest their energy, talent, and time in specific pursuits.

*pseudo-etic research* Research that assumes that research constructs and tools are etic when, in fact, they are emic; consequently, the validity of cross-cultural comparisons with pseudo-etic models is questionable.

*social cognitive theory* A theory that attributes learning to the cognitive processing of observed models of thought, action, and behavior.

## GLOSSARY

*achievement motivation* The desire to excel, that is, the impetus to strive for excellence and success.

*attribution- and cognitive-style models* Theoretical conceptions that emphasize the causal role of thinking patterns in determining motivation.

*ecological research* Research that suggests that the experiential and performance dimensions of events and situations are essential in understanding the achievements, behaviors, responses, and scores of groups and individuals.

*emic research* Research that develops culture-specific constructs and tools that may limit cross-cultural comparability.

*etic research* Research that develops universally appropriate constructs and tools and that enhances cross-cultural comparability but may limit access to emic data.

*goal theory* A theory that posits that the purposes students have for learning, such as doing better than others and improving their personal performance, influence the nature and quality of their motivation and engagement in learning.

The potential impact of culture on the nature and functions of psychological variables has long been recognized. However, until relatively recently, it was often assumed that definitions, methodologies, data, and generalizations from psychological research could be readily transferred, more or less, across cultural boundaries. Typically, this meant that psychological theory and research that was based in (and on) American or Western European settings was considered readily transferable to other cultures with few, if any, modifications. In particular, the assumption of transferability applied to a range of psychological constructs of interest in applied (i.e., real-life) settings. These constructs included intelligence, anxiety, expectancies, motivation, perceptions of success and achievement, and self-concept. However, there is now a wide recognition that such constructs are not necessarily structured in the same way across cultures. Moreover, the relative importance of various constructs may differ greatly among

cultures. For example, it has been demonstrated that the ways in which students construct their understandings of achievement and success (i.e., what “counts” for achievement/success and how achievement/success is gained) may be very different across cultures. For this reason, there has been considerable impetus to reformulate important psychological constructs (and research methodologies) to make them more relevant and appropriate to the specific cultures to be studied. This, in turn, suggests the need for development of taxonomies, nomenclatures, and conceptualizations of affect, cognition, and behavior that are culturally specific. This article begins by discussing early studies in, and conceptualizations of, achievement and achievement motivation. It then discusses more recent theoretical and methodological approaches to the study of these constructs, with particular emphasis being given to their application in cross-cultural settings.

## **1. CROSS-CULTURAL APPROACHES TO ACHIEVEMENT AND ACHIEVEMENT MOTIVATION**

### **1.1. Early Studies in Motivation and Achievement**

The study of achievement motivation owes much to the work of McClelland and associates. They suggested that achievement motivation was a personality trait that developed in some people more than in others as a result of early socialization experiences and the emotional concomitants of these experiences. Because socialization practices were presumed to vary across cultures, McClelland and colleagues’ conceptualization provided a “ready-made” theoretical basis for the examination of achievement motivation in a range of groups, societies, and cultures. As a result, McClelland and colleagues’ work gave impetus to many research programs investigating a range of issues related to achievement motivation in cross-cultural settings.

Specifically, early cross-cultural studies inspired by McClelland and colleagues typically examined the antecedent-consequent variables associated with the development of achievement motivation among diverse social and cultural groups. Some of these studies were particularly concerned with relationships between achievement motivation and variables such as racial/cultural identity, self-esteem, and academic achievement. Other studies were concerned with the sociocultural antecedents of achievement motivation,

including independence training, social class, and acculturative stress.

### **1.2. Early Conceptualizations of Achievement and Achievement Motivation**

McClelland and colleagues originally suggested that achievement behavior consisted of four distinct but related elements: competition with a standard of excellence, affective concern for goal attainment, an evaluation of performance, and some standard for the attainment of a long-term goal. These elements were typically identified in the projective responses to selected thematic apperception test (TAT) cards. In general, this approach categorized individuals as high or low in achievement motivation and determined those situations that would maximize their performance. Atkinson and colleagues introduced the concept of person (trait)/situation (state) interactions and explored the ways in which these interactions influence motivation and achievement. For example, Atkinson and colleagues’ research suggested that high-achievement need individuals are likely to be motivated by achievement situations that are moderately challenging, that is, where success is perceived to be possible but requires effort. Alternatively, low-achievement need individuals are likely to be motivated in situations that are minimally challenging and, hence, require little perceived effort for success. Moreover, low-achievement need individuals may also be attracted to maximally challenging situations where perceived effort is great but where a reasonable “excuse” for failure (i.e., situation difficulty) is inherent. Atkinson and colleagues’ research culminated in the formulation of expectancy-value theory, which systematically assessed the interaction between the value that individuals placed on perceived outcomes from achievement situations (value) and how likely they considered those outcomes to be (expectancies).

Overall, the research of McClelland and Atkinson and their colleagues suggested that achievement-oriented persons emphasized individual responsibility for their actions, were concerned with medium- to long-range goals, attributed success to their own abilities, were self-confident, and obtained most satisfaction from competitive situations where the outcomes were uncertain. Low achievement-oriented persons were characterized as being more motivated to avoid failure than to achieve success and as being most motivated by tasks that minimized the possibility of experiencing affect of failure.

This theme has been pursued up to, and including, studies by Bright in 1994 and Conchas in 2001 on student motivation and achievement.

Within its cultural context, the conceptualisation of McClelland and Atkinson and their colleagues was (and, to some extent, still is) quite an adequate interpretation of what constitutes achievement motivation. However, few studies questioned the basic premises of this conceptualization in terms of its cross-cultural validity. In contrast, most studies assumed that the conceptualization, its premises, and its associated methodologies were essentially etic (culturally transferable). Thus, when measures of differences between cultural groups on achievement and achievement motivation constructs were apparent, it was assumed that these reflected significant differences between groups on the constructs themselves rather than reflecting the applicability (e.g., appropriateness, salience) of the constructs across cultural groups. Interestingly, the non-Western cultural groups being examined often performed poorly with respect to given constructs. The conclusion drawn from such research was that “poor” performance reflected either (a) a deficiency in the group being examined (e.g., genetic inferiority, inadequate socialization for achievement) or (b) an incongruence and incompatibility between the values and norms held by the dominant cultural group and the assumptions, norms, values, and behaviors of the minority or culturally different group.

Perhaps more important, however, such studies often had the effect of “elevating” the goals, perceptions, and behaviors of Western cultural groups to the status of universal norms. In terms of cross-cultural work, this approach increasingly became recognized as both misguided and inadequate and has sometimes been termed “pseudo-etic.” For this reason, more recent cross-cultural studies have explicitly recognized the potential impact of a variety of specific cultural components on achievement and achievement motivation. This culture-specific research is often referred to as emic research. As a result of this research, a significant amount of evidence has accumulated, indicating that modes of expression, achievement, and achievement motivation are not culturally invariant but rather very much related to, and embedded in, cultural contexts. In particular, the emphasis placed on individual achievement and success in McClelland and colleagues’ theory is now questioned extensively. It is now argued that achievement must be analyzed in its total cultural context, including the various roles, both economic and noneconomic, through which achievement motivation may be directed. A wide range of studies in many different cultural settings

support this view. A common theme identified in these studies is that socialization patterns identified as important to developing achievement motivation in Western contexts (e.g., independence, self-reliance) are clearly not manifest (or, at least, not manifest in the same ways or for the same reasons) in other cultural contexts. From this style of research, researchers are able to extract common features that appear to have universal validity and then compare cultures on the basis of these features. This approach is often referred to as etic research.

## 2. TOWARD A BASIS FOR INVESTIGATING ACHIEVEMENT AND ACHIEVEMENT MOTIVATION ACROSS CULTURES

Quite early in the development of cross-cultural research into achievement and achievement motivation, Kornadt and colleagues in 1980, in accord with Maehr in 1974, suggested that, in spite of the many potential configurations of achievement and achievement motivation, some general characteristics of achievement in all cultures may be identified, thereby emphasizing the etic dimensions of achievement motivation. They suggested that these characteristics might serve as a potential basis for cross-cultural work on achievement motivation. These components include the following:

- The existence of a standard of excellence for an individual’s own goal-directed behavior
- The existence of affective reactions to success and failure (e.g., being proud and happy, being disappointed and sad)
- Individual feelings of responsibility for the outcome of achievement-related activity
- Personal incentives for achievement based on security or insecurity about an individual’s own capacity or ability to succeed

Kornadt and colleagues suggested in 1980 that the culture-bound experiences, cognitive structures, and values of each individual will determine the way in which these components are manifest in particular cultures but that the components themselves are relatively robust across cultures. Thus, achievement-motivated behavior will appear in all cultures—but in different forms, stimulated by different cues, and directed at different goals.

The preceding framework acted (and still acts) as a useful starting point for investigating achievement and

achievement motivation across cultures. However, this framework needed to be supported by more detailed conceptualizations of achievement and achievement motivation. In response to this need, many theoretical models that attempted to define such conceptualizations more explicitly emerged. Some of these models continue to be particularly useful for research in applied cross-cultural settings. Four of these are outlined briefly in the following subsections: personal investment (PI) theory, ecological theory, attribution theory, and goal theory.

### 2.1. Personal Investment Theory

Maehr and Braskamp's PI model of achievement motivation builds on and integrates various dimensions from earlier conceptualizations of the nature of motivation that are still considered to be important in any explanation and analysis of motivation, particularly in cross-cultural settings. PI theory is concerned with how persons choose to invest their energy, talent, and time. PI theory is particularly relevant for investigations into how individuals of various cultural backgrounds relate to different achievement situations. This is because it does not assume that people from a given culture will invest effort in the same achievement situations or, if they do, for the same reasons as will those from other cultures. PI theory addresses the apparent paradox between the low assessment of individuals from non-Western cultural groups on achievement tests and their obvious achievement and motivation across a range of activities as illustrated by persistence, level of activity, high performance, and continuing motivation in chosen areas of behavior.

PI theory also emphasizes the role played by social and cultural contexts in determining motivational patterns in performing achievement tasks. Moreover, it is phenomenologically based and emphasizes the subjective meanings of situations in light of individuals' culturally determined belief systems such as beliefs about self, perceptions of appropriate goals, and perceived alternatives available for pursuing these goals.

PI theory is a social cognitive theory in that it assumes that the primary antecedents of choice, persistence and variations in activity levels, are thoughts, perceptions, and beliefs that are embedded in cultural and social beliefs about self and situation. Specifically, PI theory designates three basic components of meaning as critical to determining PI in specific situations:

- Beliefs about self, referring to the more or less organized collections of perceptions, beliefs, and feelings related to who one is
- Perceived goals of behavior in given situations, referring to the motivational focus of activity, particularly what the person defines as success and failure in these situations (e.g., task, ego, social solidarity, extrinsic rewards)
- Perceived alternatives for pursuing these goals, referring to the behavioral alternatives that a person perceives to be available and appropriate (in terms of sociocultural norms that exist for the individual) in a given situation

Each of these components of PI theory may be influenced differentially by the structure of tasks and situations, personal experience, and access to information, as well as, importantly, the sociocultural context in which tasks, situations, and persons are embedded.

As a model, PI theory alleviates many of the problems inherent in monocultural research models. In particular, it conceptualizes achievement motivation in terms that recognize the possibility of diverse modes of achievement behavior across cultures. PI theory also strikes a balance between personality and situation while incorporating dimensions (e.g., locus of control) that have been found to be useful in analyzing levels of achievement motivation. Studies by McInerney and colleagues in 1995 and 1997 illustrated the use of PI theory in applied cross-cultural settings.

### 2.2. Ecological Theory

At about the same time as PI theory was being developed, Berry in 1980 attempted to redress the balance in cross-cultural research between laboratory-based programs and research set in naturalistic environments by proposing an ecological model for cross-cultural studies. This model (or at least its general approach) continues to be relevant for contemporary cross-cultural studies. Berry suggested that research that is based largely on experimentation, with little consideration of the wide range of variables operating in natural environments, holds little value. To make the task of effective naturalistic research easier, Berry proposed a multilevel model. This model describes research that moves from a naturalistic-holistic level to a controlled reductionistic level through (a) four environmental contexts (ecological, experiential, performance, and experimental) and (b) four effects (achievements, behaviors, responses, and scores).

The ecological context (also referred to as the natural-cultural habitat) refers to all of the relatively permanent characteristics that provide a context for human action. The experiential context is the pattern of recurrent experiences that provide a basis for learning. The performance context is the limited set of environmental circumstances, immediate in space and time, that may be observed to account for particular behaviors. The experimental context represents those environmental characteristics that are designed by the psychologist to elicit a particular response or test score. The experiential and performance contexts are always nested within the ecological context, whereas the experimental context may or may not be contained within the first three contexts. Berry suggested that the degree to which the experimental context is nested within the other three contexts represents the ecological validity of the experimental task.

Four effects parallel the four contexts just described: achievements, behaviors, responses, and scores. Achievements refer to the complex, long-standing behavior patterns that develop as an adaptive response to the ecological context and include established and shared patterns of behavior that can either be found in an individual or distributed in a cultural group. Behaviors are those acts that have been learned over time in the experiential context and include the skills, traits, and attitudes that have been acquired in particular roles or fostered by specific training or education. Responses, on the other hand, are those 'fleeting' performances that occur in response to immediate stimulation or experience. Finally, scores are those behaviors that are observed, measured, and recorded during a psychological experiment or testing. Berry argued that between each of the four environmental contexts and effects exists a relationship that should be considered in any effective cross-cultural research.

The complexity of Berry's model may have limited its usefulness to some researchers. However, the general ecological approach has proven to be very useful in recent research investigating motivation, achievement, and related constructs. Specifically, such research has arguably produced data with acceptable ecological validity and has specified a range of experiential, performance, and experimental variables that are associated with motivation and achievement across a range of "natural" habitats.

### 2.3. Attribution Theory

The problem of developing an attributional model that might form a more adequate basis for cross-cultural

research into achievement motivation has been a focus of older studies as well as more recent studies. These studies suggest that achievement motivation should be analyzed in terms of the subjective meaning of attributions (i.e., reasons given for success or failure) of a group or the persons who comprise that group. In other words, research into achievement motivation should be concerned with the phenomenology (i.e., definitions in context) of achievement attributions. This phenomenological approach to defining attributions cross-culturally requires that researchers elicit attributions of achievement behavior from people of different cultures rather than imposing inappropriate definitions on their achievement behavior.

In Western society, research indicates that assessments of achievement in terms of success and failure are best thought of as psychological reactions to outcomes and not as objective outcomes themselves. Thus, successes and failures are defined as much in terms of subjective feelings attached to given outcomes as in terms of the nature of the outcomes themselves. Moreover, attributions for success or failure that produce the greatest feelings of pleasure or regret are those related to ability (e.g., "I'm clever; therefore, I succeeded") and effort ("I tried hard; therefore, I succeeded"). More generally, attributions that enable a person to demonstrate to self or others that he or she possesses desirable qualities are considered to be important motivators. However, research concerned with examining the antecedents of success and failure across cultures suggests that ability and effort attributions are not necessarily salient to all groups. Hence, where research based on attribution theory emphasizes ability and effort as the most important perceived causes of perceived success and failure, it may be culture specific. For cross-cultural researchers, then, it is necessary to discover attributions that are salient to different cultural groups in eliciting feelings of success or failure.

### 2.4. Goal Theory

Goal theory developed by Ames, Dweck, Nicholls, Pintrich, and Midgley, provides a model that enables researchers to identify similar motives in diverse cultures, even if such behavior may vary in frequency and importance across cultures. Goal theory emphasizes the importance of goals and definitions of self in relationship to such goals. Goal theory differs from attribution theory in that attribution theory explores perceived reasons for success or failure while trying to succeed, whereas goal theory explores perceived reasons for trying to succeed

in the first place. These reasons for trying to succeed are labeled “goals” in the context of goal theory. Some important goals (or goal “orientations”) postulated by goal theory are the following:

- *Ability-oriented (performance) motivation.* This approach is characterized by striving to maintain a favorable perception of one’s ability to demonstrate competence. That is, the goal is to maximize the subjective probability of attributing high ability to oneself in culturally relevant areas. This approach, as with all goals in goal theory, focuses on the goal or function of behavior for the person rather than on the inevitably ambiguous external form of any given behavior itself.

- *Task-oriented (mastery) motivation.* This approach is characterized by behavior where the focus is on producing a product or solving a problem rather than on demonstrating one’s competence. In contrast to ability-oriented behavior, during task-oriented behavior, people are not attributing outcomes to ability or seeking to demonstrate ability. The completion of the task itself is its own reward, and the development of mastery is perceived as intrinsically satisfying.

- *Socially oriented (social) motivation.* This approach is characterized by behavior where the goal is to maximize the probability of attributing high effort to oneself (and to minimize the probability of attributing low effort to oneself) in conformity with perceptions of the probability of gaining social approval, respect, or recognition. This orientation, which is really a class of orientations, is based on the assumption that effort, unlike ability, is voluntary and is something that anyone can display. Therefore, it can indicate conformity to social norms or virtuous intent, and this in turn can lead to social approval, respect, or recognition. Conversely, lack of effort in conforming to social norms or virtues may result in reduced approval, respect, or recognition and, hence, is to be avoided.

Goal theory has provided, and continues to provide, a cogent and coherent platform for investigating cross-cultural issues in achievement and achievement motivation, especially where the purposes for achievement are the focus of investigation. This is because goal theory allows for a range of possible intents (e.g., purposes, reasons) for achievement and for the possibility that the range and salience of possible intents may vary considerably from group to group. As a result, significant cultural differences in goal orientations, and their impact on motivation and achievement, have been noted in a range of studies. However, there is also contrary evidence from studies by McInerney and others

suggesting that the goals that people espouse, and the effects of these goals on motivation and achievement, may be relatively stable across cultures.

### 3. APPLYING SOCIAL COGNITIVE CONCEPTUALIZATIONS OF MOTIVATION AND ACHIEVEMENT

In general, social cognitive theories of motivation, particularly goal theory and PI theory, have alerted researchers to the ways in which motivation and achievement may be differentially constructed across cultural (and other) contexts. However, a significant remaining question from this perspective is the following: Is there a single motivational environment that can be constructed in most applied settings (e.g., schools, hospitals, businesses, government organizations), especially when these settings are typically diverse in many ways with respect to individuals’ ages, genders, cultural backgrounds, and socioeconomic status? This is an important question because a range of different motivational environments cannot be easily generated in most applied settings. Theory and research have gone a considerable distance in demonstrating that mastery goals are preferable to ability performance goals in a wide range of situations. This is because mastery goals focus on an individual’s attention to the task(s) to be completed rather than on the individual’s perceived ability (or lack thereof) to complete the task(s). Psychologically, this appears to “free” individuals to perform at optimal levels, even in the face of difficulties. Indeed, it might also be that performance goals are dysfunctional in many applied settings, although not in all cases. More recently, social goals have also been shown to have adaptive effects on an individual’s engagement in a variety of achievement and social situations. These social goals may be particularly relevant in non-Western cultural settings, where the social dimensions of motivation and achievement may be more salient than in Western settings.

### 4. METHODOLOGICAL IMPROVEMENTS IN APPLIED SETTINGS

The preceding sections of this article have outlined (a) the need for a cross-cultural understanding of achievement and achievement motivation and (b) some relevant

conceptualizations and theories that may support research into achievement and achievement motivation from cross-cultural perspectives. This section deals with methods for researching these culturally variable constructs in appropriate ways in applied settings (e.g., schools, workplaces, sports settings).

It is obviously not acceptable to use “one size fits all” approaches when researching achievement and achievement motivation across cultures. Rather, culturally sensitive studies in achievement and achievement motivation are needed for the following purposes:

- To establish whether the behaviors and responses being examined across cultures are functionally, conceptually or metrically equivalent
- To establish that the constructs and tools used in research are culturally appropriate to all groups in the study
- To examine achievement and achievement motivation in “real life,” as opposed to laboratory or similar situations, whenever possible
- To consider in detail, at both theoretical and methodological levels, the potential mediating processes that might intervene among culture, achievement motivation, and achievement
- Not to conceptualize achievement orientation solely as an individual personality-based characteristic
- Not to use cultural deficit and deprivation models (or, at least, not without justification) to explain the differential performance of various cultural groups
- Not to define any culture too narrowly and fail to recognize the range of experiences, perceptions, and behaviors represented in any given culture

Following these steps should help to avoid pseudo-etic (e.g., Western individualist and competitive) ways in which to conceptualize achievement and achievement motivation in cultures where these values might not apply. At least these steps should contribute to a viable framework for examining achievement and achievement motivation cross-culturally in that they provide explicit procedures for developing culturally specific contents that may be applied in different cultural settings with comparable theoretical meanings.

#### **4.1. Informal Assessment of Motivation in Applied Settings**

Formal measurement of motivation (i.e., using psychometric instruments) occurs relatively rarely in applied settings such as schools and workplaces. However,

informal assessments are made regularly in applied settings, particularly by those responsible for “performance management” (e.g., managers, supervisors, teachers). Informal indicators of motivation typically include a demonstrated willingness to be involved in assigned activities, the selection of appropriate activities where choice is available, the level of intensity or involvement in given activities, and the quality of performance outcomes.

These informal indicators of motivation are assessed or determined, to a large extent, by the frame of reference of the individual (typically a superior) making a determination. In other words, teachers, managers, and supervisors consider their subordinates to be more or less motivated in terms of the superiors’ frames of reference. This process of evaluation, however, may cause problems in diverse settings (particularly in cross-cultural settings). For example, motivation in the workplace is often conceptualized by supervisors in terms of material or technical output (i.e., productivity). However, what happens when output is conceptualized (by subordinates) in “unconventional” ways, for example, by improving the relational climate of the workplace or classroom and by contributing to the work of others? In such cases, supervisors may consider subordinates to be “unmotivated” when the subordinates consider themselves to be highly motivated. A similar point could be made with respect to the balance between fulfilling social responsibilities (e.g., looking after family and friends) and fulfilling more conventional responsibilities (e.g., being on time for work, being available for after-work/after-school functions).

There are two major points being made here. First, motivation and the demonstration of motivated behavior may be conceptualized very differently by different cultural (or other) groups. Second, evaluations of motivation and motivated behavior are dependent on one’s frame of reference. This adds considerable complexity to the evaluation of motivation in applied settings and argues against defining motivated behavior too narrowly, especially as the cultural diversity or other diversity of any setting increases.

## **5. CONCLUSION**

This article has reviewed basic approaches to cross-cultural studies of achievement and achievement motivation, especially in applied settings such as schools, classrooms, and workplaces. It did so by highlighting some major limitations of the concepts and methods used to guide early studies while still recognizing the

value of this early work. It also presented a range of theoretical and methodological options that are currently being developed to enhance cross-cultural research in general and cross-cultural research into achievement and achievement motivation in particular.

A perennial problem for cross-cultural research is generating data that are universally comparable while also being culturally valid. Recent studies have demonstrated much more sophisticated and sensitive approaches to the issues of achievement and achievement motivation in cross-cultural settings.

### See Also the Following Articles

Motives and Goals ■ Subjective Culture

### Further Reading

- Ames, C. (1984). Competitive, cooperative, and individualistic goal structures: A motivational analysis. In R. Ames, & C. Ames (Eds.), *Research on motivation in education*, Vol. 1: *Student motivation* (pp. 177–207). San Diego: Academic Press.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261–271.
- Atkinson, J. W., & Feather, N. T. (Eds.). (1966). *A theory of achievement motivation*. London: Wiley.
- Atkinson, J. W., & Raynor, J. O. (1974). *Motivation and achievement*. Washington, DC: V.H. Winston.
- Berry, J. W. (1980). Ecological analyses for cross-cultural psychology. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 2, pp. 157–189). London: Academic Press.
- Bright, J. A. (1994). Beliefs in action: Family contributions to African-American student success. *Equity and Choice*, 10, 5–13.
- Conchas, G. Q. (2001). Structuring failure and success: Understanding the variability in Latino school engagement. *Harvard Educational Review*, 71, 475–504.
- Conner, R. F. (1998). Toward a social ecological view of evaluation use. *American Journal of Evaluation*, 19, 237–241.
- Gomez, M. J., & Fassinger, R. E. (1994). An initial model of Latina achievement: Acculturation, biculturalism, and achieving styles. *Journal of Counseling Psychology*, 41, 205–215.
- Graham, S., & Taylor, A. (2002). Ethnicity, gender, and the development of achievement values. In A. Wigfield, & J. Eccles (Eds.), *The development of achievement motivation* (pp. 121–146). San Diego: Academic Press.
- Graham, S., Weiner, B., & Benesh-Weiner, M. (1995). An attributional analysis of the development of excuse-giving in aggressive and nonaggressive African American boys. *Developmental Psychology*, 31, 274–284.
- Johnson, G. M. (1994). An ecological framework for conceptualizing educational risk. *Urban Education*, 29, 34–49.
- Kaplan, A., & Maehr, M. L. (1999). Enhancing the motivation of African American students: An achievement goal theory perspective. *Journal of Negro Education*, 68, 23–41.
- Kornadt, H. J., Eckensberger, L. H., and Emminghaus, W. B. (1980). Cross-cultural research on motivation and its contribution to a general theory of motivation. In H. C. Triandis & W. Lonner (Eds.), *Handbook of cross-cultural psychology* (Vol. 3, pp. 223–320). Boston: Allyn & Bacon.
- Maehr, M. L., & Braskamp, L. A. (1986). *The motivation factor: A theory of personal investment*. Lexington, MA: Lexington Books.
- Maehr, M. L. (1974). Culture and achievement motivation. *American Psychologist*, 29, 887–896.
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The achievement motive*. New York: Appleton–Century–Crofts.
- McClelland, D. C., Baldwin, A. L., Bronfenbrenner, U., & Strodbeck, F. L. (1958). *Talent and society: New perspectives in the identification of talent*. Princeton, NJ: Van Nostrand.
- McInerney, D. M. (1994). Psychometric perspectives on school motivation and culture. In E. Thomas (Ed.), *International perspectives on culture and schooling*. London: London University, Institute of Education.
- McInerney, D. M. (1995). Goal theory and indigenous minority school motivation: Relevance and application. In P. R. Pintrich, & M. L. Maehr (Eds.), *Advances in motivation and achievement* (Vol. 3, pp. 153–181). Greenwich, CT: JAI.
- McInerney, D. M., Roche, L., McInerney, V., & Marsh, H. W. (1997). Cultural perspectives on school motivation: The relevance and application of goal theory. *American Educational Research Journal*, 34, 207–236.
- McInerney, D. M., & Swisher, K. G. (1995). Exploring Navajo motivation in school settings. *Journal of American Indian Education*, 34(3), 28–51.
- McInerney, D. M., Yeung, S. Y., & McInerney, V. (2001). Cross-cultural validation of the Inventory of School Motivation (ISM). *Journal of Applied Measurement*, 2, 134–152.
- Moeller, A., & Kramer, E. (1995). The teaching of culture and language in the second language classroom: Focus on the learner. *International Journal of Educational Research*, 23, 571–652.
- Neuman, S. B., & Celano, D. (2001). Access to print in low-income and middle-income communities: An ecological study of four neighborhoods. *Reading Research Quarterly*, 36(1), 8–26.
- Niles, S. (1998). Achievement goals and means: A cultural comparison. *Journal of Cross-Cultural Psychology*, 29, 656–667.
- Ogbu, J. U., & Simons, H. D. (1998). Voluntary and involuntary minorities: A cultural–ecological theory of school

- performance with some implications for education. *Anthropology & Education Quarterly*, 29, 55–88.
- Salili, F., & Hoosain, R. (2002). Cross-cultural differences in affective meaning of achievement: A semantic differential study. In D. M. McInerney, & S. Van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (Vol.2). Greenwich, CT: Information Age.
- Strohschneider, S., & Guss, D. (1998). Planning and problem solving: Differences between Brazilian and German students. *Journal of Cross-Cultural Psychology*, 29, 695–716.
- Villani, C. J. (1999). Community culture and school climate. *School Community Journal*, 9, 103–105.
- Watkins, D., McInerney, D. M., Lee, C., Akande, A., & Regmi, M. (2002). Motivation and learning strategies. A cross-cultural perspective. In D. M. McInerney, & S. Van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (Vol. 2). Greenwich, CT: Information Age.
- Weiner, B. (1994). Integrating social and personal theories of achievement striving. *Review of Educational Research*, 64, 557–573.



# Motivation to Diet and Exercise

Larry M. Leith

University of Toronto, Toronto, Ontario, Canada

1. Introduction
2. Motivational Considerations in Diet and Exercise
3. Individual Intervention Strategies to Improve Diet and Exercise
4. Concluding Remarks  
Further Reading

## GLOSSARY

- behavior therapy** A field of psychology that provides information telling people how to change their individual habits; it involves self-control to make people's lives better.
- motivation** Term used to explain why people do, or do not do, certain behaviors.
- obesity** When an individual has a body mass index (BMI) of at least 30.
- protection motivation theory** Theory suggesting that people are motivated to protect themselves from harmful or stressful life events.
- self-determination theory** Viewpoint evaluating the relative importance of intrinsic and extrinsic rewards for human behavior.
- self-efficacy** Term referring to situation-specific self-confidence.

People are continually trying to exercise more and eat diets that are conducive to good health. The problem of childhood and adolescent obesity continues to be a major focus of concern for health care professionals in North America. This article describes the necessary

tools to make lifestyle changes with the potential to reverse this troublesome trend.

## 1. INTRODUCTION

The problems associated with obesity and chronic degenerative diseases, such as heart disease, arteriosclerosis, hypertension, osteoarthritis, and type 2 diabetes, are now reaching epidemic proportions. The prevalence of obesity, where an individual has a body mass index (BMI) of at least 30, is increasing worldwide. Currently, there are approximately 300 million obese individuals around the world. More alarmingly, in the United States, approximately 30% of all adults are categorized as obese. Despite this prevalence being somewhat lower in Canada, the rate has increased significantly there during recent years as well. In 2003, Bouchard reported that obesity costs the unregulated U.S. health care system approximately \$100 billion annually. Therefore, it is time to take a serious look at how to reverse this troublesome trend.

In 2000, data collected by the U.S. Department of Health revealed that nearly 30% of the population is physically inactive. It is interesting to note how this percentage corresponds exactly with the percentage of individuals who are considered to be obese. Also, an additional 40 to 50% of the U.S. population does not engage in the necessary amount of physical activity to achieve health benefits. Through a process known as tracking, these statistics are reported to have remained

relatively consistent for the past 10 years or so. This is unfortunate given that frequent physical activity has been shown to benefit every system of the human body. For example, exercise physiologists have documented several important changes in the body that accompany participation in an ongoing exercise program. These include a decreased chance of heart disease and stroke, an increase in good (HDL) cholesterol, a decrease in bad (LDL) cholesterol, a lowered risk of osteoporosis (i.e., brittle bones), a decreased percentage of body fat, an increased lean muscle mass, lower blood pressure, and an improved immune system. In addition, in 1998, Leith reported that regular participation in physical activity has been found to be associated with several psychological benefits such as lowered anxiety, lowered depression, improved self-concept, and better overall mood.

In 1996, Manley, the former U.S. surgeon general, made the following observation. Because physical activity is so effective in preventing diseases, preventing premature death, and promoting a high quality of life, it should be accorded the same level of attention that is given to other important health practices that affect the nation. This assertion suggests that there is absolutely no doubt that physical activity can be described as efficacious in the prevention and treatment of many medical conditions, including obesity. However, despite this and extant research highlighting the benefits of physical activity and a healthy diet, people remain inactive and eat poorly. Therefore, it is important to consider the “why” of this phenomenon. The next section summarizes contemporary motivational theories relating to these topics and examines their relevance for influencing the health behaviors of diet and exercise.

## 2. MOTIVATIONAL CONSIDERATIONS IN DIET AND EXERCISE

The term “motivation” is actually a derivative of the Latin word *movere*, meaning “to move.” It is the psychological term used to explain why people do, or do not do, certain things. For example, why do some individuals exercise regularly and eat healthy diets, whereas others choose the options of inactivity and eating junk foods? This is not an easy question to answer, but the bottom line is that people’s underlying motivation can be understood by looking at three characteristics of behavior. First, people’s selection of certain activities over others

indicates something about their motivation. To paraphrase, what is it that is important to people personally? Their answers to that question will indicate their likelihood or reluctance to engage in a specific behavior. Second, the effort that people put into any activity also reveals their underlying motivation. In other words, the more effort people put into any behavior, the more motivated they are to accomplish that goal. The harder people work toward accomplishing a task, the more they want it. Finally, the degree of persistence that people demonstrate in their pursuit of a goal also reveals their level of motivation. If they refuse to give up and really want it but still give up easily, they did not want it enough in the first place. Therefore, it is important that each and every individual is aware of these three benchmarks of motivational theory.

Now that we have examined the basic components of motivation, we look at theories of motivation that have specific relevance for dieting and exercise. It is far beyond the scope of this article to cover the whole range of motivational theories. In 2003, Carron and colleagues published a book summarizing a wide range of motivational theories. This section focuses only on self-efficacy theory, protection motivation theory, and self-determination theory. In each case, a summary of the theory is provided with specific implications for diet and exercise.

### 2.1. Self-Efficacy Theory

The term “self-efficacy” can be summarized with an individual asking himself or herself the following two questions: “Am I capable of doing this activity?” and “Will I be successful in achieving my goals if I engage in this behavior?” The internal question specific to this article would be “Am I capable of successfully participating in an ongoing diet and exercise program?” Self-efficacy can also be described as situation-specific self-confidence.

In 1997, Bandura reported that an individual’s concept of self-efficacy is molded by his or her past experiences, vicarious experiences, verbal persuasion, and physiological/emotional states. To summarize this theory in light of the article’s objectives, self-efficacy can be enhanced in a variety of ways. First, people should remember to take “baby steps.” In other words, they should start an exercise program slowly and try to improve their overall diet in small stages. Most people fail because they try to do too much too soon. Second, people should take advantage of watching demonstrations, television shows, and videos that show other people succeeding in their efforts. These options will improve people’s belief in their ability to succeed in

their quest. Third, people should learn everything they can about the exercise and diet topics. More information is always better. They should not be shy about soliciting support from their significant others; doing so can go a long way toward improving their chances of success. Finally, people should be aware of how their physiological/emotional states can influence their resolve to diet or exercise. It is only natural to question the increased heart rate associated with exercise and/or the effect of mood on exercise or diet compliance. People should become familiar with these outcomes and recognize them as harbingers of success in their future efforts.

## 2.2. Protection Motivation Theory

This theory suggests that people are motivated to protect themselves from harmful or stressful life events. It also considers coping strategies for these occurrences. To summarize, the protection motivation theory maintains that an individual's motivation to engage in a health-related behavior is influenced by two considerations. First, threat appraisal represents a conscious evaluation of a person's perceived vulnerability and the possible severity of a medical problem. The internal questions ask "Am I at risk, and how serious are the possible outcomes of not exercising and of eating a poor diet?" Second, coping appraisal reflects an individual's perception of whether he or she will be successful in implementing the appropriate behavior and, if so, whether these actions will remove the threat. The internal questions ask "Am I capable of engaging in regular exercise and maintaining a healthy diet and, if so, will I be less at risk for developing chronic degenerative diseases?" The answers to these questions will provide a good indication of the person's intention to improve exercise and diet habits. Intention remains a critical ingredient in predicting human behavior.

In 2000, Floyd and colleagues and Milne and colleagues published meta-analytic reviews suggesting the efficacy of the protection motivation theory in predicting health behaviors. The former study examined 65 published articles with 29,650 participants and concluded that increased threat severity, threat vulnerability, self-efficacy, and response efficacy resulted in facilitated health behaviors. The latter study reviewed 29 studies with 7700 participants and also concluded that the protection motivation theory is predictive of facilitated adaptive intentions and improved health behaviors. Therefore, it appears that an individual's perceived severity of threat and vulnerability from inaction, as well as the chances of being successful in changing

these health behaviors, will seriously affect the desire to exercise more and to maintain a healthy weight.

## 2.3. Self-Determination Theory

This theory evolved in an attempt to understand the relative importance of intrinsic and extrinsic rewards for human behavior. Originally, it was believed that if intrinsic rewards were present, extrinsic rewards would be absent, but empirical research disproved this hypothesis.

Deci's 1992 development of self-determination theory attempted to address the discrepancy. This approach maintains that extrinsic and intrinsic motivations exist on a continuum. At one end of the continuum is amotivation, indicating the complete absence of motivation toward engaging in a particular behavior such as exercising or losing weight. At this point on the continuum, the person has no desire to change his or her lifestyle. Extrinsic motivation lies in the middle of the continuum and is multidimensional in nature. It consists of three additional dimensions. Consider the scenario where a family doctor tells a patient that he or she must lose weight and start exercising to avoid serious medical problems. In this case, the patient may try to follow the doctor's advice simply to avoid illness. Slightly farther along the continuum, the patient may lose enough weight and exercise to no longer be considered a health risk, and the patient continues the program only because he or she believes that it must be done. Finally, the individual may continue to "count calories" and exercise three times per week, but only because doing so is necessary to avoid illness.

At the farthest point on the continuum lies intrinsic motivation, and it is also multidimensional. This type of motivation occurs when a person performs a behavior for the sole pleasure it provides. Intrinsic motivation is also a multidimensional construct. For example, acquiring the knowledge of how to count calories effectively, eat a more healthy diet, and perform a new exercise represents the first stage of intrinsic motivation. During the second stage, the satisfaction of losing weight and becoming a regular exerciser becomes more important. Finally, once an individual reaches the intrinsic motivation point on the continuum, he or she will continue to exercise solely for the pleasant sensations associated with the physical activity. Likewise, the person will continue to maintain a healthy weight simply because it feels good to do so.

Now that we have examined the theories of motivation most relevant to diet and exercise, it is time to learn how to apply this important knowledge. The next

section discusses the behavioral tools that people need to take control of their lifestyle choices.

### **3. INDIVIDUAL INTERVENTION STRATEGIES TO IMPROVE DIET AND EXERCISE**

The field of behavior therapy provides important information concerning the making or breaking of individual habits. Practitioners who endorse this approach maintain that people often fail in their self-control efforts because they do not consider environmental factors or the “ABCs of behavior.” Environmental stimuli (i.e., cues) that affect people’s behavior represent the As. The behaviors themselves are termed the Bs, and the rewards or punishments that closely follow the behaviors are labeled the Cs. The following specific interventions use this ABCs approach to behavior change.

#### **3.1. Strategy 1: Removing Potent Cues That Are Associated with a Problem Behavior**

One serious handicap to the average weight control program is the typical pairing of eating and snacks that are readily visible. Not many people can walk by a bowl of chips or a box of chocolates without sampling these treasures. For this reason, if people are attempting to lose weight, they should not have the fattening snacks showcased in a well-traveled area of the house. In terms of exercise, many individuals report that televisions and computers are major impediments to being physically active. If this is a problem for people, they should make a deal with themselves not to walk by their televisions or computers until they have completed their exercise sessions. This will eliminate the automatic pairing of watching television or surfing the Internet with inactivity.

#### **3.2. Strategy 2: Making the Problem Behavior Unlikely or Impossible**

Following from the previous example, the simplest way not to indulge in high-calorie snacks is simply not to have them in the house. This way, if people feel that they are weakening, it takes a good deal of effort to make a special trip to the store, so they are less likely to do so. When people shop for groceries, they should not buy these fattening foods. Similarly, if people find that they fail to exercise on days when they go to a tavern or to a

friend’s place, they should make an agreement with themselves to avoid these venues on exercise days. People will still have many opportunities to enjoy these situations that are getting in the way of their exercise programs.

#### **3.3. Strategy 3: Developing New Behaviors to Replace Maladaptive Ones**

This intervention invites people to identify the existing connection between the usual cues to action and the problem behaviors. For example, if people find that they tend to overindulge in snacking when they are feeling angry or lonely, they should consider an alternative behavior. Many individuals report that writing down their thoughts or feelings on a pad of paper can be therapeutic. If people are trying to lose weight, this is a better option than snacking. Taking this example to the next level, people might want to consider going for a walk or going to a fitness facility to deal with their feelings. Exercise has been shown to improve mood, and exercising in a social setting can help to alleviate feelings of anger or loneliness.

#### **3.4. Strategy 4: Developing Cues That Will Foster a Desired Behavior**

People should think in terms of using environmental cues that can work for them in pursuit of their goals. Individuals who want to lose weight should consider taping pictures of their dream tropical island getaway destinations on the fronts of their refrigerators. This way, each time they are tempted to snack, they can visualize themselves looking good in their swimwear on the beach. This will give them the needed strength to resist temptation. People who want to exercise regularly should consider placing their walking/running shoes and/or sweat suit at the point of entry into their residences or on their bedposts. This way, the first thing that they see when they get home from work or wake up in the morning is exercise equipment. This will prove to be a strong reminder for them to walk and/or jog.

#### **3.5. Strategy 5: Rewarding Oneself for Exhibiting a Good Behavior**

The whole idea of whether or not people should reward themselves is contingent on the principle of their success in achieving the target behavior. In other words, if

people maintain their daily caloric intake or achieve their exercise goals, they need to give themselves a “pat on the back.” It is important that this reward closely follows the desired behavior. Effective self-control requires a reorientation of the way in which people reward themselves. If they limit themselves to a certain number of calories or engage in their exercise programs on any given day, they are entitled to a reward. People should take this reward sooner rather than later because they deserve it and it will lead to greater compliance.

### **3.6. Strategy 6: Rewarding Oneself with What One Really Wants to Do Anyway**

If there is a behavior that people really like to do on any given day, such as watching television or surfing the Internet, they should use one or both of these activities as potential rewards for achieving their daily goal. So, if people like to watch a favorite television show or use their computers, they should allow themselves to do so only after they have gone for their walks or reduced their caloric intake for that day. This represents a way in which people can reward themselves with things they have already identified as important to them personally.

### **3.7. Strategy 7: Using a Point-Reward System to Motivate Oneself**

Sometimes, achieving a goal can seem a long way off. For this reason, it is important for people to use a point-reward system so as to continue to motivate themselves. For example, if people’s goal is to lose 20 pounds or to be able to jog 2 miles, it is important to remember that these goals can be reached only with “baby steps.” This means that people should set smaller, short-term goals to achieve their desired results. Losing 1 pound per week or walking/jogging a certain distance during that week allows people to give themselves a “point” or “star” for their accomplishments. People should record these points or stars in diaries or on their monthly calendars as a form of self-monitoring. They should make a deal with themselves that once they have accumulated a certain number of points or stars, they can trade these in for something they really want such as a night on the town, a new video, or a CD that was recently advertised. They should continue to set new reward limits and to treat themselves every time they have achieved their short-term goals.

### **3.8. Strategy 8: Being Creative in Finding New Ways in Which to Reward Oneself**

One of the most often overlooked sources of reinforcement is of a social nature, for example, people’s friends, families, and coworkers. People should not be shy about enlisting their families, friends, and coworkers for support. A pat on the back or a “thumbs up” from a significant other can go a long way toward helping people to achieve their behavioral goals of losing weight and exercising more.

### **3.9. Strategy 9: Using a Behavioral Contract to Help Achieve One’s Goals**

Writing up an exercise and/or diet contract can be a very effective strategy. It requires people to put in writing their diet or exercise aspirations. For example, the number of days per week people are going to exercise and the number of daily calories they are going to consume are formalized with the written contract. It is also a good idea to include the mention of a reward that people will receive if successful as well as a response cost if they slack off in their mission. People sign the contract and have significant others also sign the document. This will increase the chances that they will be successful in their efforts.

## **4. CONCLUDING REMARKS**

This article has examined the problems of increasing obesity trends and declining exercise habits in the North American population. Although these problems are indeed factual, it is important to put both topics into a healthy perspective. This article is not intended to suggest that everyone should become obsessive-compulsive about their diets and exercise patterns. The best approach is for people to find a happy medium where they exercise moderately and eat a healthy diet that is low in fats and includes lots of fruits and vegetables. If people buy into this philosophy, the intervention strategies outlined in the preceding section will provide them with the behavioral tools needed to achieve healthier lifestyles. Most experts would suggest that behavior can be changed within a matter of weeks using the behavior therapy approach.

### See Also the Following Articles

Eating, Psychology of ■ Health and Culture ■ Health Psychology, Cross-Cultural ■ Obesity

### Further Reading

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bouchard, C. (2003, September). *Alarming facts and potential causes of the growing obesity epidemic*. Paper presented at the "Obesity: Weighing the Options" symposium, Toronto, Canada.

Carron, A. V., Hausenblas, H. A., & Estabrooks, P. A. (2003). *The psychology of physical activity*. New York: McGraw-Hill.

Deci, E. I. (1992). On the nature and function of motivation theories. *Psychological Science*, 3, 167–171.

Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology*, 30, 407–429.

Leith, L. M. (1998). *Exercising your way to better mental health*. Morgantown, WV: Fitness Information Technology.

Manley, A. E. (1996). Preface. In U.S. Department of Health and Human Services, *Physical activity and health: A report of the surgeon general executive summary*. Washington, DC: U.S. Department of Health and Human Services.

Milne, S., Sheeran, P., & Orbell, S. (2000). Prediction and intervention in health related behavior: A meta-analytic review of protection motivation theory. *Journal of Applied Social Psychology*, 30, 106–143.

U.S. Department of Health and Human Services. (2000). Physical activity and fitness. In *Healthy People 2010*. [www.health.gov](http://www.health.gov).



# Motives and Goals

Uwe E. Kleinbeck

University of Dortmund, Bochum, Germany

1. Introduction
  2. Goal Effects: Mechanisms and Moderators
  3. Developing a Personal Goal
  4. Acceptance of Goals
  5. How to Pursue Goals That Are (Still) Unknown
  6. Managing Successful Goal-Oriented Work Behavior: The Participatory Productivity Management
  7. Competencies in Managing Goal-Directed Behavior
  8. An Integrative Model for the Effects of Motives and Goals
- Further Reading

## GLOSSARY

**competencies** Capabilities needed by leaders for supporting goal setting, translating goals into action, committing to goals, and dealing with failure in managing human resources at work in a goal-directed way.

**dealing with failure** Dealing with not reaching a goal; failure can lead to strengthened approach tendencies with a higher motivation to reach the goal, but it can also result in withdrawal from the task.

**goal acceptance** When acting persons accept goals set by others; if such persons experience positive influence potentials, they have the chance of negotiating new goals.

**goal commitment** The commitment that keeps people persistent in striving for goals; the degree of goal commitment is dependent on changes in current motivation.

**management system** Instruments that are able to measure performance, offer feedback, and stimulate participative goal setting to help people reach organizational goals.

**moderators** Variables that moderate the translation of goals into actions; the most important ones are task complexity, goal commitment, and individual ability.

**motivating potentials** Opportunities in a situation to reach the goals of a motive; the interaction of motivating potentials with motives determines work motivation.

Motives and goals work together to control the intensity, direction, and persistence of behavior at work. Motives are defined as classes of goals that determine, in interaction with motivating potentials of the work situation, the kind of goals people use to control their actions. Goal characteristics, mechanisms, and moderators of the translating process are the determinants of behavioral outcomes. Motives also play a role during this action phase. They affect the degree of goal commitment that people show while performing. Especially interesting is the impact of motives on the way in which people deal with failure. People differ in their ways of reacting to failure. Some are better equipped to deal with failure because they tend to increase motivation, stay with their goal, and finally reach it. Others withdraw from the task.

## 1. INTRODUCTION

Goals are the basic concepts used to understand and describe the direction and persistence of behavior at

work. Goals are effective on various levels of abstractness, with very specific goals controlling specific behaviors (e.g., motor acts). For example, a person can have the goal to press a button 300 milliseconds after a signal appears on a screen. On a more abstract level, goals are aggregated and often are called motives. A motive is defined as a highly aggregated class of goals, with variants of a basic theme being common to all of them. For example, a set of social goals forms a social motive acting as a stable disposition in persons across situations and over long time periods. Examples of other motives when the focus is on the development of personal goals are provided later in this article. In this context, motives are introduced as central determinants of personal goal setting in concrete situational contexts.

Goals can be set by an acting person internally or can be set externally by others. In the following, the focus is on internally set goals. The acceptance of externally set goals is described later in the article. Next, several functional properties of goals and three phases of goal-oriented action processes are discussed, and how goals and motives interact during these phases of action is discussed throughout the article.

Setting goals for oneself means anticipating expected outcomes of action that the person wants to reach. Goals also function as motivational states, directing behavior, influencing commitment to those goals, and keeping action more or less persistent toward those goals. Moreover, goals serve as reference values in a feedback loop in that they are used as standards to

compare behavioral states in the course of actions with the intended goals. Finally, goals trigger the forming of emotions such as pride (in the case of reaching goals) and shame (in the case of failing to reach goals).

## 2. GOAL EFFECTS: MECHANISMS AND MODERATORS

According to Locke and Latham, goals determine the effort that people invest in their actions, goals direct behavioral acts toward the intended outcomes, and goals are responsible for the degree of persistence in keeping actions goal oriented. These are the mechanisms by which goals affect performance and outcomes. The effect of goals is moderated by a set of variables, and in this context the three most important variables are goal commitment, task complexity, and ability (Fig. 1).

Before going into details of the moderator effects, some characteristics of goals influencing their effect on performance should be noted. Hundreds of studies have shown that setting a specific difficult goal leads to higher performance than does a general intention to do one's best. This is true for individuals who have the ability necessary to perform a task. For people who do not have the necessary knowledge to perform a task effectively, setting a high outcome goal can have detrimental effects on performance.

In 2001, Seijts and Latham showed that in the case where people do not have the necessary knowledge to

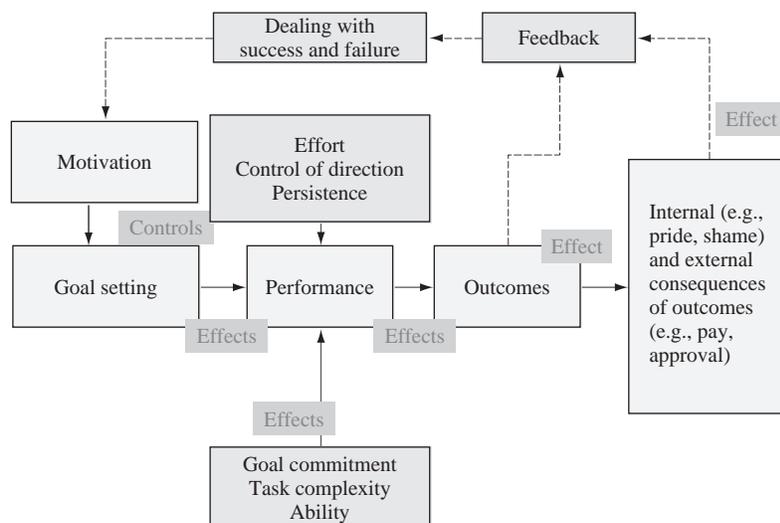


FIGURE 1 Concepts for understanding the translation of goals into action.

perform a task effectively, it is helpful to set a learning goal instead of an outcome goal. They defined a high learning goal as specific and difficult in terms of the number of strategies to be discovered to learn how to perform a task. In their study, they showed that a specific difficult learning goal leads to higher performance than does urging people to do their best on a task where they initially lack the requisite knowledge to perform it. Specific difficult outcome goals should be avoided because they appear to have a detrimental effect on performance during the learning process. The practical implication of these findings is that practitioners should focus on ways in which to increase workers' self-efficacy on tasks that the workers perceive as difficult and complex. This helps to avoid failure experiences with detrimental consequences for further goal setting and performance, and it also reduces negative effects on self-esteem.

Two additional variables in this model deserve some attention: goal commitment and the way in which actors deal with failure. Goal commitment is important to keep actions going so as to reach a goal. Highly committed persons stay with goals persistently, even when they encounter difficulties and hindrances along the way. They keep focused on the current task and resist temptations raised by other motivating potentials or motives.

Commitment to a goal actually controlling ongoing behavior directly depends on the motivation while pursuing the goal (Fig. 2). If there is no competitive motivational state, a persistent striving toward the original goal is not threatened at all. Although the strength of goal commitment for the ongoing action is affected by personal motives and situational

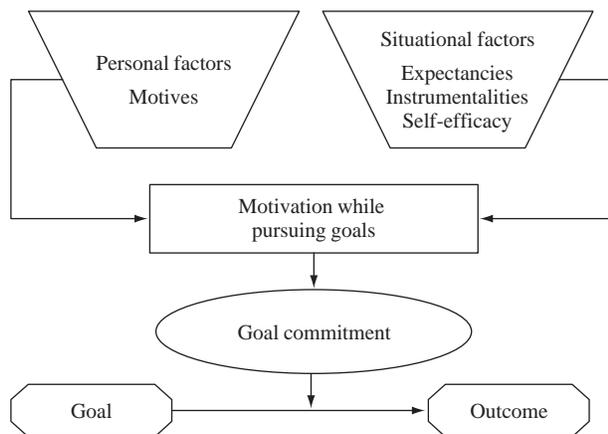


FIGURE 2 Determinants of goal commitment.

characteristics when starting the action, there may be changes in both groups of variables that are responsible for either a decrease in goal commitment or an increase in motivation for a different action. This could lead to an early change in activity (due to changing goals) before the originally intended goal has been reached. Other determinants of a high goal commitment are high self-efficacy, a high expectancy to reach the goal, and a high instrumentality of the outcome aimed for higher order outcomes in the future. A success-oriented achievement motive could also be helpful for keeping goal commitment high if the motivating potential of work is high with respect to the achievement motive system.

Setting high goals at work increases the probability of high performance. But high goals also carry the risk of failure. Although occasional failure should not be dangerous for successful work, it can be dangerous in the long run, especially when it has a negative influence on the self-esteem of acting persons. In the positive case, failure strengthens approach tendencies and goal commitment. The original goal stays as high as it was before the failure, or it is set even higher to invest more effort than before. This increases the chances of reaching the goal (Fig. 3). In the negative case, failure urges people to withdraw from the task, goal commitment goes down, and the acting persons give up on original goal. This leads to lower performance.

A large amount of research clearly shows the convincing effects of goals on behavior at work. Not only can performance be increased by high and specific

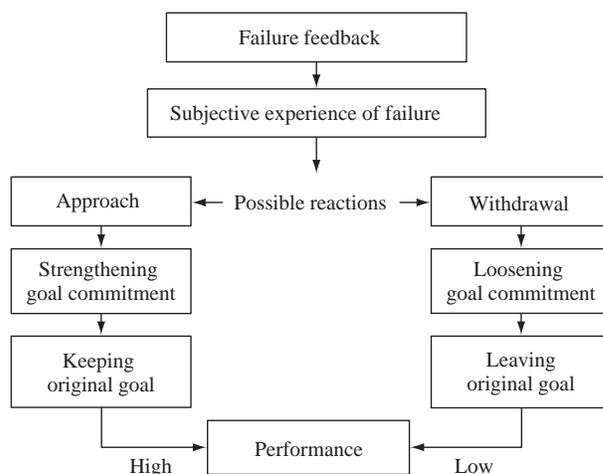


FIGURE 3 Possible reactions to failure: Approach or withdrawal.

outcome or learning goals, but withdrawal behaviors, such as absences and fluctuations, also can be influenced by goals. Important moderators are goal complexity, goal commitment, and the competence to deal with failure in an adequate mode. Motives are especially effective as central determinants of goal commitment. Of course, they unfold their influence on goal commitment in interactions with motivating potentials, changing with various situational aspects occurring in the course of an action.

### 3. DEVELOPING A PERSONAL GOAL

Whether a person sets high or low goals, whether the person directs his or her behavior in this or that direction, and whether the person persists more or less with one goal depend on the interaction of personal and situational factors. In motivational theory, these factors are called motives as personal factors and motivating potentials on the side of the situational factors. As stated previously, motives are highly aggregated classes of goals, and motivating potentials provide the occasions for realizing motive goals. Important motives in modern industrial and service-oriented societies include the following:

- Curiosity
- Achievement
- Affiliation
- Power
- Aggression
- Hunger, thirst, and sexual motives

For each motive, there are motivating potentials in the situation in which persons are acting. Motivating potentials for persons with a high curiosity motive are high when a situation enables them to increase their knowledge. For more information about motivating potentials related to various motives, see [Table I](#).

Motives and motivating potentials interact in determining goal direction and goal difficulty ([Fig. 4](#)). Empirical evidence has been reported for several motive domains. For purposes of illustration, the interaction of the achievement motive with the corresponding motivating potential of the work situation is described. The achievement motive is defined as striving for competing with a standard of excellence. Persons with a high achievement motivation like to know how efficient they are and how to improve their efficacy. To set high goals, these persons need challenging tasks that give them the opportunity to see for themselves how efficient they are and that allow them to

**TABLE I**  
**Important Motives and the Corresponding Motivating Potentials of Situations**

Motive	<i>Motivating potentials offer an opportunity to:</i>
Curiosity	Get better knowledge of what is going on
Achievement	Enhance one's competence and feel pride about having reached a goal
Affiliation	Build a new social relation and keep it in a good state
Power	Influence the behavior of others
Aggression	Perform harmful acts toward others
Hunger, thirst, and sexual motives	Get food and water and get a sexual partner

find ways in which to increase their efficacy. Such tasks are characterized by a variety of task components. They are important, offer a certain amount of autonomy, and allow continuous feedback during task performance together with clear outcome feedback. In 1976, Hackman and Oldham referred to the combination of these factors as "motivating potential of tasks."

Outcome and performance feedback can be used by the acting person to calculate expectancies about the chances of reaching a goal. In the case of achievement motivation, expectancies are also important for the strength of motivating potentials because it makes sense for those with a high achievement motive to work with tasks that give them at least a moderate probability of success. In this way, it is possible for persons with a high achievement motive to learn as much as possible about their own efficacy.

### 4. ACCEPTANCE OF GOALS

Very often in their lives, especially at work, persons are asked to accept goals that have not been set by themselves. In such cases, the question arises as to whether they accept these external goals for controlling and regulating their own behavior. Reacting on externally set goals, a person has at least three choices: accepting the goals, internalizing the goals, or using the goals to control goal-oriented actions ([Fig. 5](#)). This happens, of course, when the externally set goals are not in conflict

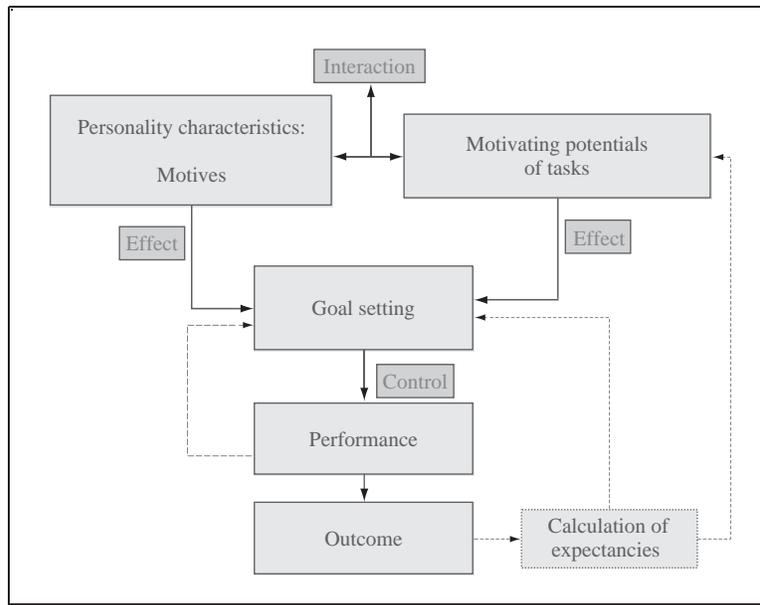


FIGURE 4 The formation of goals as the result of an interaction between personal and situational characteristics.

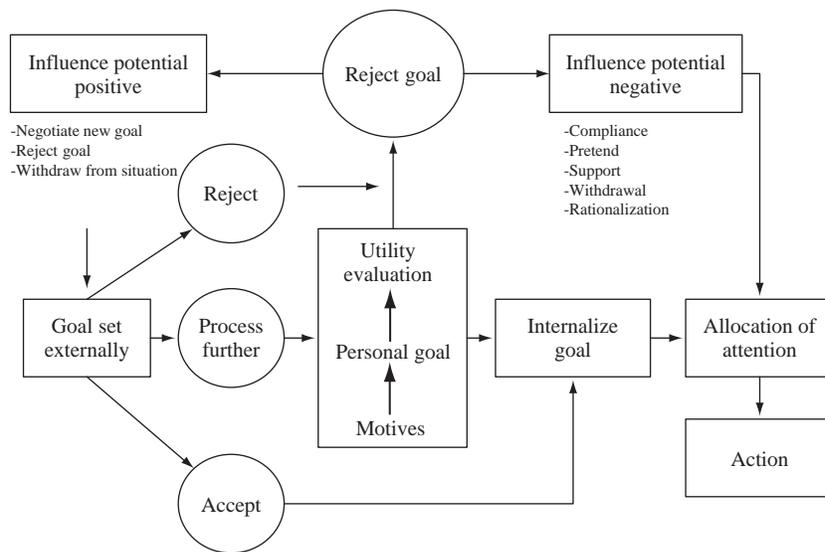


FIGURE 5 Concepts used to describe the process of accepting and internalizing or rejecting externally set goals. Adapted from Earley and Shalley (1991).

with personal goals, when they might even lead to outcomes with high personal valence, or when the costs of a conflict between personal and organizational goals is compensated by a stimulating reward.

On the other hand, one can reject goals directly when it appears clear from the beginning that there will be conflicts with personal goals and no compensation is in

sight. In this case, the person begins to think about the possibility of influencing the leader who set the goal. If influence appears to be possible, negotiation is started to change the original goal, possibly leading to a new goal. If the influence potential of the situation is negative, the person must decide either to comply and accept the goal or to withdraw from the situation.

These considerations have some practical implications for motivating people at work. When setting goals externally, it makes sense to be aware of the motives of workers and employees because these motives are of essential importance to their personal goals. If managers know these goals, they can calculate in advance what effects their externally set goals will have. If conflict between externally set goals and personal goals is expected, managers either can reformulate their goals to minimize conflict or can use other techniques that make clear how important it is to work on the goals set by them even if such goals are not in line with personal goals. One technique could be to introduce a participative goal-setting program, as discussed later in the article.

### 5. HOW TO PURSUE GOALS THAT ARE (STILL) UNKNOWN

Activities at work in general are goal oriented. Often, goals are specific and difficult, leading to high performance. Goal-setting theory offers evidence for understanding these effects. But nowadays, more and more work activities are no longer related to clear and specific goals because the specific goals are not yet known when the activities are started. One classic example is given with design tasks. They are characterized by the fact that outcomes are not at all clear for a long time during work on the tasks. They are similar to the learning tasks mentioned earlier in the article. In design tasks, people cannot do the following:

- Analyze the conditions of action necessary for reaching the goals exactly.
- Run trial actions.
- Form language terms to describe actions.
- Coordinate component goals toward ultimate goals.
- Compare outcomes with goals stored in memory.

Although these activities cannot be regulated by specific goals as anticipated outcomes, design actions are not without direction given that the outcomes must be developed first. During activities of this kind, people must do the following:

- Follow up more than one problem-solving alternative.
- Build up evaluation criteria to be used in iterative processes of error correction.
- Look for similar cases to be used as models for the current ones.

The practical consequences of these considerations are easy to describe. Although it is known that high specific and concrete outcome goals lead to high performance, it is not always possible to set such goals, especially when the outcome of work is not quite clear at the beginning. In such a case the chances for success will increase through setting learning goals such as the following:

- Discover a number of strategies to learn how to perform the task.
- Develop systematic testing of task-relevant strategies.
- Identify evaluation criteria.
- Use information from similar cases.

After having learned about strategies to solve the task, it could be helpful to again relate goals to outcomes.

### 6. MANAGING SUCCESSFUL GOAL-ORIENTED WORK BEHAVIOR: THE PARTICIPATORY PRODUCTIVITY MANAGEMENT

During the second half of the 20th century, many management systems that focused on goal setting were presented and used. The management by objectives system has been used successfully for five decades. Goal-oriented management systems help leaders to coordinate the goal-setting process continuously. The Balanced Scorecard (BSC) is well known for its ability to support successful management in this respect. Another powerful instrument with a more solid empirical foundation is the Participative Productivity Management (PPM), originally developed by Pritchard in 1990 as the Productivity Measurement and Enhancement System (ProMES). PPM is different from BSC in its very elaborate bottom-up approach that makes sure that workers accept the measurement and feedback system used to prepare adequate goal setting. So, it effectively helps to avoid goal conflicts between external or task-driven goals and personal goals set by the workers themselves.

In the framework of this management system, work group members develop their own measurement system coordinated with all responsible actors in the company, enabling them to participate as well-informed partners in the goal-setting process. In weekly or monthly systematic feedback sessions, group members and their leaders come together to discuss feedback delivered by the measurement

system that they had developed themselves. On the level of various performance indicators (e.g., amount of items produced, quality of work, customer satisfaction, scheduling effectiveness), group members use this system to describe the discrepancies between their previously set goals and their actual outcomes for each indicator. The system also helps the workers to weigh each indicator score with respect to its importance for the organization's productivity. On this basis, they discuss the reasons why they succeeded or failed to reach their goals. As a result of this discussion, they set new goals for the next measurement period. These procedures keep them in a continuous improvement process. The task of their leaders becomes more and more one of facilitators of a self-controlled productivity process. Meta-analyses show that using this management system has an enormous effect on productivity.

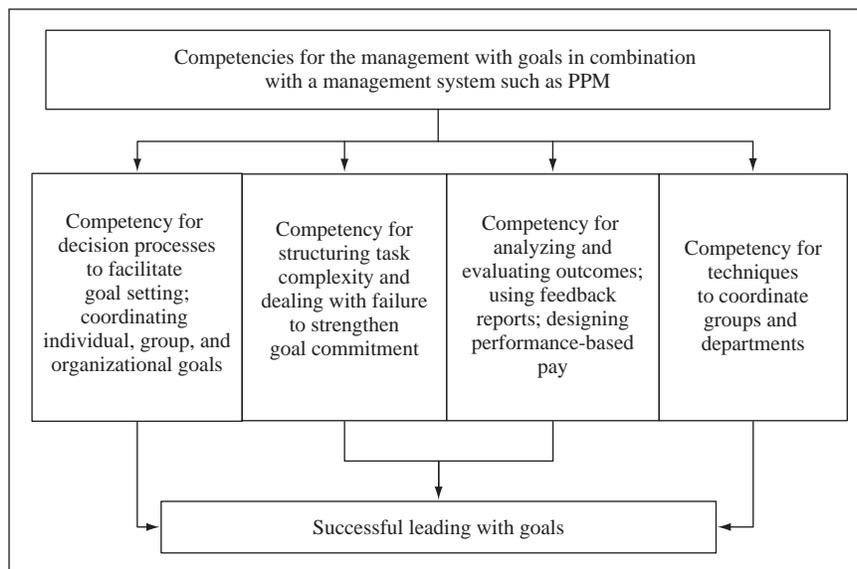
### 7. COMPETENCIES IN MANAGING GOAL-DIRECTED BEHAVIOR

Empirical data show that goal setting can be used as a technique to manage people effectively toward the goals of an organization. As was shown in the preceding section, the goal-setting process will profit from the support of a goal-oriented management system. For example,

PPM offers the structural systemic background for successful goal setting. It helps leaders to do the following:

- Set goals participatively so that acceptance of organizational goals can be expected.
- Delegate responsibilities to work group members.
- Clarify goal–outcome relations.
- Define clear criteria for success and failure.

But PPM does not replace leaders. The tasks of leaders will change, but leadership will still be necessary. Leaders using a management technique such as PPM should have some special competencies to support workers' goal setting and the effect of goals during various phases of the action process. During the first phase of the action process, when the goals are developed, special competencies are required to identify the personal goals (and motives) of workers and to take them into account while discussing the necessary goals for the next work and feedback period defined in the PPM measurement system (Fig. 6). Another competency is required to structure complex tasks and identify subtasks with corresponding subgoals. In the case of evaluating negative outcomes, leaders need some competencies as well. Dealing with failure can be dangerous for leaders because it is necessary to feed back failure without trying to make things seem better as they are. The danger is that feeding back such information about failure can be seen as an indicator of low ability by failing persons. This might be a reason for lowering



**FIGURE 6** Competencies needed or leaders using goal setting in combination with a goal-oriented management system.

one's self-esteem, with further effects on goal setting. Goals will be set lower, or high external goals will not be accepted, due to fears of failing again. Although work motivation can be decreased by failure feedback, it does not make sense to hold back such information. Instead, leaders should be able to talk about failure as an opportunity to learn about new strategies to avoid failure and, as such, to increase their self-efficacy. These competencies help to strengthen goal commitment. Another set of competencies for successful leadership involves coordinating persons in groups as well as between groups and departments.

### 8. AN INTEGRATIVE MODEL FOR THE EFFECTS OF MOTIVES AND GOALS

Motives and goals work together to control the intensity, direction, and persistence of behavior at work. Figure 7 summarizes the effects of motives and goals during all phases of the action process. Motives are defined as classes of goals that determine, in interaction with motivating potentials of the work situation, the kinds of goals people use to control their actions. The way in which they do this is described by expectancy-value theories. One example of an empirically sound specific

theory of this kind is achievement motivation. It describes a motive (achievement motive) and a motivating potential (task difficulty) in explaining the development of individual goals in achievement situations. The translation of a goal into action and performance is the topic of goal-setting theory. Goal characteristics and mechanisms and moderators of the translating process are the determinants of outcomes. During this action phase, motives play a role as well. They affect the degree of goal commitment persons show while performing. Even the evaluative phase of an outcome motive can be seen at work. Motives influence the way in which causal attributions are used to explain the outcome. Especially interesting is the impact of motives on the way in which people deal with failure.

Failure delivers two kinds of information. First, it informs about the discrepancy between an outcome and the goal. Second, it says something about a person's efficiency. In case of failure, this second kind of information is dangerous for one's self-esteem. It tells a person that he or she is not able to succeed at this kind of task, especially if the failure repeats. Persons with a high achievement motive are better equipped to deal with failure because they tend to increase motivation, stay with their goals, and finally reach them. The integrated model helps to understand how goals are established, how they affect ongoing action, and how they influence the evaluation of outcomes of actions. It

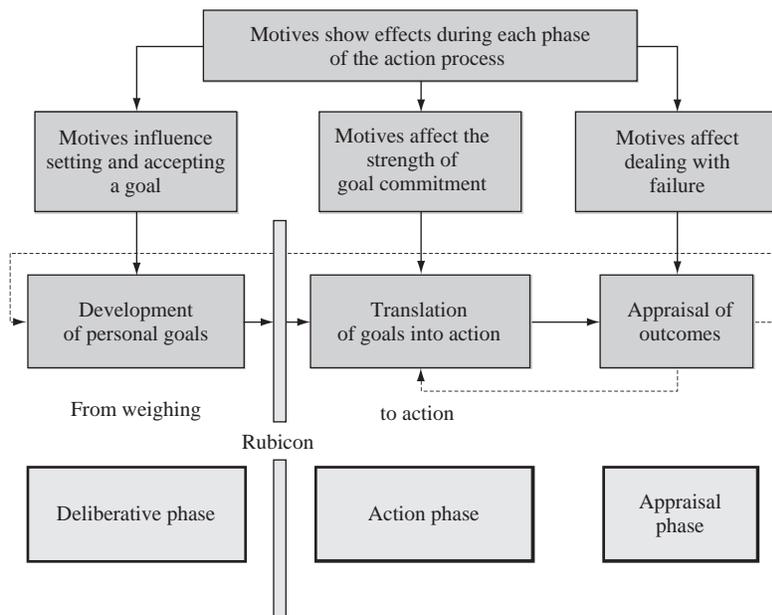


FIGURE 7 The role of motives in each phase of an action process.

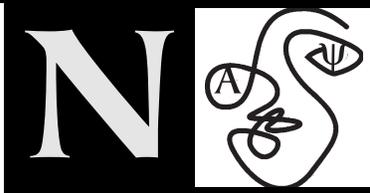
summarizes a fundamental process that can be differentiated during all of its phases to describe the many facets of goal-directed behavior. Most of the concepts in the model have been tested empirically and have proven to be valid. On such an empirical ground, the model allows one to develop psychologically sound techniques to be used in applied psychology for increasing productivity and for adapting tasks and organizational contexts to the structure and functions of human resources at work.

### See Also the Following Articles

Compensation ■ Cooperation at Work ■ Goal Setting and Achievement Motivation in Sport ■ Intrinsic Motivation and Self-Determination ■ Motivational Taxonomies ■ Work Motivation ■ Work Role, Values Sought in the

### Further Reading

- Atkinson, J. W. (1974). Strength of motivation and efficiency of performance. In J. W. Atkinson, & J. O. Raynor (Eds.), *Motivation and achievement* (pp. 193–218). Washington, DC: Winston.
- Austin, J. T., & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin*, *120*, 338–375.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge, UK: Cambridge University Press.
- Earley, P. C., Connolly, T., & Ekegren, G. (1989). Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. *Journal of Applied Psychology*, *74*, 24–33.
- Earley, P. C., & Shalley, C. E. (1991). New perspectives on goals and performance: Merging motivation and cognition. In G. R. Ferris, & K. M. Rowland (Eds.), *Research in personnel and human resources management* (Vol. 9, pp. 121–157). Greenwich, CT: JAI.
- Hacker, W. (1998). *Allgemeine Arbeitspsychologie: Psychische Regulation von Arbeitstätigkeiten*. Bern, Germany: Hans Huber.
- Hacker, W. (2004). Wie verfolgt man Ziele, die man noch nicht kennt? Ergebnisorientierte opportunistische Tätigkeitsregulation. In J. Wegge, & K-H. Schmidt (Eds.), *Förderung von Arbeitsmotivation und Gesundheit in Organisationen* (pp. 27–37). Göttingen, Germany: Hogrefe.
- Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, *16*, 250–279.
- Heckhausen, H. (1989). *Motivation und Handeln*. Berlin, Germany: Springer.
- Ilgen, D. R., & Davies, C. A. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology*, *49*, 550–565.
- Kanfer, R., & Ackerman, P. (1989). Motivation and cognitive abilities: An integrative aptitude–treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657–690.
- Kleinbeck, U., Wegge, J., & Schmidt, K-H. (2001). Work motivation and performance in groups. In M. Erez, U. Kleinbeck, & H. Thierry (Eds.), *Work motivation in the context of a globalizing economy* (pp. 181–191). Mahwah, NJ: Lawrence Erlbaum.
- Kuhl, J. (2001). *Motivation und Persönlichkeit*. Göttingen, Germany: Hogrefe.
- Latham, G. P., & Frayne, C. A. (1989). Self-management training for increased job attendance: A follow-up and a replication. *Journal of Applied Psychology*, *74*, 411–416.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*, 705–717.
- McClelland, D. C. (1987). *Human motivation*. New York: Cambridge University Press.
- Pinder, C. C. (1998). *Work motivation in organizational behavior*. Upper Saddle River, NJ: Prentice Hall.
- Pritchard, R. D. (1990). *Measuring and improving organizational productivity: A practical guide*. New York: Praeger.
- Pritchard, R. D., Holling, H., Lammers, F., & Clark, B. D. (Eds.). (2002). *Improving organizational performance with the productivity measurement and enhancement system: An international collaboration*. New York: Nova.
- Rodgers, R., & Hunter, J. E. (1991). Impact of management by objectives on organizational productivity. *Journal of Applied Psychology*, *76*, 322–336.
- Schneider, K., & Schmalz, H-D. (2000). *Motivation*. Stuttgart, Germany: Kohlhammer.
- Seijts, G. H., & Latham, G. P. (2001). The effect of distal learning, outcome, and proximal goals on a moderately complex task. *Journal of Organizational Behavior*, *22*, 291–307.
- Vroom, V. H. (1964). *Work and motivation*. New York: John Wiley.
- Weiner, B., Heckhausen, H., Meyer, W. U., & Cook, R. E. (1972). Causal ascriptions and achievement motivation: A conceptual analysis of effort and reanalysis of locus of control. *Journal of Personality and Social Psychology*, *21*, 239–248.
- Winters, D., & Latham, G. P. (1996). The effect of learning versus outcome goals on a simple versus a complex task. *Group and Organization Management*, *21*, 235–250.



# Neuropsychological Assessment in Schools

Daniel C. Miller

Texas Woman's University, Denton, Texas, USA

1. School-Based Problems That Can Be Addressed by Neuropsychological Assessment
2. State of the Art of Neuropsychological Assessment in Schools
3. Conceptual Framework for School Neuropsychological Assessment
4. Summary  
Further Reading

## GLOSSARY

**pediatric neuropsychologist** Uses standardized tests and observes behavior to define a child's pattern of cognitive development. The pediatric neuropsychologist uses knowledge of brain development, brain organization, and the effects of various forms of brain injury on development to guide this assessment and interpret results.

**school neuropsychologist** Integrates neuropsychological and educational principles in assessments and interventions with infants, children, and adolescents to facilitate learning and behavior within the school and family environments. School neuropsychologists also play an important role in curriculum development, classroom design, and the integration of differentiated instruction that is based on brain-behavior principles in order to provide an optimal learning environment for every child.

**specific learning disabilities** Disorder in one or more of the basic psychological processes involved in understanding

or using spoken or written language that may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations.

**traumatic brain injury** An acquired injury to the brain caused by an external physical force—resulting in total or partial functional disability, psychosocial impairment, or both—that adversely affects a child's educational performance. The term applies to open or closed head injuries resulting in impairments in one or more areas, such as cognition; language; memory; attention; reasoning; abstract thinking; judgment; problem solving; sensory, perceptual, and motor abilities; psychosocial behavior; physical functions; information processing; and speech. The term does not apply to brain injuries that are congenital or degenerative or to brain injuries induced by birth trauma.

Increasingly, school-aged children and adolescents have cognitive processing deficits (e.g., memory, learning, and attention) or behavioral/emotional deficits (e.g., autism, mood disorders, anxiety, and conduct problems) that are a result of known or suspected neurological conditions. Educational personnel, including school psychologists, are not always prepared to assess and intervene with these neurologically impaired populations; however, they can be trained. Traditionally, psychoeducational assessments measure estimates of overall intelligence and academic achievement but do not assess the underlying cognitive processes and sensory-motor

functions associated with learning and behavior. Neuropsychological assessments are designed to assess discrete cognitive processes (e.g., visual short-term memory, auditory short-term memory, and sustained attention) that help facilitate or inhibit the learning process. Ideally, a neuropsychological assessment of a school-aged child should link the assessment results with instructional and behavioral recommendations or interventions that improve the child's academic and/or social-emotional/behavioral functioning. This article presents several examples of children who would benefit from a neuropsychological evaluation, discusses several reasons why there is growing interest in integrating neuropsychological principles into psychoeducational assessment and intervention, discusses the current state of the art of school neuropsychological assessment, and presents a conceptual framework for school neuropsychological evaluations.

## **1. SCHOOL-BASED PROBLEMS THAT CAN BE ADDRESSED BY NEUROPSYCHOLOGICAL ASSESSMENT**

### **1.1. Case Examples**

Increasingly, children and adolescents are coming to school with known or suspected neurological conditions. Two examples are presented here that illustrate the types of cases being identified by educators and assessment specialists within the schools.

First, Mary is 17 years old and a junior in high school. While on a date with her boyfriend, a drunk driver struck the car in which they were riding head on. Her boyfriend escaped the accident with minor injuries, but Mary was not wearing her seat belt. She was thrown through the windshield. Mary suffered a severe head injury and was hospitalized. Several weeks later, after her medical needs had become stabilized, she was released from the hospital. Her medical insurance did not cover any cognitive rehabilitation so she was sent home. When Mary returned to school a month after the accident, her physical wounds had healed and she looked as if she had healed completely. She returned to her 11th-grade regular education classes. Prior to the accident, Mary was an "A" and "B" student. When Mary returned to school after the accident, her grades declined dramatically, she complained about having trouble focusing attention and concentrating for long periods

of time, she could not remember things, and she seemed to tire easily.

Second, Sam is in the fifth grade. He was diagnosed 3 years earlier with a specific learning disability in the areas of reading comprehension and written expression. Sam has average cognitive capabilities, but he has severe problems with understanding what he has read and putting his thoughts down on paper. Sam has made very little academic progress during the past 3 years. His teachers express frustration that they cannot seem to "find the right buttons to push" to help Sam make academic progress.

Sam and Mary represent two examples of children who have disabilities defined in federal legislation that are of particular interest to neuropsychologists: specific learning disability (SLD) and traumatic brain injury (TBI). In 1975, the U.S. Congress passed landmark legislation with the Education of All Handicapped Children Act (P.L. 94-142), designed to provide free and appropriate educational services to children. P.L. 94-142 was reauthorized in 1990 and became known as the Individuals with Disabilities Act (IDEA). IDEA was reauthorized in 1997 and is under revision with an expected adoption date in 2004.

### **1.2. Specific Learning Disabilities**

SLD was included in the 1975 P.L. 94-142 as a student disability. SLD is a generic term that refers to a heterogeneous group of disorders due to identifiable or inferred central nervous system dysfunction. Such disorders may be manifested by delays in development and/or difficulties in one or more of the following areas: reading accuracy, reading comprehension, mathematical calculations, mathematical reasoning, written expression, listening comprehension, and oral expression. Since the inception of the concept of a SLD, educators have struggled with how to identify children who have these processing disorders caused by some type of central nervous system dysfunction. Commonly used methods of SLD identification, such as aptitude/achievement discrepancies, have failed to address the underlying brain processing disorder(s).

The revision of IDEA under consideration encourages states to deemphasize the use of discrepancy models in the identification of SLD and move toward a three-tiered intervention model. Tier 1 operates on the principles that high-quality instructional and behavioral supports are required for all students in general education. Tier 2 would require targeted, intensive prevention or remediation services for students whose performance and

rate of progress lag behind the norm for their grade and educational setting. Tier 3 would require a comprehensive evaluation by a multidisciplinary team to determine eligibility for special education and related services. Within the proposed third tier of the intervention model, neuropsychological assessments could provide valuable information about the child's cognitive strengths and weaknesses and the links to prescriptive interventions.

### 1.3. Traumatic Brain Injury

In the 1990 revision of IDEA, TBI was included for the first time as an identifiable student disability. TBI is an injury to the brain caused by the head being hit or shaken violently. When a head injury occurs in a child or adolescent, the brain has not fully developed. As a result, recovery from loss of function due to a head injury is often uneven and unpredictable. School-aged children who have sustained a head injury require organized, cooperative, and closely monitored special education services. As recently as 10 years ago, a child with a head injury would typically move from an acute care hospital setting, in which his or her physical needs were addressed, to an intermediate rehabilitation setting for a few months, in which cognitive rehabilitation took place. Today, it is typical for a child to be released from a hospital as soon as his or her medical condition is stabilized and to return to his or her regular classroom soon thereafter, without any cognitive rehabilitation. During the past 15 years, managed care within the insurance industry has led to a reduction of cognitive rehabilitation services offered to TBI victims. Educators have been faced with how to educate children with increasingly complicated learning and emotional needs.

Neuropsychological assessment can play an important role with TBI students by determining the mental and behavioral effects of the brain injury, identifying the functional assets and deficits for instructional planning, and establishing baseline data to evaluate subsequent change. Because of the rapid changes in recovery of function, it is often necessary to meet frequently as an individualized education plan team to review the TBI student's progress and to adjust the educational goals. Educational personnel, including school psychologists, speech and language pathologists, occupational therapists, physical therapists, adaptive physical education specialists, and regular and special education teachers, must all work collaboratively with the family to meet the educational needs of the student.

### 1.4. Other Disabilities That May Warrant a Neuropsychological Evaluation

Table I presents a list of other suspected or known disabilities or neurological impairments that would benefit from neuropsychological assessments. In the 1990 and subsequent IDEA definitions of TBI, only injuries to the brain as a result of trauma are classified as TBI. Acquired brain injuries as a result of infection, disease, chemotherapy, or high fever are not considered under the current definition as TBI. Children with acquired brain injuries are typically identified as "other health impaired" if there are significant academic and/or behavioral concerns that need to be addressed.

Research has shown that children born with neurodevelopmental risk factors, such as low birth weight, prenatal exposure to drugs or alcohol or both, or exposure to environmental toxins (e.g., lead), are likely to present with neurocognitive deficits (e.g., memory, attention, processing speed, and language skills). Children with medical conditions such as epilepsy or children undergoing chemotherapy due to cancer may also be at risk for neurocognitive deficits associated with their illnesses or treatments.

**TABLE I**  
Suspected Handicaps and Neurological Conditions  
for Which Neuropsychological Assessment  
May Be Beneficial

Children and adolescents who are having academic or behavioral difficulties due to
Severe learning disabilities that have not responded to traditional intervention techniques
Processing discrepancies on psychoeducational measures (e.g., memory deficits, language processing deficits, attentional processing deficits, and processing speed deficits)
A documented rapid decline in academic achievement that cannot be explained by social, environmental, or psychological causes
Congenital or acquired brain damage
Neuromuscular diseases such as cerebral palsy
Chronic epilepsy
Neurodevelopmental risk factors such as low birth weight, prenatal exposure to drugs or alcohol or both, or exposure to environmental toxins (e.g., lead)
A treated or suspected brain tumor
An infection of the central nervous system
An unusually large scatter among test scores that cannot be explained

Unfortunately, many children who have known or suspected neurological conditions are misclassified [often as learning disabled (LD) or emotionally disturbed (ED)], which leads to planning that does not fully address their educational needs. For example, an undiagnosed TBI student can develop secondary emotional and behavioral problems due to the head injury and the associated cognitive deficits. Labeling this TBI student as emotionally disturbed and focusing the majority of the educational efforts on regulating the student's behavior will not address the underlying cognitive processing problems.

Traditional psychoeducational assessments conducted with school-aged children do not fully address the types of cognitive processing deficits associated with children having known or suspected neurological conditions. Neuropsychological assessments measure discrete cognitive processes (e.g., visual short-term memory) that help facilitate or inhibit the learning process. The results of a neuropsychological assessment can provide a profile of a child's functional strengths and weaknesses that will facilitate instructional planning.

## 2. STATE OF THE ART OF NEUROPSYCHOLOGICAL ASSESSMENT IN SCHOOLS

### 2.1. Neuropsychological Assessment in the Past and Present

From the 1940s to the 1970s, one of the main functions of a neuropsychological evaluation was to identify the source of a possible lesion or area of dysfunction within the brain. With the advent of modern neuroimaging (e.g., computerized axial tomography, magnetic resonance imaging, functional magnetic resonance imaging, and positron emission tomography), these specialized techniques are now used to directly assess the localization of lesions or tumors to millimeter precision. The goal of neuropsychological assessment should be to identify the functional strengths and weaknesses of the examinee and to translate those data into evidenced-based intervention strategies rather than to try to localize possible lesion sites within the brain. A lengthy neuropsychological report to the schools with the conclusion that "Johnny has a right parietal lobe lesion" will not facilitate educational planning. The neuropsychological report must summarize implementable prescriptive recommendations that address the student's functional strengths and weaknesses for parents and teachers.

Evaluators who wanted to administer neuropsychological measures to children prior to the 1990s were very limited in their assessment choices. The fixed batteries that were standardized 20–40 years ago (e.g., Halstead–Reitan Neuropsychological Batteries and Luria–Nebraska Neuropsychological Battery) were the principle measures available for neuropsychologists. Unfortunately, these test batteries were not designed for children but were downward extensions of adult measures. Also, the fixed batteries of the 1950s and 1960s had limited standardization samples for children and would not stand up to the scrutiny of test development today. In the 1990s, test authors and publishers provided school-based assessment specialists with a wide variety of new tests of memory and learning designed for and standardized on large numbers of children and adolescents (Table II).

Since the late 1990s, there has been an increased number of fixed battery tests developed specifically for school-aged children (Table III). Some of the tests, such as the NEPSY, are neuropsychological batteries based on modern theory and research and designed specifically for young children. Other fixed batteries are intelligence tests that have been reformulated to take into account advances in cognitive psychology, neuropsychology, and intelligence test theories [e.g., Woodcock–Johnson Revised Tests of Cognitive Abilities (WJIII) and the Wechsler Intelligence Scale for Children—Fourth Edition]. Tests such as the Delis–Kaplan Executive Functions System have revised and restandardized traditional tests (e.g., Trail Making and Stroop) and added to their interpretative depth.

**TABLE II**  
**Tests of Memory and Learning for School-Aged Children**

<i>Test</i>	<i>Age range (years)</i>
Wide Range Assessment of Memory and Learning (WRAML; Sheslow & Adams, 1990; WRAML-2; Adams and Sheslow, 2003)	WRAML: 5–17 WRAML2: 5–90
Test of Memory and Learning (TOMAL; Bigler and Reynolds, 1994)	5–19
Children's Memory Scale (CMS; Cohen, 1997)	5–16
Wechsler Memory Scale—Third Edition (WMS-III; Wechsler, 1997)	16–89
California Verbal List Learning—Second Edition (CVLT-II; Delis <i>et al.</i> , 2000)	5–16

**TABLE III**  
**Selected Fixed Batteries for School-Aged Children and Adolescents**

Test	Age range (years)
Cognitive Assessment System (CAS; Naglieri & Das, 1997)	5–17
NEPSY (Korkman <i>et al.</i> , 1997)	3–12
WISC-III as a Process Instrument (WISC-III PI; Kaplan <i>et al.</i> , 1999)	6–16
Delis–Kaplan Executive Functions System (D-KEFS; Delis <i>et al.</i> , 2001)	8–89
Woodcock–Johnson III Tests of Cognitive Ability and Achievement (WJIII; Woodcock <i>et al.</i> , 2000)	2–90+
Dean–Woodcock Neuropsychological Battery (DW; Dean and Woodcock, 2002)	4–90+
Diagnostic Supplement to the WJIII Tests of Cognitive Abilities (DS-WJIII; Woodcock <i>et al.</i> , 2003)	2–90+
Wechsler Pre-School and Primary Scale of Intelligence—Third Edition (WPPSI-III; Wechsler, 2002)	2–7
Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003)	6–16
Kaufmann Assessment Battery for Children—Second Edition (KABC-II; Kaufman and Kaufman, 2004)	3–18

There are three major trends in neuropsychological assessment. The first is a subtle shift toward providing the evaluator qualitative information about the cognitive processes a child used to obtain a score, in addition to the score itself. For example, publishers are providing base rates for each age range within the standardization sample for common qualitative behaviors (e.g., asking for repetitions and failure to comprehend). This “process approach” is important because the qualitative behaviors often provide valuable insight into how a child’s brain is processing information and assist in the diagnoses of neuropsychological conditions.

The second trend in neuropsychological assessment is ensuring the assessment is grounded in theory. Test authors and publishers are more consistently providing a theoretical foundation for their tests, which aids in test interpretation (e.g., NEPSY based on Lurian theory or the WJIII based on the Carroll–Horn–Cattell model). Finally, the third trend will be an added

emphasis during the next decade on linking assessment with evidenced-based interventions. Researchers and clinicians will need to establish and verify linkages between assessment and intervention.

## 2.2. Who Can Complete a Neuropsychological Assessment?

The most highly trained assessment specialists in schools are school psychologists. However, the majority of school psychologists are trained at the specialist level (master’s degree +30 hours) of training, which provides minimal training in the biological bases of behavior and rarely covers neuropsychological assessment techniques. School psychologists can only conduct neuropsychological assessments if they have expertise in neuropsychology. To ensure proper expertise, school psychologists must complete an organized competency-based, post-graduate training or a doctoral program in which they can specialize in school neuropsychology.

If a school district does not have a school neuropsychologist on staff, which is the norm, the district will often request a neuropsychological evaluation from a clinical neuropsychologist in private practice or working in a hospital setting. Clinical neuropsychologists are licensed psychologists who are trained at the doctoral level, typically with a degree in clinical psychology and a specialization in clinical neuropsychology. However, the majority of clinical neuropsychologists have expertise in completing adult assessments rather than school-based neuropsychological evaluations.

A pediatric neuropsychologist is a clinical neuropsychologist specifically trained to work with children. If school districts seek an outside neuropsychological assessment, it is best to contract a pediatric neuropsychologist. The downside to contracting with any specialist outside of schools is that the specialist may not be familiar with the IDEA rules and regulations that guide special education placements and may operate under a clinical model using clinical nomenclature (e.g., *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, diagnoses).

For neuropsychological principles to be best integrated into schools, there needs to be more cross-training. Many of the new cognitive ability tests and tests of memory and learning being administered by school psychologists have strong foundations in neuropsychological theory. School psychologists and educators need to improve their knowledge base about brain–behavior relationships if they are going to

appropriately interpret these newer tests. There are more than 25,000 school psychologists who have direct access to children in schools. Training to use and interpret neuropsychological assessments is important. At a minimum, school psychologists and educators should be taught to recognize the warning signs of neuropsychological conditions and know when to refer to another specialist for assistance.

As a general rule, pediatric neuropsychologists consulting with schools should learn to integrate their neuropsychological assessments within the context of IDEA eligibility and recommend realistic school-based interventions. A school neuropsychology specialist can write an insightful report about a child's cognitive strengths and weaknesses and make a set of educational recommendations that are evidence based and practical. It is possible that the neuropsychological assessment report may simply be filed away in a child's academic folder if the classroom teachers do not understand how to implement the educational recommendations. An essential role that a school neuropsychology specialist can play is to consult with classroom teachers regarding the integration of the neuropsychological assessment results into the instructional planning.

### 3. CONCEPTUAL FRAMEWORK FOR SCHOOL NEUROPSYCHOLOGICAL ASSESSMENT

A neuropsychological evaluation for a school-aged child should focus on answering the referral question. A conceptual framework for a school neuropsychological evaluation is outlined in Table IV. A sensory perceptual examination across the auditory, visual, and kinesthetic modalities should be a required baseline assessment in all evaluations. If the child is not able to adequately receive and process sensory perceptual information, then performance on higher order cognitive functions could be compromised. The next major component to be assessed is attentional processing skills. Attention serves an important baseline for the rest of the higher order processing. Attention is not a unitary construct. It is recommended that a school neuropsychological evaluation assess the following elements of attention for both the visual and auditory modalities: initiating, sustaining, inhibiting, shifting, and divided.

Assessment of executive functions is also a recommended component. For a thorough evaluation of executive functions, the assessment must include cognitive

TABLE IV  
Suggested Components of a School Neuropsychological Assessment<sup>a</sup>

Sensory perceptual
Auditory
Visual
Kinesthetic/tactile
Attention
Auditory (initiating, sustaining, inhibiting, shifting, and divided)
Visual (initiating, sustaining, inhibiting, shifting, and divided)
Executive functions
Cognitive (e.g., processing speed and efficiency, initiating and shifting cognitive sets, conceptual shift, planning, organizing, self-monitoring, and verbal/nonverbal retrieval fluency)
Motor programming
Behavioral/emotional regulation [e.g., interest, motivation, initiative, disinhibition (impulsivity), affect, judgment, insight, and activity level]
Memory
Auditory modality
Auditory, verbal
Explicit acquisition
Rote (word lists) (immediate and delayed)
Associative (paired learning) (immediate and delayed)
Discourse (words, sentences, paragraphs, and stories) (immediate and delayed)
Implicit acquisition
Incidental (passive)
Retrieval (recall—primacy and recency effects—and recognition)
Auditory, nonverbal
Explicit acquisition (immediate and delayed)
Implicit acquisition
Retrieval (recall—primacy and recency effects—and recognition)
Visual modality
Visual, verbal [letters, words, picture or shapes with verbal labels (rebus), etc.]
Acquisition (immediate and delayed)
Initial registration [sensory memory to short-term memory (immediate)]
Consolidation [short-term memory to long-term memory (delayed)]
Retrieval (recall and recognition)
Visual, spatial (e.g., Rey–Osterieth)
Acquisition (immediate and delayed)
Retrieval (recall and recognition)

Continues

*Continued*

---

Motor memory
Procedural [e.g., literal apraxias (imitation) and paragraphias (spelling)]
Complex motor tasks (e.g., riding a bicycle and tying shoes)
Ideomotor and ideational apraxias
Writing and typing
Language
Receptive (phonology, semantics, grammar/syntax, pragmatics, and listening comprehension)
Expressive [phonology, semantics, grammar/syntax, pragmatics, and discourse (e.g., articulation)]
Reading (phonology, semantics, grammar/syntax, pragmatics, decoding, and comprehension)
Written expression (not writing) (phonology–sequential processing, semantics, grammar/syntax, pragmatics, decoding, comprehension, and organization of thought and content)
Motor
Fine (coordination and complex mechanical processes)
Gross
Speed
Strength
Mathematics
Verbal (e.g., algebra)
Number knowledge
Counting
Computation/problem solving
Reasoning
Nonverbal (e.g., geometry)
Math notation [e.g., knowledge of math signs (+, −, ×, /)]
Computation
Reasoning in the time–space continuum
Social–emotional
Personality characteristics
Temperament
Psychopathology

---

<sup>a</sup>Adapted from a proposed model by Miller *et al.* (manuscript in preparation).

integrity, behavioral regulation, and motor programming components. Memory assessment is next in the conceptual framework. Memory is multifaceted as well. A comprehensive evaluation must include assessment of auditory and verbal memory and the associated subcomponents (immediate, delayed, recognition, recall, etc.).

Language assessment is next in the conceptual framework. Language assessment includes measures of receptive, expressive, reading, and written expression. An

evaluation of motor functions, if warranted, would include fine, gross, speed, and strength components. Mathematical skills are conceptualized as a separate area that includes verbal components (number knowledge, counting, computation, and reasoning) and nonverbal components (math notation, reasoning in time and space, and computation). Finally, a comprehensive school neuropsychological assessment should include assessment of the student's social–emotional functioning (e.g., personality characteristics, temperament, and psychopathology). This conceptual framework serves as a starting point when considering a comprehensive school neuropsychological evaluation. The component parts that are selected should vary from student to student based on the referral questions to be answered and the presenting behavioral and/or learning problems.

#### 4. SUMMARY

Children and youth who have known or suspected neurological conditions are increasingly being identified in schools. Traditional psychoeducational measures do not fully address the underlying neurocognitive deficits associated with SLD, TBI, acquired brain injury, epilepsy, etc. However, neuropsychological assessments measure discrete neurocognitive functions (e.g., visual short-term memory, processing speed, and sustained attention) that provide valuable insight into the neurocognitive strengths and weaknesses of the child and guide prescriptive interventions to improve instruction or provide positive behavioral supports.

Neuropsychological assessments for school-aged children have changed dramatically in the past 60 years. The purpose of a neuropsychological evaluation has evolved from identifying potential brain lesions to identifying functional profiles of strengths and weaknesses that link to prescriptive interventions. There has been a tremendous increase in the number and quality of neuropsychological assessment measures available for use with school-aged children. The strengths of these measures are their theoretical foundations, reliance on empirical studies for construct development, broad standardization samples for children, inclusion of qualitative and quantitative data, and good psychometric properties.

Preferably, neuropsychological assessments should be completed by a pediatric neuropsychologist or a school neuropsychologist. There is an urgent need for more cross-training between specialists in neuropsychology and educators to ensure that the recommendations that are made in a neuropsychological assessment report can

be implemented by classroom teachers. A conceptual framework for a school-based neuropsychological evaluation was also presented in this article. The selected component parts should be based on the referral questions and the presenting behavioral and/or learning problems.

The application of neuropsychological assessments in schools is an emerging area within the broader field of psychology and school psychology. For neuropsychological principles to be fully integrated within schools, brain-based research needs to be more fully translated into practice. More specialists in neuropsychology need to be trained to be readily available to educators for consultation, assessment, and prescriptive intervention.

### See Also the Following Articles

Clinical Assessment ■ Diagnostic and Statistical Manual of Mental Disorders ■ Emotional and Behavioral Problems, Students with ■ Psychological Assessment, Standards and Guidelines for ■ Psychometric Tests ■ Psychoneuroimmunology ■ Psychophysiological Assessment

### Further Reading

- Baron, I. S. (2003). *Neuropsychological evaluation of the child*. New York: Oxford University Press.
- Feifer, S. G., & DeFina, P. A. (2000). *The neuropsychology of reading disorders*. Middletown, MD: School Neuropsych Press.
- Feifer, S. G., & DeFina, P. A. (2002). *The neuropsychology of written language disorders*. Middletown, MD: School Neuropsych Press.
- Hebben, N., & Milberg, W. (2002). *Essentials of neuropsychological assessment*. New York: Wiley.
- Levine, M. D., & Reed, M. (1998). *Developmental variation and learning disorders* (2nd ed.). Cambridge, MA: Educators Publishing Service.
- Reynolds, C. R., & Fletcher-Janzen, E. (Eds.). (1997). *Handbook of clinical child neuropsychology* (2nd ed.). New York: Plenum Press.
- Semrud-Clikeman, M. (2001). *Traumatic brain injury in children and adolescents: Assessment and intervention*. New York: Guilford Press.
- Teeter, P. A., & Semrud-Clikeman, M. (1997). *Child neuropsychology: Assessment and interventions for neurodevelopmental disorders*. Needham Heights, MA: Allyn & Bacon.



# Nicotine Addiction

Neil E. Grunberg and Jennifer M. Phillips

Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA

1. Introduction
  2. Nicotine Overview
  3. Role of Nicotine in Tobacco Use
  4. Nicotine as a Treatment for Tobacco Cessation
  5. Directions in Nicotine Addiction Research and Application
- Further Reading

## GLOSSARY

**addiction** A behavior pattern of drug use characterized by overwhelming involvement with the use of the drug, the securing of its supply, and a high tendency to relapse after withdrawal.

**classical conditioning** Learning in which the conditioned stimulus is paired with and precedes the unconditioned stimulus until the conditioned stimulus alone is sufficient to elicit the response.

**dependence** The highly controlled or compulsive use of a psychoactive drug that is reinforced by the effects of that drug and includes the occurrence of tolerance and withdrawal.

**nicotine** A chemical alkaloid ( $C_{10}H_{14}N_2$ ) that is naturally present in tobacco.

**operant conditioning** Learning in which the desired behavior or increasingly closer approximations to it are followed by a rewarding or reinforcing stimulus.

**tolerance** The need for markedly increased amounts of a substance to achieve the desired effect or a markedly diminished effect of a drug with continued use of the same amount of the substance.

**withdrawal** The syndrome of often painful or uncomfortable physiological and psychological effects that result when the use of a drug is discontinued.

Tobacco use and cigarette smoking are widespread behaviors that involve the self-administration of the addictive drug nicotine. These behaviors directly contribute to premature morbidity and mortality from a variety of diseases, including cardiovascular diseases, cancers, and respiratory diseases. A great deal of information has accumulated relevant to the reasons people use tobacco products and the role of nicotine addiction. In addition, nicotine replacement therapies are used in conjunction with behavioral and cognitive therapies to help people abstain from tobacco use.

## 1. INTRODUCTION

Tobacco use and cigarette smoking are worldwide health problems. It is estimated that more than 1 billion adults, or one-third of the global adult population, smoke cigarettes. In the United States, approximately 46.5 million adults (23.3% of the adult population) are cigarette smokers and 7.6 million adults are smokeless tobacco users. Cigarette smoking is an especially difficult problem in the adolescent population. Although smoking among high school students is beginning to decline, more than 2000 people younger than age 18 begin smoking every day in the United States. In addition to

age, factors such as gender, race, and socioeconomic status affect the likelihood that an individual will use cigarettes or other tobacco-containing products.

The high prevalence of tobacco use results in significant costs to society. The medical consequences of smoking and tobacco use are extensive and result in 430,000 premature deaths per year in the United States and more than 5 million premature deaths per year worldwide, making tobacco use the leading preventable cause of death. Smoking is significantly related to an increased risk for stroke, heart attacks, vascular diseases, and aneurysms. It is estimated that approximately one-fifth of all heart disease deaths are smoking related. Tobacco use accounts for one-third of all cancers and has been associated with approximately 90% of all lung cancer cases, the leading source of cancer-related deaths in both men and women. In addition, smoking causes several other types of cancer and pulmonary diseases, and smoking increases overall morbidity rates. Smoking and tobacco use in the United States cost approximately \$80 billion per year in health care costs and \$138 billion per year when factors such as smoking-related fires, medical care for low-birth-weight infants born to smokers, and the medical effects of second-hand smoke are taken into account.

Psychological and biological processes contribute to the initiation, maintenance, and cessation of tobacco use. People begin to use tobacco as a result of peer pressure, role modeling, responses to advertising, and other social and psychological influences that expose them to the addictive drug nicotine, which is responsible for the development of tobacco dependence. Self-administration of nicotine via tobacco becomes interwoven with environmental stimuli. Principles of learning, including classical conditioning, operant conditioning, and paired associations, act in concert with pharmacological actions of nicotine to positively and negatively reinforce this self-administration. Treatment of tobacco use, therefore, involves pharmacological and psychological interventions that work together.

## 2. NICOTINE OVERVIEW

Nicotine is one of the 4000 chemical components of tobacco smoke and is also present in smokeless tobacco. It is an alkaloid that is derived from the leaves of several plants, most commonly tobacco. Nicotine's molecular formula is  $C_{10}H_{14}N_2$ , and structurally it is composed of a pyridine and a pyrrolidine ring. The fact that nicotine

is a relatively small molecule and composed of elements that are compatible with biological systems contributes to its distribution and actions in the body. The most common route of administration of nicotine is by smoking cured and processed tobacco leaves in the forms of cigarettes, cigars, or pipe tobacco. The pH of cigarette smoke is slightly acidic (5.5–6.5) and nicotine's dissociation constant ( $pK_a$ )—the pH at which the chemical is in its ionized versus deionized form—is 8.0. As a result, nicotine in cigarette smoke does not readily diffuse across the mucous membranes of the mouth or throat. The smoke must instead be inhaled into the lungs to be absorbed across the membranes of the pulmonary alveoli. Nicotine passes into the pulmonary vascular circuit and is transported directly to the brain within 5–10 seconds of inhalation. The nicotine acts in the brain and is carried in the bloodstream throughout the body, metabolized in the liver, passed to the kidneys, and then eliminated from the body.

Psychological and behavioral variables (including stress, diet, and activity) can alter the distribution of nicotine and its rate of elimination from the body, which helps explain why people smoke more in certain situations (e.g., when under stress). The pH of chewing tobacco, snuff, and cigar and pipe smoke differs from that of cigarette smoke; as a result, nicotine is absorbed across mucosal membranes from these noncigarette tobacco products. When nicotine is absorbed across membranes in the mouth, rather than in the lungs, it is slower to act on the brain.

In the brain, nicotine's effects are mediated by several subtypes of nicotinic acetyl cholinergic receptors. These receptors are pentameric (five-unit) peptides that are found in the brain and central nervous system. The stimulation of these receptors by nicotine results in cholinergic activation and an increased release of dopamine and norepinephrine. In addition, nicotine affects the release and availability of endogenous opioid peptides, serotonin, and a variety of other biochemicals. The effects of nicotine on dopamine and endogenous opioid peptides are particularly relevant to nicotine's rewarding effects and play a major role in the biological mechanisms underlying nicotine addiction. It is well established that dopamine is involved in reward pathways that reinforce the self-administration of drugs of addiction, including nicotine, cocaine, and opiates. In addition, the endogenous opioid peptides are the naturally occurring biochemicals that mimic effects of the opiates, including morphine and heroin.

### 3. ROLE OF NICOTINE IN TOBACCO USE

To understand the role of nicotine addiction in tobacco use and cigarette smoking, it is important to recognize and define the classical triad of drug addiction: dependence, tolerance, and withdrawal. The terms drug addiction and drug dependence are often used interchangeably. According to the 1988 U.S. Surgeon General's report on nicotine addiction, drug dependence refers to the highly controlled or compulsive use of a psychoactive drug that is reinforced by the effects of that drug. The American Psychiatric Association uses a similar definition for drug dependence but also includes the occurrence of tolerance and withdrawal as necessary to establish drug dependence. Tolerance is the need for markedly increased amounts of a substance to achieve the desired effect or a markedly diminished effect of a drug with continued use of the same amount of the substance. Withdrawal is the syndrome of often painful or uncomfortable physiological and psychological effects that result when the use of a drug is discontinued.

Nicotine fits the classic profile of an addictive drug resulting in dependence as characterized by the occurrence of controlled and compulsive use as well as tolerance and withdrawal in a variety of species of animals and humans. Nicotine has been identified as the addictive drug in tobacco that is largely responsible for development of chronic tobacco use and dependence. It is an addicting drug with multiple reinforcing actions that are relevant to the maintenance of cigarette smoking and tobacco use as well as relapse during quit attempts. Cigarette smokers report positive, rewarding effects of nicotine and negative effects following abstinence from nicotine self-administration. As a result, nicotine has positive and negative reinforcing actions. These reinforcing actions are mediated by neurochemical (including dopamine and opioid peptides), electrophysiological (including brain stimulation and muscular relaxation), and psychobiological mechanisms. Administration of nicotine to animals produces similar effects and alleviates symptoms of withdrawal. Psychological processes and principles (including classical conditioning, operant conditioning, and paired associations) contribute to the self-administration of nicotine. Nicotine's biobehavioral and psychobiological effects (including body weight control, attentional focusing, and stress management) also reinforce nicotine self-administration. In addition, nicotine has potential use as a treatment for a wide variety of medical conditions.

### 4. NICOTINE AS A TREATMENT FOR TOBACCO CESSATION

Although nicotine causes addiction, it does not cause most of the health risks associated with tobacco use. Nicotine replacement therapies (NRTs) are logical treatments to help people abstain from tobacco use because they help to relieve withdrawal symptoms associated with smoking cessation. In addition, the lower dosages of nicotine supplied by current NRTs do not produce the pleasurable effects associated with tobacco products and abuse potential is relatively low.

Nicotine polacrilex gum was the first pharmacological agent approved for smoking cessation treatment by the Food and Drug Administration (FDA). In 1984, it was only available by prescription. A higher dosage gum was introduced in 1992. In 1996, the FDA approved nicotine gum for over-the-counter sales. Nicotine patches that deliver nicotine transdermally were introduced in the United States in 1991. Two variations of these patches became available over-the-counter in 1996.

Today, there also are nicotine nasal sprays, inhalers, lozenges, and a variety of other nicotine-containing products, such as lollipops and water, that are marketed as NRTs. The FDA estimates that more than 1 million U.S. smokers have been successfully treated for nicotine addiction since the introduction of NRTs. The availability of over-the-counter NRTs and media campaigns to increase smoking cessation have produced an estimated 20% increase in successful cessation attempts each year. In addition, there are other nonnicotine pharmacological treatments to help people abstain from smoking. It is important to note that abstinence success is maximized when pharmacological treatments are used in conjunction with behavioral, cognitive, and psychosocial treatments.

### 5. DIRECTIONS IN NICOTINE ADDICTION RESEARCH AND APPLICATION

Psychologists, pharmacologists, neuroscientists, molecular biologists, epidemiologists, and other health care scientists and practitioners are actively studying nicotine addiction and tobacco use with the goal of developing additional and more effective ways to prevent and treat tobacco use. Ongoing research is examining various forms of NRT, novel nicotine analogs (i.e., drugs that

are similar to nicotine), and various pharmacological treatments that alter different mechanisms underlying nicotine's actions. In addition, psychosocial and cognitive-behavioral treatments are being coupled with pharmacotherapy to optimize tobacco cessation. Progress in understanding the mechanisms and treatment of nicotine addiction is constantly being made through innovative new research and techniques in the field. Many researchers are employing various methods of brain imaging to better understand how nicotine affects specific areas of the brain and how the sites and mechanisms of action might contribute to nicotine's addictive effects. At the molecular level, a variety of strains of genetically manipulated knockout mice provide information regarding the receptor biology associated with nicotine's effects.

In addition to pharmacological and psychological studies and strategies, immunopharmacotherapy is an active focus of inquiry for tobacco treatment. Substances are being developed and tested that can immunologically seek nicotine as it enters the body, complex with it, and prevent it from binding to its receptors. Of course, such treatment requires patient compliance, so psychological variables again become relevant and important for effective treatments.

The tailoring of treatments to people who are at the greatest risk of tobacco use is also an area of emphasis in nicotine research. People with mental health conditions, including depression, anxiety disorders, and thought disorders, are at high risk for tobacco use. In fact, approximately 50% of people with mental health conditions smoke. Therefore, it is important to determine why these high comorbidities exist and how to help these people quit smoking.

Nicotine has also been proposed as a potential treatment for disorders other than nicotine addiction.

Scientists are examining the use of nicotine to treat disorders such as Alzheimer's disease, dementias, attention deficit hyperactivity disorder, skeletal muscular diseases, Parkinson's disease, obesity, anxiety, affective disorders, and schizophrenia.

### See Also the Following Articles

Alcohol Dependence ■ Drug Dependence

### Further Reading

- Benowitz, N. L. (1999). Nicotine addiction. *Primary Care*, 26, 611-631.
- Grunberg, N. E., Brown, K. J., & Klein, L. C. (1997). Tobacco smoking. In A. Baum *et al.* (Eds.), *Cambridge handbook of psychology, health and medicine* (pp. 606-611). Cambridge University Press, Cambridge, UK.
- Grunberg, N. E., Faraday, M. F., & Rahman, M. A. (2001). The psychobiology of nicotine self-administration. In A. Baum *et al.* (Eds.), *Handbook of health psychology* (pp. 249-262). Mahwah, NJ: Lawrence Erlbaum.
- Hays, J. T. (2000). Tobacco, nicotine and addiction. *Journal of Substance Use*, 5, 159-164.
- Henningfield, J. E. (1995). Nicotine medications for smoking cessation. *New England Journal of Medicine*, 333, 1196-1203.
- U.S. Department of Health and Human Services. (1988). *The health consequences of smoking: Nicotine Addiction. A report of the Surgeon General* (DHHS Publication No. CDC 88-8406). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Health and Human Services. (2001). *National Institute on Drug Abuse Research report series: Nicotine addiction* (DHHS Publication No. 01-4342). Washington, DC: U.S. Government Printing Office.



# Nondiscriminatory Assessment in Schools

Samuel O. Ortiz

St. John's University, Jamaica, New York, USA

1. Introduction
  2. Brief History of Nondiscriminatory Assessment in Schools
  3. Issues and Definitions of Test Bias
  4. Issues in Psychological Testing
  5. Nondiscriminatory Assessment
  6. Summary
- Further Reading

## GLOSSARY

**acculturation** The process by which individuals from a different culture learn about and assimilate the values, beliefs, and attitudes of the mainstream culture in which they now live, as reflected by their knowledge of or participation in the new culture, including use of the dominant language. Acculturation is like any other developmental process and is largely a function of time and exposure to the mainstream culture so that individuals who have lived in the new cultural milieu longer tend to display higher levels of acculturation—that is, more knowledge of the culture and more proficiency and use of the dominant language of the culture.

**assumption of comparability** When tests are normed on particular populations, there is an assumption that all the individuals share similar or comparable, but not necessarily identical, background or acculturative experiences. This is the reason why 10-year-olds are compared to other 10-year-olds and not to 15-year-olds. If it can be assumed that there is comparability in the backgrounds of

the individuals on whom the test was normed (namely in terms of level of acculturation and language proficiency/development), then along with other controls for variation (age, grade, socioeconomic status, region, etc.), the norms can be said to establish an appropriate base for comparing performance of individuals against the group as a whole.

**bilingual** Refers to any individual with some experience or exposure to at least two different languages. Being bilingual is not an either/or proposition; rather, it is a continuum that ranges from people with very low levels of proficiency in both languages to people with very high levels of proficiency in two or more languages, or balanced bilinguals. Generally, bilingual individuals will have a “dominant” or preferred language and their proficiency is not often identical; being very proficient in one language and limited in another is quite common. Nonetheless, bilingual individuals do not share and do not have comparable backgrounds to individuals who have experience primarily with only one language.

**culture** The idiosyncratic values, beliefs, attitudes, customs, traditions, and language that belong to a circumscribed group of people and that are transmitted as part of the collective societal knowledge and behaviors from one generation to the next. Culture thus represents a powerful environment that shapes individuals in a unique way and differently from one group to the next. Culture is often equated with or confused with race or ethnicity, but it is neither. People of varying ethnorracial backgrounds can share the same culture. Likewise, people from the same ethnorracial background can come from markedly different cultures. It is culture, not race or ethnicity, that is

important because it is the force that defines and shapes an individual's behavior and cognition.

**psychometrics** Literally means "measurement of mental abilities" and refers to the field in psychology that is devoted to establishing statistical and mathematical procedures for quantifying a wide range of human cognitive abilities. Such abilities are latent; that is, they cannot be measured directly and their existence is inferred on the basis of manifest attributes that can be directly measured.

**saturation** This is similar to the concept of acculturation and refers specifically to the degree to which an individual's knowledge, experiences, and development closely align with expectations for individuals of the same age in the society. When an individual has acquired this knowledge or experience (become acculturated) to the degree that would be considered developmentally appropriate, his or her level of saturation can be said to be normal or comparable to that of other individuals of the same age or grade.

**test validity** The degree to which a test is valid is an indication of the degree to which it measures what it purports and is designed to measure. Validity is related to reliability, or the accuracy with which a test can measure something, because a test must first be reliable (or accurate) before it can be valid. No test that is unreliable or inaccurate can be reliable.

Although the need and rationale behind nondiscriminatory assessment have been clear from the outset of the development of psychometrics, the actual practice of providing fair and equitable assessment has yet to be fully embraced by practitioners in school psychology and other major areas of psychological practice. Indeed, many of the issues related to the use of standardized tests, particularly regarding individuals from diverse linguistic and cultural backgrounds, remain poorly or incompletely understood by school psychologists. There are many reasons for this gap, including lack of training standards and models of practice; a relative dearth of practitioners and academicians with expertise in the field; confusion regarding the nature of bias or discrimination in assessment; the lack of systematic guidelines for everyday practice; and the need to develop transdisciplinary, collaborative approaches that span other professional domains, such as speech-language and education. This article provides a brief overview of these issues including a historical perspective on nondiscriminatory assessment; the consequences of inequitable evaluations; the relationship between culture, language, cognition, and education; distinguishing definitions of bias in testing; and preassessment considerations. This article also specifies steps that can be taken

to conduct a comprehensive and systematic evaluation that is in keeping with best practices on nondiscriminatory assessment.

## 1. INTRODUCTION

In its simplest form, nondiscriminatory assessment refers to assessment that is conducted in a manner so as to ensure the validity of obtained results. Generally, the tests and procedures employed in assessment have been designed for use with a wide range of diverse individuals so that adherence to standardization and other guidelines effectively render the obtained results valid. However, the issues are not quite so straightforward. Even at the very early stages of psychometric development, psychologists were well aware of the potential problems with regard to testing diverse individuals. George Sanchez (as cited in Valdés and Figueroa, 1994), a Mexican American psychologist, noted with acute insight that "as long as tests do not at least sample in equal degree a state of saturation [assimilation of fundamental experiences and activities] that is equal for the 'norm children' and the particular bilingual child it cannot be assumed that the test is a valid one for the child" (p. 106).

The issue of validity is at the heart of nondiscriminatory assessment and the failure on the part of psychologists to fully understand and delineate this variable has led to far-ranging ramifications that may be widely recognized but are often misunderstood. Despite some attention to this issue, the effects of biased and discriminatory assessment (e.g., attribution of culture-specific behavior to various forms of psychopathology, overrepresentation of African American children in special education, and misdiagnosis of the normal process of second language acquisition as a disability) continue to be seen in virtually all aspects of psychological practice, particularly in schools. How and in what manner these variables have affected perspectives and practices in nondiscriminatory assessment will be discussed further in the following sections.

## 2. BRIEF HISTORY OF NONDISCRIMINATORY ASSESSMENT IN SCHOOLS

It is important to recognize that testing, evaluation, and assessment are not interchangeable terms. In the

school setting, testing represents the most narrow activity and refers to the actual use of tests to measure some trait or characteristic. Evaluation is the process of making inferences and judgments based on data collected during testing and generally takes the form of diagnosis, disability determination, and program eligibility consideration. The main focus of the activities that constitute evaluation is on “how much” a person exhibits some trait relative to same age or grade peers. Assessment is the broadest activity of all and is defined generally as the process of problem solving based on information gathered during testing, any evaluative inferences made regarding test data, as well as on other data, including observation, interviews, work samples, and review of existing records. Assessment activities ultimately lead to a corrective solution and nondiscriminatory assessment ensures that the solution is responsive to the particular individuals, elements, and systems that comprise the problematic situation.

One of the first legal battles regarding issues of potential test bias occurred in California in *Diana v. State Board of Education* (1970). The gist of the lawsuit by the plaintiffs was that the “educable mentally retarded” (EMR) placement and diagnosis were incorrect, particularly because the manner by which the latter was determined was invalid—an IQ test administered in English to children who were predominantly Spanish speakers. However, the case was settled out of court and never went to trial because the California State Board of Education agreed to adopt certain specific remedies to the satisfaction of the plaintiffs. The remedies were largely ineffective but set certain precedents regarding the testing of students with limited English proficiency, including testing in the primary language and in English, use of nonverbal IQ, reevaluation of children previously diagnosed as EMR with particular attention to nonverbal performance, development of an intelligence test appropriate for Mexican American children (never accomplished), and monitoring of the representation of EMR Hispanic students in special education throughout the state.

The remedies in the Diana case demonstrated the growing concern over linguistic issues and the potential discriminatory aspects and outcomes in assessment and education in schools. This sentiment was further reinforced by the U.S. Supreme Court in its *Lau v. Nichols* (1974) ruling that Chinese-speaking students were being discriminated against by having school instruction provided in English only—a language they could not understand—and thus were not

receiving an equal educational opportunity. Although the primary impact of the Lau decision was the implementation of a system for evaluating schoolchildren’s linguistic proficiency and need for equitable access to the curriculum, it also carried ramifications that extended to the assessment of students with limited English proficiency. Foremost among these implications is the appreciation of the potential bias inherent in evaluating students in a language they do not comprehend or comprehend fully as well as the possible consequences of failure to evaluate performance in the native language.

In 1975, P.L. 94-142 [known as the Education of Handicapped Children Act and reauthorized several times since then as the Individuals with Disabilities Education Act (IDEA)] contained specific wording that addressed issues in assessment in schools and whether in fact it was nondiscriminatory in nature. Several stipulations in IDEA make it clear that assessment has to account for possible areas of bias or discrimination, as evidenced by the following statement that remains in current versions: “Tests and other evaluation materials used to assess a child . . . are selected and administered so as not to be discriminatory on a racial or cultural basis; and are provided and administered in the child’s native language or other mode of communication” (IDEA §300.532, Evaluation Procedures). The emphasis in the law on fair and equitable assessment is quite clear, as is the notion that “tests” in particular may prove to be discriminatory if not properly used. Soon after the implementation of IDEA, the question of the fairness of IQ tests emerged prominently, as did concerns regarding the apparent negative consequences of their use, as evidenced most often by the overrepresentation of one cultural or ethnic group in special education.

The potential discriminatory aspects of psychological tests and testing in schools is exemplified in the landmark California case known as *Larry P. (Larry P. v. Riles, 1979)*. In this case, it was argued by the plaintiffs that the use of IQ tests with African Americans resulted in improper special education placements largely because the tests were biased. Judge Robert Peckham, the presiding jurist, provided an initial ruling that still stands today (in California) that prohibited the use of IQ tests for the purpose of determining placement in “dead-end” special education classes. In other words, IQ tests could not be used to establish the intelligence level of African American children in cases of suspected mental retardation because it had not been demonstrated that such tests had been validated for this purpose.

A later order by Judge Peckham banning the use of IQ tests altogether with African American children was rescinded on procedural grounds. However, the original order remains intact and the California Department of Education later adopted policy in keeping with the restriction of IQ tests for any special education purposes involving African American children.

At approximately the same time as the Larry P. case, the issue of IQ test bias and overrepresentation was also being considered in *Parents in Action for Special Education (PASE) v. Hannon* (1980). In this case, Judge John Grady limited the question of test bias to whether IQ tests were culturally biased against African American students and even examined for himself every item and correct answer from the Wechsler Intelligence Scale for Children (WISC), WISC-Revised (WISC-R), and the Stanford–Binet IQ tests. In rendering his judicial opinion, Judge Grady found only a handful of test questions (eight on the WISC and WISC-R and only one on the Binet) that he believed were biased and subsequently ruled in favor of the defendants, citing that the plaintiffs (PASE) failed to prove the question of bias in IQ tests. The more interesting aspect of the case, however, is that Judge Grady cited that in Chicago special education placement was not based primarily or exclusively on the IQ score but rather on a rigorous and careful placement process with well-trained and knowledgeable professionals, including sufficient numbers of African American school psychologists to manage the process fairly and equitably. Judge Grady's decision, although rather tenuous on the issue of test bias, reinforced the notion that nondiscriminatory assessment is a process and includes many aspects that may create or contribute to bias—all of which must be attended to in order to ensure fairness.

Criticism and concern regarding the discriminatory aspects of assessment in schools continues almost unabated primarily because of statistics that still indicate a disproportionate representation of African Americans (and other ethnic groups) in special education programs throughout the United States. Typically, African Americans are overrepresented in the Mentally Retarded and Emotionally Disturbed categories while conspicuously underrepresented in the Specific Learning Disabled and Speech–Language Impairment categories. The U.S. Office of Civil Rights has pointed to this overrepresentation as evidence of discriminatory assessment practices (which presumably include biased testing) and has routinely used its authority to enforce compliance with procedures and practices that

do not result in such inequitable outcomes. Nonetheless, overrepresentation continues to be a national problem in schools.

### 3. ISSUES AND DEFINITIONS OF TEST BIAS

The issue of bias in testing has revolved almost exclusively around statistical or psychometrically based definitions—that is, that bias exists as an inherent flaw within the test that results in a systematic but unintended pattern of results for a particular group of individuals. Thus, bias has been examined from various aspects of test construction, including as a function of such characteristics as (i) item content or novelty; (ii) item sequence, order, or difficulty; (iii) reliability in terms of measurement error or accuracy; (iv) stability of the underlying factorial structure; (v) examiner–examinee or atmosphere effects; and (vi) predictive ability, usually related to academic achievement or success.

Various types of studies, including correlation, multiple regression, factor analysis, prediction, and other types of analyses on test characteristics, have been conducted for decades and, by and large, evidence of bias along these lines has been rather meager. For example, items on tests with clear cultural connotations have been abandoned in favor of content that is more general and universal. Sophisticated techniques for creating scales and tests whereby the items show measurably progressive levels of difficulty (e.g., Rasch scaling) have been in use for more than a decade. Virtually all IQ tests have demonstrated more than adequate forms of reliability and their underlying factor structures have not been found to change as a function of cultural or ethnically based group differences. However, perhaps the most often cited type of research with respect to issues of bias (or lack thereof) revolves around studies that examine fairness via prediction methods. Generally, a test is considered to be fair when its use does not lead to different predictions for different groups as determined by comparison of the regression lines between a predictor test. Only when the characteristics of such regression lines differ across groups is it believed that a test may be unfair. Although this method of examining bias appears intuitive and appealing, it should be noted that even perfectly matched regression lines do not automatically imply that the construct or ability underlying the test used for prediction is in fact the construct or ability that

was actually measured. A test may well predict equally well for native English speakers as it does for limited English speakers, but the nature of exactly what ability was measured in each case can differ substantially.

Overall, these results are not surprising because the sophistication and psychometric rigor of modern-day tests virtually guarantee that bias in this form will not operate because careful attention has been paid to each of these issues in the development and construction of the test. Combined with norms that are stratified on the basis of age, grade, gender, geographical location, socioeconomic status, race/ethnicity, and other assorted variables, use and application of current assessment instruments would appear to carry little risk of bias defined as such. Indeed, when bias is viewed only as a function of the psychometric properties of a test (generally related to issues of reliability), it is unlikely to be found.

However, psychometric bias is by no means that simple a matter. Tests must not only be reliable but also be valid—that is, they must measure the very constructs they were designed to measure. By definition, a test that is valid must already be reliable, but a test that is reliable is not automatically valid. Hence, the expectation that a very reliable test will show evidence of psychometric bias related to its accuracy is essentially a straw man argument because the variables that might affect reliability have already been controlled or eliminated during test development. What has not been recognized, however, is that there are powerful variables that may not affect reliability but can indeed significantly affect validity. Inequity in assessment is most likely to occur in cases in which the obtained results no longer reflect the constructs that were purportedly measured but, rather, different constructs that were unintentionally measured. In its most simple form, this concept is illustrated in the administration of a test in English to an individual with no or an incomplete understanding of English. The test no longer provides an accurate measure of ability or achievement but, rather, reflects the individual's comprehension of and proficiency in the English language. Nondiscriminatory assessment demands recognition of the fact that variations in acculturation and language backgrounds and experiences can and do affect test validity—a historically underresearched area of test bias.

According to Scarr (1978), “intelligence tests are not tests of intelligence in some abstract, culture-free way. They are measures of the ability to function intellectually by virtue of knowledge and skills in the culture of which they are a sample” (p. 339). High-quality tests such as those currently available may indeed not be

“culturally biased” per se but they remain “culturally loaded”—that is, they have embedded content that is specific and often unique to the culture that gave rise to the test. For tests developed in the United States, this would encompass content and knowledge that is known as “mainstream” or popular—information that would be acquired either incidentally through exposure and contact with the culture or formally through instruction. The process by which an individual learns the common elements of a particular culture is known as acculturation, and it is a process that advances naturally and over time, as do all developmental processes. The more frequent, simple, and common elements of the culture are learned first and the more complex ones later.

Similarly, tests that rely on an individual's ability to comprehend English can be construed as being “linguistically demanding”—that is, they require certain expected levels of language proficiency or facility with the language in order to adequately or successfully understand directions for the task or even perform the task. Learning any language, particularly a second language, is also a developmental process. Usually, the process of learning a language begins at birth, but in many cases it may begin at some point after birth, putting the acquisition and learning of that language ostensibly “behind” that of native speakers. This means that the normal expectations for language proficiency (as well as those for acculturative knowledge acquisition) for a given age or grade may not yet have been attained by those who began learning the language at a point other than from birth or who were not indoctrinated into the mainstream culture from birth. Because neither language nor acculturation are finite, there is no point at which individuals who began at points other than birth will ever “catch up” or become equivalent to their mainstream, native speaking counterparts—except in cases in which individuals were well schooled and educated in their native language first, usually up to at least age 11 or the sixth grade.

In summary, both the level of acculturation and language proficiency (generally, English language proficiency) can significantly affect the process of assessment and can alter the psychometric properties of a test with regard to its validity even though the inherent reliability of the test remains intact. Tests in particular are constructed on an assumption of comparability, which presumes a relatively similar background and experiences for the individuals on whom the test was normed. Variations in background or experiences that violate this assumption therefore render test results

effectively invalid. This is most likely to occur in cases involving culturally or linguistically diverse individuals because the test may well degenerate into one that measures level of acculturation or English language proficiency more so than the intended or expected constructs. In such situations, the tests used may not be biased, but they are not likely valid either.

Apart from the more commonly defined areas of bias, there are other definitions that are important relative to nondiscriminatory assessment, albeit they are perhaps not as widely accepted. Bias in tests is often cited as a function of the fact that a test results in (i) mean differences (generally between different cultural or ethnic groups); (ii) variable selection ratios or disproportionate classification, usually in special education; and (iii) social consequences. Indeed, mean differences and disproportionate classification formed the basis of the arguments that prompted the advancement of the historical legal challenges to the use of IQ tests in schools discussed previously. Although the result may well be pernicious, it is important to note that bias definitions of this type reflect more the application and use of tests than any inherently discriminatory property of tests. This shift in the nature and definition of bias is particularly evident with respect to the issue of social consequences, or more accurately the negative social consequences of the use of IQ tests, which generally arises out of recognition of a more general failure on the part of special education to provide positive outcomes. Frequently, use of IQ tests leads to disability classification, which then leads to placement in special education organized around the disability label. When combined with the negative findings on attempts to match instruction to cognitive or neuropsychological strengths, the efficacy of labeling children via use of IQ tests results leads to decidedly negative and discriminatory social consequences.

#### **4. ISSUES IN PSYCHOLOGICAL TESTING**

Perhaps the area that psychologists tend to focus on most when considering nondiscriminatory assessment in general is that which pertains to what is the best test to use. This may be viewed as somewhat of a search for the Holy Grail. Sattler (1992) expressed this illusion best when he stated that “probably no test can be created that will entirely eliminate the influence of learning and cultural experiences. The test content

and materials, the language in which the questions are phrased, the test directions, the categories for classifying the responses, the scoring criteria, and the validity criteria are all culture bound” (p. 579).

As noted previously, there are no tests that have norms suitable for individuals who are either bilingual or bicultural (or, frequently, both). Level of acculturation and English language proficiency have yet to be stratified along with other variables in the construction of norm samples. This is problematic because in the United States, whenever an individual who does not speak English or is limited in English proficiency enters the public school system, he or she will invariably be required to learn English, or in a sense become a circumstantial bilingual. Even when such individuals become “English dominant” or “fully English proficiency,” they do not cease to be bilingual. Their dual-language backgrounds are not erased simply because they are now able to communicate at a conversational level and with no trace of accent. Even nonverbal tests, which reduce the inherent problems with communication and linguistic loading, do not address the issues of acculturation and the norms remain based primarily if not exclusively on monolingual English speakers. Native language tests also have the same problem. That is, although they tend to reduce bias that may be due to language comprehension or proficiency issues, they nevertheless are normed on monocultural, monolingual speakers of the native language—a group that is no more an appropriate standard for comparison than monocultural, monolingual English speakers when the examinee happens to be bilingual and bicultural.

The only practical recommendation that may be offered with respect to nondiscriminatory assessment is to first recognize that use of any test with diverse individuals will result in some form of bias but that the degree of bias can vary greatly. Initially, psychologists seemed intent on finding ways in which an actual IQ could be “adjusted” or modified so as to offer a better estimate of an individual’s true functioning, often by taking into account acculturation issues as well as norm sample issues in estimating the eventual IQ. However, such practices, as well as others that have utilized nonstandardized administrations, altered norms, and the like, have been met with harsh criticism and have failed to provide a viable framework for nondiscriminatory assessment. A more promising practical approach to dealing with the potential bias in the use of tests seems to rest in simply recognizing that all tests carry differing levels of expectations regarding language ability, embedded cultural content,

or both. By selecting tests that have less cultural loading and lower linguistic demands, more confidence can be placed in the fact that what is being measured is more a function of actual ability than lack of acculturation or English proficiency. This is not to say that the entire process of assessment has been rendered equitable but, rather, that the discriminatory aspects and potential in the use of tests have been considered and accommodated to a certain degree.

An example of a systematic framework for engaging in this type of selection and interpretive process is available via use of the CHC Culture–Language Matrix. Initially, the matrix provides a visual representation of cognitive ability tests as classified according to the degree of cultural loading and linguistic demand that allows practitioners to select tests that are lower along these dimensions and thus more likely to be equitable or fair. The matrix also provides a method for evaluating actual vs expected patterns of performance that draws upon research in order to guide interpretations regarding the validity of the results relative to the primary or contributory effects of cultural or linguistic differences.

By allowing practitioners to select tests that may provide fairer estimates of true ability as well as a practical way to gauge the extent to which cultural or linguistic factors may have played a role in influencing test performance, potential bias and discrimination in test use can be effectively reduced, albeit never eliminated entirely. Nonetheless, when it can be ascertained that cultural or linguistic factors did not play a primary role in determining the obtained pattern of test results, a significant portion of the potential inequity (related to validity) has been resolved and practitioners can exercise more confidence in the validity of their results. Ultimately, nondiscriminatory assessment is not dependent on so-called culture-free or nonverbal tests but rather on a process that recognizes the nature and source of possible invalidity and that seeks to reduce potential bias to the maximum extent possible.

## 5. NONDISCRIMINATORY ASSESSMENT

Several recommendations have been provided in the literature for engaging in various forms of nondiscriminatory assessment. Although many discrete methods for reducing bias in assessment have been described

throughout the years, much less attention has been paid to the development of a broad, comprehensive framework for nondiscriminatory assessment. Within the field of school psychology, Ortiz provided a framework that detailed a 10-step process for nondiscriminatory assessment. This process included guidelines and recommendations on how to assess and evaluate the learning ecology; assess and evaluate language proficiency; assess and evaluate opportunity for learning; assess and evaluate educationally relevant cultural and linguistic factors; evaluate, revise, and retest hypotheses; determine the need for and language(s) of assessment; reduce bias in traditional testing practices; utilize authentic and alternative assessment procedures; evaluate and interpret all data within the context of the learning ecology; and link assessment to intervention.

The framework delineated by Ortiz integrates salient and promising procedures and recommendations for nondiscriminatory assessment culled from research and practice. It is intended to be both a linear guide to assessment and a recursive model that allows a return to previous steps as may be necessary when newly collected data suggest alternative hypotheses. The framework can accommodate assessment activities at the individual and team levels and is particularly suited toward collaborative assessment in which members of an assessment team (including parents) can work together, sharing data and information and drawing conclusions or making decisions jointly. The need to consider multiple sources of information and multiple points of view in any assessment underlies the principles of nondiscriminatory assessment and serves to enhance the validity and ultimately the fairness and equity of the evaluation.

## 6. SUMMARY

Nondiscriminatory assessment should not be viewed as a single procedure, method, or test. Rather, fair and equitable assessment results from the integration of a wide range of approaches that together form the basis for drawing valid conclusions regarding the meaning of the collected data. Decisions regarding functioning and performance will likely prove to be less discriminatory in cases in which the sources of bias have been recognized a priori and systematic steps were taken to control their influence in both the collection and the interpretation of data. Nondiscriminatory assessment should not be seen merely as a search for nonbiased

tests—an impossibility—or a form of assessment that is applied only in cases of cultural or linguistic diversity. The provisions in IDEA and other guidelines for nondiscriminatory assessment do not assert that it is intended for select groups only but rather that it should be conducted in a way that ensures that all individuals are evaluated in the least discriminatory manner possible. As such, nondiscriminatory assessment is an approach to assessment that is equally applicable and appropriate to any individual and there is no reason it cannot be applied as such.

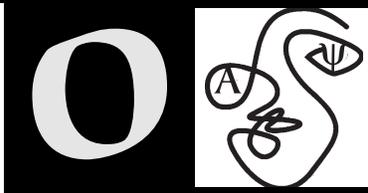
### See Also the Following Articles

Acculturation ■ Cross-Cultural Psychology, Overview ■ Educational and Child Assessment ■ Intelligence and Culture ■ Intelligence Assessment ■ Personality Assessment ■ Psychological Assessment, Standards and Guidelines for ■ Psychometric Tests ■ Training, Cross-Cultural ■ Translation

### Further Reading

Denbo, S., Beaulieu, L. M., & Coleman, J. L. (2002). *Improving schools for African American students: A reader for educational leaders*. Springfield, IL: Thomas.

- Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: Wiley.
- Kavale, S. R., & Forness, K. A. (2000). *The efficacy of special education*. Washington, DC: American Association on Mental Retardation.
- Ortiz, S. O. (2002). Best practices in nondiscriminatory assessment. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology IV*. Washington, DC: National Association of School Psychologists.
- Patton, J. M. (1998). The disproportionate representation of African Americans in special education: Looking behind the curtain for understanding and solutions. *Journal of Special Education*, 32(1).
- Sandoval, J., Frisby, C. L., Geisinger, K. F., Scheuneman, J. D., & Grenier, J. R. (1998). *Test interpretation and diversity: Achieving equity in assessment*. Washington, DC: American Psychological Association.
- Sattler, J. (1992). *Assessment of children* (Rev. 3rd ed.). San Diego: Sattler.
- Scarr, S. (1978). From evolution to Larry P., or what shall we do about IQ tests? *Intelligence*, 2, 325–342.
- Skiba, R. J. (2001). When is disproportionality discrimination? The overrepresentation of black students in school suspension. In W. Ayers, B. Dohrn, & R. Ayers (Eds.), *Zero tolerance: Resisting the drive for punishment in our schools* (pp. 176–187). New York: New Press.
- Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex.



# Obesity

Shannon M. Kehle, Jessica S. Benas, and Michael A. Friedman

Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

1. Basic Knowledge of Obesity
  2. Problems Associated with Obesity
  3. Treatments
  4. Future Directions
- Further Reading

## GLOSSARY

**binge eating disorder (BED)** Recurrent episodes of eating large quantities of food while feeling a loss of control without the use of compensatory purging behaviors.

**body image** A person's concept of his or her physical image.

**body mass index (BMI)** A common measurement to assess obesity; BMI is the relation between a person's height and weight, expressed by weight in kilograms per height in meters squared.

**obesity** An accumulation of excess adipose tissue in relation to lean body mass.

**reasonable weight loss** A 5 to 10% loss of current weight; losses of this magnitude are achievable for most people and will lower their risk of developing obesity-related diseases.

**stigma** Being viewed as different from normative expectations based on cultural and situational factors; obese individuals who are stigmatized are viewed as possessing certain undesirable characteristics that are not necessarily accurate of every obese person.

**toxic environment** An environment in which high-fat, high-calorie foods are advertised heavily, offered at low costs, and readily available.

**weight cycling** Recurring instances of weight loss and regain, also known as yo-yo dieting.

Obesity is chronic condition that has recently reached epidemic proportions worldwide. Obesity is associated with serious medical and social consequences. Although the psychological consequences of obesity are less consistent, body image disturbances and binge eating disorder are common. While there are efficacious medical and behavioral treatments, maintenance of weight loss continues to be a problem and is a direction for future research.

## 1. BASIC KNOWLEDGE OF OBESITY

### 1.1. Definition

Obesity is defined as possessing a body mass index (BMI) of at least 30 kilograms per square meter ( $\text{kg}/\text{m}^2$ ), and morbid/extreme obesity is defined as possessing a BMI of at least 40  $\text{kg}/\text{m}^2$ . Besides using BMI, there are various ways in which to measure total body fat and classify obesity such as imaging techniques to assess localization of body fat, underwater weighing, skinfold thickness, and waist/hip circumference ratio. Obesity is often a chronic condition in which individuals who attempt to diet subsequently regain their lost weight, leading to "weight cycling."

### 1.2. Prevalence

Obesity is a health risk that has reached epidemic proportions worldwide, with more than 300 million adults

considered to be obese. Although prevalence estimates have differed due to various classification methods, recent statistics cite that nearly one-third of adult Americans are obese, a twofold increase over the past decade. Childhood obesity has also increased, with estimates suggesting that 15% of children and adolescents were obese in 2000, a 4% increase over the previous decade. In addition, there has been a marked increase in rates of morbid obesity; the prevalence rose threefold between 1990 and 2000.

### 1.3. Causes

Overall, obesity is the result of an individual consuming more calories than he or she expends, with an additional 3500 calories eaten leading to 1 pound gained. A complex interaction of environmental, genetic, and behavioral factors influences both consumption and expenditure. Factors related to consumption include food choices as well as food availability. For example, a diet high in fat will provide more energy than will carbohydrates or proteins and will produce fewer satiety signals, thereby disabling the individual's ability to suppress hunger cues. Currently, we live in a "toxic environment" such that the availability of unhealthy foods increases the likelihood that poor decisions are made. For example, there is an increasing number of commercials for calorie-dense foods, such as candy and high-sugar cereal geared toward children, and an abundance of easily accessible fast-food restaurants offering fatty foods at low costs.

Caloric expenditure is influenced by several factors, including resting metabolic rate and level of activity. Although a range of genetic factors most likely influences metabolism, the U.S. society is moving toward an increasingly sedentary lifestyle given that many Americans are engaging in less planned and lifestyle exercise. Fewer than one-third of Americans get the minimum recommended amount of exercise of light to moderate activity five times per week. Thus, as a society, Americans are consuming more and expending less, causing a rise in rates of obesity.

## 2. PROBLEMS ASSOCIATED WITH OBESITY

### 2.1. Medical Consequences

Obesity is strongly associated with diabetes mellitus; close to 50% of all individuals with type 2 diabetes have a BMI of at least 30 kg/m<sup>2</sup>. This relation results from insulin's decreased ability to act at a cellular level

when there is an excess of fatty tissue in the body. Obesity is also a risk factor for the development of coronary heart disease (CHD). Obesity contributes to the development of CHD through increasing rates of risk factors such as hypertension and increased cholesterol. Recently, an association between obesity and cancer has also been found. Obesity is associated with increased risk for certain cancers, including colon, endometrial, and breast cancers, and is also related to increased mortality rates among cancer patients. The causal mechanisms linking obesity to cancer are currently unknown.

Obesity is also associated with other complications such as disruption of sleep, difficulties with daily activities (e.g., climbing stairs, walking, bending, kneeling), increased levels of lower back pain, shortness of breath, and fatigue. As a result, impairments in daily business, work productivity, and role functioning are significantly more common. Overall, obesity accounts for approximately 300,000 deaths nationwide each year, and treatment costs were as high as \$117 billion in 2000.

### 2.2. Social Stigma

Obese individuals are not only condemned for their appearance but also blamed for their condition. In a society that espouses a thin and lean body ideal, obese individuals are viewed as less attractive. Furthermore, body weight is viewed as controllable, leading obese individuals to be perceived as lacking the self-discipline necessary to maintain thinness. As a result, obese individuals are often thought to be lazy, gluttonous, and stupid. These assumptions have led to discrimination against the obese in areas such as employment, education, housing, and medical care.

### 2.3. Psychological and Behavioral Consequences

Efforts aimed at clarifying psychological consequences of obesity have provided inconsistent results. The obese population appears to be heterogeneous with respect to psychological functioning, with a subset of individuals experiencing psychological consequences of their weight and others exhibiting healthy mental functioning. To gain a more comprehensive understanding of the psychological correlates of obesity, researchers have suggested an approach focused on identifying risk factors that may explain the occurrence of negative

psychological consequences of obesity. Potential risk factors for psychological distress include being female, being adolescent, and being severely overweight. In addition, a subgroup of obese individuals who suffer more psychological distress are binge eaters. Binging refers to eating a large quantity of food (an amount larger than an individual would normally eat) in a discrete amount of time and feeling a loss of control. When individuals engage in binge eating more than twice a week and do not engage in compensatory purging behaviors (e.g., vomiting), they may be diagnosed with binge eating disorder (BED). Obese binge eaters tend to display higher levels of psychological distress and social dysfunction than do nonbinging obese individuals.

One area where obesity carries a more consistent risk for suffering is body image dissatisfaction. Body image concerns have been found to increase steadily with BMI, and body image disorders are thought to be present in approximately 10% of the obese population. Body image disorders often result in emotional distress due to preoccupation with appearance, negative self-worth, and social isolation.

### 3. TREATMENTS

#### 3.1. Medical Approaches

There are two medical approaches currently available for the treatment of obesity: surgery and pharmacological treatments. Because of potentially life-threatening complications, surgery is recommended only for the morbidly obese (those with a BMI of at least 40 kg/m<sup>2</sup>) or for those with a BMI of at least 35 kg/m<sup>2</sup> and a comorbid physical illness. Bariatric surgery produces its effects through either limiting the stomach's capacity or interfering with the digestion of food. Gastric bypass is widely accepted as the surgery of choice due to its low mortality rate and the ability to produce losses of up to 70% of excess body fat. Surgical treatments have also been shown to effectively improve type 2 diabetes, cardiac conditions, and quality of life among obese individuals. Although the initial effects of the surgery are drastic, weight regain is likely without significant behavior modifications; the average regain is 15% of initial excess weight.

Pharmacological treatment may be appropriate for individuals with a BMI of 30 to 40 kg/m<sup>2</sup> and for the morbidly obese who do not opt for surgical treatment. Although weight loss drugs have proven to be unsafe in the past, significant advances have been made in this area. There are currently two weight loss drugs that

have been approved by the Food and Drug Administration: sibutramine and orlistat. Treatment with sibutramine increases satiety levels and produces average losses of approximately 5% of total body weight with a relatively low side effect profile. Orlistat blocks absorption of approximately 30% of dietary fat consumed daily and has weight loss effects approximately equal to those seen with sibutramine. Both drugs have been shown to be more efficacious when combined with dieting, and losses of more than 10% of original body weight are common. The long-term effects of pharmacological treatments are less promising. There is evidence that continued long-term use may lead to decreased efficacy, and discontinuation of the drugs is associated with weight regain.

#### 3.2. Very Low-Calorie Diets

Very low-calorie diets (VLCDs) are defined as programs that allow fewer than 800 calories per day. The goal of such programs is to produce extremely rapid weight loss while maintaining muscle mass. This is generally accomplished through consuming small amounts of prepackaged foods and powdered supplements. VLCDs are appropriate only for those with a BMI of at least 30 kg/m<sup>2</sup> due to the muscle depletion that would occur in lower weight individuals. For the obese, VLCDs have been found to be effective in short-term weight loss and safe when used under medical supervision. Weight losses of approximately 20 to 25% of initial body weight typically occur within 16 weeks of treatment. However, individuals who use VLCDs find it difficult to maintain their losses over a 1-year period. When considering both the risk of malnutrition and the large level of weight regain following VLCD treatment, there is substantial doubt as to whether VLCDs are preferable to more moderate diets.

#### 3.3. Behavioral Approaches

Behavioral weight loss (BWL) approaches, also known as cognitive-behavioral treatments, are currently the psychological treatment of choice for obesity. BWL treatment for obesity is generally conducted in a group format and occurs once per week for 4 to 6 months. The primary goals of BWL treatments are to increase physical activity levels and decrease caloric restriction so as to create a negative energy balance. Patients are encouraged to set reasonable health goals and follow moderate dieting plans that provide sufficient nutrition and allow for flexibility. Some of the core components of

BWL for obesity include (a) self-monitoring, (b) identification of factors that influence eating, (c) stimulus control, and (d) cognitive restructuring. Self-monitoring is an essential component of behavior change. Patients record factors such as their food intake, emotions while eating, and daily exercise. These records allow patients to identify and work toward changing behavior patterns. Once eating cues are discovered, stimulus control is used to limit exposure to such cues. Finally, cognitive restructuring is used to reshape attitudes about dieting, weight, and body image.

The effects of BWL have been widely studied. BWL has been shown to produce average weight losses of 7 to 10% of total body weight and to significantly improve both the medical and psychological consequences of obesity. However, as with other treatments, maintaining initial losses has proven to be difficult for patients. Approximately one-third of the losses are typically regained within 1 year, and slow weight gain tends to continue until baseline levels are met. It appears as though patients are unable to continue with the behavioral changes over long periods of time. One possible reason is that the degree of weight loss that patients have experienced is not rewarding enough to continue with the challenging lifestyle modifications.

### 3.4. Binge Eating Disorder

During the short time that BED has been recognized as a distinct eating disorder, several effective treatments have emerged. Two specific psychological treatments for BED have been tested: cognitive-behavioral therapy (CBT) and interpersonal therapy (IPT). The focus of CBT for BED is similar to the focus of CBT (BWL) for the non-bingeing obese; however, an additional emphasis is placed on moderation of food intake so that the amount consumed is neither restrictive nor excessive. Initial studies of CBT for BED have shown that treatment cuts binge frequency in half and that approximately 60% of patients abstain from bingeing for 1 year. Self-help versions of CBT also show promising results, with abstinence rates approximately equal to those of group CBT treatments.

IPT for BED is a directive treatment that focuses on managing interpersonal difficulties that are thought to result in emotional states that trigger binge episodes and maintain problem behaviors. Specific eating behaviors are rarely discussed within treatment, and weight loss strategies are not included; eating and weight are addressed only as they relate to interpersonal issues. Controlled studies of the efficacy of IPT have revealed short-term and follow-up results approximately equal

to those of CBT. Finally, BWL strategies as used in the treatment of obesity have been shown to be as efficacious as CBT and IPT in reducing binge frequency. In addition, BWL programs show an added benefit of weight loss that is not seen in CBT or IPT. For this reason, BWL programs are considered by many to be the preferred treatment for BED.

### 3.5. Strategies for Weight Loss Maintenance

Effective weight loss treatments are available; however, the maintenance of losses continues to be a challenge. Three primary strategies aimed at increasing maintenance rates have been proposed recently. The first involves combining medical approaches with behavioral lifestyle modifications. It has been shown that combining surgery with psychological approaches aimed at healthy eating and exercise may be effective in maintaining initial losses. Also, the addition of lifestyle modification has been shown to more than double the effects of drugs alone and to prevent weight regain during the maintenance phase of treatment. Therefore, to provide optimal treatment, behavioral treatments need to be considered a standard adjunct to pharmacological interventions.

The second strategy is extending the length of existing CBT treatments. This has been shown to result in continued adherence to behavioral changes and, in turn, to improved rates of maintenance. The focus of the extended contact would be teaching problem-solving techniques that could be used in challenging situations and stressing the importance of maintaining behavioral changes. Although such strategies have been proven to be efficacious, attrition rates following initial weight loss are high; therefore, interventions aimed at increasing retention rates are necessary.

The third proposed strategy is to modify patients' expectations regarding the amount of weight loss that can be expected. For example, current CBT weight loss programs are typically able to produce losses of approximately 10% of initial body weight. Although a loss of 10% substantially improves health-related quality of life and medical complications related to obesity, patients typically expect greater weight losses and are disappointed by modest results. It has been proposed that the dissatisfaction experienced with modest losses, coupled with the effort required to maintain such losses, leads to an abandonment of behavioral changes and subsequent weight regain. Therefore, experts are

now suggesting that patients aim for modest losses rather than “ideal” losses. A new form of CBT for obesity that includes the modification of beliefs regarding the desired amount of weight loss and places an emphasis on acceptance of modest losses is currently being tested. Although realistic expectations regarding weight loss may increase maintenance rates, convincing obese individuals to engage in treatment for minimal losses will undoubtedly prove to be difficult.

#### 4. FUTURE DIRECTIONS

Given the growing prevalence of obesity worldwide, the demand for effective treatments continues to increase. In response, researchers are continually working to improve existing treatments and develop new approaches. Medical breakthroughs will undoubtedly continue, and new psychological treatments are currently being developed and evaluated. However, obesity likely will remain a difficult disorder to treat. Brownell recently argued that a preventive approach, aimed at combating the toxic environment that supports obesity, is necessary to reverse the obesity epidemic. Brownell suggested that a public health approach with government initiatives, such as taxing fattening foods, removing unhealthy foods from public schools, and providing federal funding for nutrition programs, will be necessary to eliminate the environment that supports overeating and inactivity. Until such steps are taken, the ability of weight loss treatments to produce enduring change may remain limited. Public health policy might prove to be necessary in preventing obesity and maintaining long-term weight loss.

#### See Also the Following Articles

Health and Culture

#### Further Reading

- Agras, W. S., Telch, C. F., Arnow, B., Eldredge, K., & Marnell, M. (1997). One-year follow-up of cognitive behavioral therapy for obese individuals with binge eating disorder. *Journal of Consulting and Clinical Psychology, 65*, 343–347.
- Allison, D. R., Fontaine, K. R., Manson, J. E., Stevens, J., & Van Itallie, T. B. (1999). Annual deaths attributable to obesity in the United States. *Journal of the American Medical Association, 27*, 1530–1538.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Brownell, K. D. (1991). *The LEARN program for weight control*. Dallas, TX: American Health Publishing.
- Brownell, K. D. (2004). *Food fight*. New York: McGraw-Hill.
- Carter, J. C., & Fairburn, C. G. (1998). Cognitive-behavioral self-help for binge eating disorder: A controlled effectiveness study. *Journal of Consulting and Clinical Psychology, 66*, 616–623.
- Centers for Disease Control and Prevention. (2003). *Defining overweight and obesity*. [www.cdc.gov](http://www.cdc.gov).
- Fairburn, C. G., & Brownell, K. D. (Eds.). (2002). *Eating disorders and obesity* (2nd ed.). New York: Guilford.
- Fairburn, C. G., & Wilson, G. T. (Eds.). (1993). *Binge eating*. New York: Guilford.
- Foster, G. D., Wadden, T. A., Vogt, R. A., & Brewer, G. (1997). What is a reasonable weight loss? Patients' expectations and evaluations of obesity treatment outcomes. *Journal of Consulting and Clinical Psychology, 65*, 79–85.
- Friedman, M. A., & Brownell, K. D. (1995). Psychological correlates of obesity: Moving to the next research generation. *Psychological Bulletin, 117*, 3–20.
- Gortmaker, S. L., Must, A., Perrin, J. M., Sobol, A. M., & Dietz, W. H. (1993). Social and economic consequences of overweight in adolescence and young adulthood. *New England Journal of Medicine, 329*, 1008–1012.
- Jeffery, R. W., Drewnowski, A., Epstein, L. H., Stunkard, A. J., Wilson, G. T., Wing, R. R., & Hill, D. R. (2000). Long-term maintenance of weight loss: Current status. *Health Psychology, 19*, 5–16.
- National Institute of Health. (2003). *Statistics related to overweight and obesity*. [www.niddk.nih.gov](http://www.niddk.nih.gov).
- National Task Force on the Prevention and Treatment of Obesity. (1993). Very low-calorie diets. *Journal of the American Medical Association, 270*, 967–974.
- Perri, M. G., Nezu, A. M., Patti, E. T., & McCann, K. L. (1989). Effect of length of treatment on weight loss. *Journal of Consulting and Clinical Psychology, 57*, 450–452.
- Pi-Sunyer, F. X. (1993). Medical hazards of obesity. *Annals of Internal Medicine, 119*, 655–660.
- Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology, 51*, 390–395.
- Puhl, R., & Brownell, K. D. (2003). Ways of coping with obesity stigma: Review and conceptual analysis. *Eating Behaviors, 4*, 53–78.
- Rosner, S. (2002). Obesity: The disease of the twenty-first century. *International Journal of Obesity, 26*, S2–S4.
- Wadden, T. A., Berkowitz, R. J., Sarwer, D. B., Prus-Wisniewski, R., & Steinberg, C. (2001). Benefits of lifestyle modification in the pharmacologic treatment of obesity. *Archives of Internal Medicine, 161*, 218–227.
- Wadden, T. A., Brownell, K. D., & Foster, G. D. (2002). Obesity: Responding to the global epidemic. *Journal of Consulting and Clinical Psychology, 70*, 510–525.
- World Health Organization. (2003). *Nutrition*. [www.who.int](http://www.who.int).



# Occupational Choice

Jo-Ida C. Hansen and Sarah K. Stoeberl

University of Minnesota, Minneapolis, Minnesota, USA

1. Career Theories
  2. Career Counseling for Occupational Choice
  3. Conclusion
- Further Reading

## GLOSSARY

**maxi-cycle** A series of life stages characterized as a sequence of growth, exploration, establishment, maintenance, and disengagement.

**satisfaction** A perception of satisfaction from the perspective of the employee.

**satisfactoriness** A perception of satisfaction from the perspective of the employer.

**Self-Directed Search** A self-administered inventory that helps people to explore what occupations to follow and to estimate related abilities.

**Strong Interest Inventory** A self-report inventory that inquires about interests in a wide range of occupations, occupational activities, hobbies, leisure activities, school subjects, and types of people.

What makes people who are big lottery winners keep their jobs even if they are in low-paying positions? Of 16 women in a small town (population 736) in Minnesota who won \$95.5 million (approximately \$2 million each after taxes), 15 claimed that the

money was not reason enough to quit their morning jobs as school cafeteria workers or their second jobs driving school buses, cleaning houses, or milking cows. And what about the 16th woman? At 70 years of age, she already was considering retirement. Their story is not unusual. In 2004, Arvey and colleagues found that few lottery winners quit working completely, although some do take sabbaticals or change positions. These women certainly did not stay in their jobs due to the high wages they earned; their pay ranged from \$7.75 to \$10.68 per hour. Rather, research shows that a variety of internal and external factors keep people on the job—some grudgingly, of course, but others with passion and zeal for what they do. Obviously, people who are not big lottery winners do work for financial reasons—to meet basic needs or to be able to afford nonwork activities that cost money. But beyond income, work gives people a feeling of belonging and provides structure. It gives people a sense of satisfaction and productivity while satisfying their interests, promoting their self-esteem, and providing them with status and recognition. Work also provides a personal identity and a mechanism for describing oneself to others. Often, what people do is who they are. Most adults spend one-third to one-half of their waking hours at work. It follows, then, that work can have a tremendous impact on psychological and physiological well-being. It is little wonder that the process of making occupational choices can create anxiety as one searches for the perfect match.

## 1. CAREER THEORIES

Since the early 1900s, scholars have studied occupational choice and career decision making. They focused first on improving personal satisfaction, and later on increasing performance and lowering turnover, through informed occupational choices. Several theories have emerged that identify the variables important to the process of occupational choice and work adjustment. The theory of work adjustment focuses on the congruence between the abilities and needs of the individual and the requirements and rewards of the job, Super's life span theory focuses on the developmental process of occupational choice, and Holland's theory of vocational personality type focuses on prediction of the types of work that people will choose. These three theories complement one another and offer a framework for understanding occupational choice and adjustment. Career theories also provide a mechanism for conceptualizing career concerns and client goals. For example, the theory of work adjustment can guide conceptualizing clients with adjustment problems in the workplace, Super's theory can lend understanding to the influence of developmental stages in occupational choice, and Holland's theory can provide a model for describing the world of work.

### 1.1. Theory of Work Adjustment

The theory of work adjustment, stemming from the Work Adjustment Project at the University of Minnesota, first appeared during the mid-1960s and has subsequently been revised and extended a number of times. The theory describes the dynamic interaction between persons and their work environments that influences work adjustment and occupational choice.

According to Dawis and Lofquist, the architects of the theory of work adjustment, individuals "inherently seek to achieve and maintain correspondence with their environment," where correspondence is defined as a well-balanced relationship between the characteristics of the individual and the requirements of the environment. Indeed, much of the emphasis of the theory concerns the matching of the abilities and values of an employee with the abilities required by a job and the reinforcers provided by the job. Consequently, the theory provides a model for career counseling that is designed to help individuals achieve (a) a clear understanding of their own abilities, values, personalities, and interests; (b) a knowledge of the

requirements and conditions of success in a work environment; and (c) a true understanding of the interaction between these two groups.

Dawis and Lofquist maintained that abilities are general dimensions that underlie groupings of required skills, including general learning ability, verbal ability, numerical ability, spatial ability, form perception, clerical ability, eye-hand coordination, finger dexterity, and manual dexterity. Similarly, values are seen as representing a grouping of needs and are defined as achievement, comfort, status, altruism, safety, and autonomy. Two additional components to the prediction of work adjustment and occupational choice are satisfaction (i.e., a perception of satisfaction from the perspective of the employee) and satisfactoriness (i.e., a perception of satisfaction from the perspective of the employer). Put another way, satisfaction refers to the extent to which employees' needs are fulfilled by the work they do, whereas satisfactoriness concerns the appraisal by others, usually employers, of the extent to which the employees adequately complete the work that is assigned to them.

The theory of work adjustment is heavily based on the premise that individuals differ in their abilities, needs, values, and interests, among other variables. This principle views people as complex individuals who differ on a number of dimensions and suggests that counseling needs to be individualized to allow everyone to make a personalized occupational choice. Although the theory is broader than many career theories and has important implications for career adjustment and occupational choice, it focuses on only one point in time. Super's life span theory, on the other hand, focuses on the developmental process of occupational choice.

### 1.2. Super's Life Span Theory

Probably no one has written as extensively about career development or influenced the study of the topic as much as has Super. His earliest theoretical statements were influenced by researchers in various areas of psychology, sociology, and personality theory. In fact, Super described his theory of career development as a "segmental" theory or one that includes the work of many other theorists such as Freud, Maslow, and Rogers. From the works of these theorists, Super derived basic assumptions that allowed him to develop his own developmental process of occupational choice.

A hallmark of Super's theory is the view that occupational development is a process of making several decisions over time, culminating in occupational

choices that represent an implementation of the self-concept. Super viewed self-concept as a combination of biological characteristics, social roles people play, and evaluations of the reactions that others have toward the individual. For Super, self-concept served to organize roles that people play throughout their lifetimes.

An important idea throughout Super's career development theory is the concept of role. Super described six major roles—homemaker, worker, citizen, leisurite, student, and child—and stressed that these roles, as well as the importance of these roles, vary throughout one's life. For instance, he argued that during childhood, the roles of leisurite, student, and child are particularly important, while the roles of worker, citizen, and homemaker are minimal. During adolescence, citizen and worker may become more important roles, but work is not often directly related to one's eventual career. It is during adulthood that one has more occupational choices. Consequently, it is important that professionals understand the concept of life roles (in conjunction with developmental stages) when helping individuals to make occupational choices.

The notion of developmental stages is essential to Super's life span theory. Super proposed a series of stages or "maxi-cycles" over the life span, beginning with the growth stage during early childhood. During this stage, children become curious about life and begin to explore their environments. The information gained through this exploration will be an important component of the children's self-concept. The exploratory stage begins with adolescents' awareness that an occupation will be an aspect of life and ends with young adults choosing occupations that they believe are within reach and provide important opportunities. The establishment stage, as one would expect, relates to adults' early encounters within actual work experiences. During the maintenance stage, employees attempt to continue or improve their occupational situations. The final stage, disengagement, includes the preretirement period during which employees' emphasis on work is focused on retaining their positions rather than enhancing them. This period ends when older adults withdraw from the world of work.

Super proposed characteristic developmental tasks within each stage. Mastery of these tasks allows individuals to function effectively in their life roles within that stage and prepares them for the next task. Successful coping with the requirements of each stage is dependent on the individual's career maturity—a group of physical, psychological, and social characteristics that represent the individual's readiness and ability to deal with the

developmental problems and challenges that are faced. The person whose maturity is equal to the problem probably resolves it with minimal difficulty or concern. However, when the maturity is not sufficient for the task, responses of procrastination, ineptness, and/or indecision are likely to occur.

As was mentioned previously, a fundamental aspect of Super's theory is that occupational choice is an implementation of the self-concept. The implication for career counselors is that, in addition to providing clients with information about who they are, counselors need to integrate objective information about the self (e.g., interests, values, abilities). Holland's theory of vocational personality type focuses on this objective information.

### 1.3. Holland's Theory of Vocational Personality Types

Holland's theory continues to enjoy attention as it celebrates more than four decades of popularity. Holland's person–environment typology and theory of career choice clearly is the most widely studied career theory in history. In addition, the concepts that derive from this theory are integral to the vocabulary, tools, and processes of career counseling.

Holland's theory of occupational choice is based on several assumptions. One assumption is that persons and environments can be categorized according to six types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional.

1. Realistic individuals are likely to enjoy using tools or machines in their work and often approach problems in a practical or problem-solving manner. Likewise, a Realistic environment is one in which the work setting includes tools, machines, and/or animals that the individual manipulates and where the ability to work with "things" is more important than the ability to interact with others.

2. Investigative persons are likely to enjoy challenges that require the use of intellect and are apt to enjoy learning and feelings of confidence about their ability to solve mathematical and scientific problems. Similarly, an Investigative environment is one in which people search for solutions to problems through mathematical and scientific interests and competencies.

3. Artistic individuals like the opportunity to express themselves in a free and unsystematic way by creating music, art, and/or writing. Correspondingly, an Artistic environment is one that is free and open and is encouraging of creativity and personal expression.

4. Social persons are interested in helping people through teaching, helping with personal or vocational problems, and/or providing personal services. Similarly, a Social environment is one that encourages people to be flexible and understanding of each other and where people can work with others through helping with personal or career problems, teaching others, and being socially responsible.

5. Enterprising individuals enjoy being with others and like to use verbal skills to sell, persuade, and/or lead. Correspondingly, an Enterprising environment is one in which people manage and persuade others to attain organizational or personal gains.

6. Conventional persons are typically dependable and have the ability to follow rules and orders in unambiguous situations. Likewise, a Conventional environment is one in which organization, planning, and ability to follow directions (often in an office environment) are of utmost importance.

Another assumption, coined Holland's calculus hypothesis, describes the relationship among the six person-environment types. The calculus hypothesis states that the six types can be arranged around a hexagon, with the types most similar to each other (e.g., Social and Enterprising) falling next to each other and those most dissimilar (e.g., Social and Realistic) falling directly across the hexagon from one another.

Holland's theory has been applied extensively in career counseling, most notably using the six-category typology to categorize individuals in the interpretation of interest inventories such as the Self-Directed Search (SDS) and the Strong Interest Inventory (SII). In addition, Holland's typology has been used to classify occupational information and college majors in an attempt to facilitate individuals' success in making occupational decisions.

### 1.4. Summary of Theories

The theory of work adjustment, Super's life span theory, and Holland's theory of vocational personality types all have received a great deal of attention and empirical support over the years. Research on the theory of work adjustment has generated many promising findings; however, research specifically using constructs from this theory has proceeded at a slow pace. Super's life span theory is intuitively appealing and offers assistance in understanding the richness of an individual's career and life. Nevertheless, it remains to be seen whether it will maintain its position of preeminence

in the future. Finally, Holland's theory of vocational personality types stands as the most influential of the existing theories in both research and practice, and research supporting Holland's theory is very robust.

## 2. CAREER COUNSELING FOR OCCUPATIONAL CHOICE

Career theories suggest that people make their occupational decisions in an orderly and systematic fashion. Yet research shows that for most people, the path often is serpentine. On the one hand, barriers to occupational entry, such as stereotyping, prejudice, and family, economic, and educational constraints, thwart people's opportunity to follow their interests. On the other hand, roughly 60% of adults report that serendipity plays a role in helping them to make good career decisions. Career counseling is a process, facilitated by a counselor, designed to help people find their way through the decision-making process. The ultimate goal is to help people obtain satisfaction and effectiveness in their jobs, occupations, and careers.

Career counseling can be structured in several different delivery modes. In some cases, the intervention is one-on-one between the client and the counselor. In other instances, the intervention may be conducted for groups and may resemble an educational experience with assignments and "in-class" discussions and information sharing. Career counseling also can occur in a variety of settings. For example, it is not unusual for college students to seek career counseling to help them decide on majors or to help them focus their job searches at the time of graduation. Many high schools also offer students the opportunity to explore future careers either with their school counselors or in life studies courses. Some employers offer career exploration opportunities within the work setting or arrange referrals for clients to work with career counselors. In addition, community organizations (e.g., the YWCA/YMCA) and adult education centers offer career counseling. Many people also seek out career counselors (now sometimes called career coaches) who are in private practice or work for consulting firms.

The desire to learn more about oneself and to ponder career questions is not the privilege of only the young. Many adults turn to career counseling for a variety of reasons. Some people are genuinely unhappy in their jobs and seek career counseling with the goal of finding better matches for their interests,

personalities, and values. Some people are bored and believe that their abilities are underused or that they have reached career plateaus. Others, in spite of their achievements, may question whether they are making a difference. Still others believe that they drifted into their jobs and never made deliberate choices. Not all people will discover, even with the help of career counseling, that they have a passion for work. For some people, work is necessary simply to provide for their basic needs, and other aspects of their lives will give them more pleasure and satisfaction than will their work. However, career counseling can help even those not driven to career achievement to maximize the match between their personal characteristics and those of the jobs they choose.

Regardless of the format, setting, life stage, or goals and purposes, most career counseling approaches involve psychological assessment. The ideal situation allows assessment of interests, abilities, personality, values, and needs. However, time and financial constraints often limit testing to the administration of a vocational interest inventory such as the SII, the Campbell Interest and Skill Survey (CISS), the SDS, the Career Assessment Inventory (CAI), or the Sixteen Personality Factor Questionnaire (16PF) (Table 1).

The University of Minnesota's Vocational Assessment Clinic (VAC) offers one model for combining comprehensive assessment with career planning services during a time-limited intervention: intake, assessment, interpretation of inventories, and integration with plan of action. In addition to taking the SII to assess vocational interests, clients who use the VAC services take the Minnesota Ability Test Battery (MATB), the California Psychological Inventory (CPI), and the Minnesota Importance Questionnaire (MIQ). During the interpretive sessions, counselors use the test results to help structure career exploration, to stimulate new possibilities and directions, and generally to help clients develop an integrated whole of their self-knowledge. Interest inventories can provide clients with a wealth of information, and the number of possibilities to explore often can be overwhelming. Thus, additional information from the assessment of abilities, personality, values, and needs can be useful for focusing their exploration and for providing additional information to determine the extent to which their characteristics match those of various jobs and occupations. After exploring the assessment results, the counseling process moves to developing a plan of action, often in the form of outside assignments, for gathering more information to inform the career decisions.

TABLE 1  
Comparison of Interest Inventories

<i>Vocational interest inventory</i>	<i>Focus</i>	<i>Usage setting(s)</i>	<i>Number of items</i>	<i>Scales</i>
<b>Strong Interest Inventory</b>	Professional populations Separate sex norms	High school/College	317	6 General Themes 25 Basic Interest 211 Occupational 4 Personal Style
<b>Self-Directed Search</b>	Student populations Self-administered Self-scored	High school/College	228	Daydreams Activities Competencies Occupations Self-Estimates
<b>Campbell Interest and Skill Survey</b>	Professional populations Combined sex norms Addresses self-estimated skills	Business	320	7 Orientation 29 Basic Interest 60 Occupational 2 Special
<b>Career Assessment Inventory</b>	Nonprofessional populations Combined sex norms	High school/College	370	6 RIASEC Themes 25 Basic Interest 111 Occupational 4 Nonoccupational
<b>Sixteen Personality Factor Questionnaire</b>	Professional populations Addresses personality	Business	185	16 Primary Factors 5 Global Factors 3 Response Styles

### 3. CONCLUSION

Research has been done to determine the effectiveness of career counseling. The results across several studies indicate that an average client has an outcome that is better than the outcomes of 80% of untreated control groups. Research has shown that congruence between a person's interests, abilities, personality, and values and the work environment—a desirable outcome of career counseling—leads to success and performance on the job and in school, life and work satisfaction, persistence on the job (i.e., tenure), and positive self-image and self-esteem. Conversely, poor correspondence (i.e., person–environment incongruence) is related to burnout, increased levels of anxiety, and somatic complaints. The implication of these outcome results is that whether people view their work as a job, a career, or a calling, career counseling can be a powerful aid in making occupational choices that lead to well-being.

#### See Also the Following Articles

Career Counseling ■ Holland's Theory (Vocational Personality Types) ■ Indecision, Vocational ■ Job Analysis, Design, and Evaluation ■ Job Search ■ Vocational Interests ■ Vocational Psychology, Overview

#### Further Reading

Arvey, R. D., Harpaz, I., & Liao, H. (in press). Work centrality and post-award work behavior of lottery winners. *Journal of Psychology*.  
 Betsworth, D., & Hansen, J. C. (1996). The categorization of serendipitous career development events. *Journal of Career Assessment*, 4, 91–98.  
 Brown, D., & Brooks, L. (Eds.). (1996). *Career choice and development* (3rd ed.). San Francisco: Jossey-Bass.

Campbell, D. P., Hyne, S. A., & Nilsen, D. L. (1992). *Manual for the Campbell Interest and Skill Survey*. Minneapolis, MN: National Computer Systems.  
 Cattell, R. B., Cattell, A. K., & Cattell, H. E. (1993). *Sixteen Personality Factor Questionnaire* (5th ed.). Champaign, IL: Institute for Personality and Ability Testing.  
 Dawis, R. V., & Weiss, D. (1994). *The Minnesota Ability Test Battery technical manual*. Minneapolis: University of Minnesota, Vocational Psychology Research.  
 Gough, H. (1995). *California Psychological Inventory manual*. Palo Alto, CA: Consulting Psychologists Press.  
 Hansen, J. C., & Dik, B. (2004). Measures of career interests. In M. Hersen, & J. Thomas (Eds.), *Handbook of psychological assessment*, Vol. 4: *Industrial/Organizational assessment*. New York: John Wiley.  
 Harmon, L., Hansen, J. C., Borgen, F., & Hammer, A. (1994). *Strong Interest Inventory applications and technical guide*. Stanford, CA: Stanford University Press.  
 Holland, J. L., Powell, A. B., & Fritzsche, B. A. (1994). *The Self-Directed Search professional user's guide*. Odessa, FL: Psychological Assessment Resources.  
 Judge, T. A., & Watanabe, S. (1993). Another look at job satisfaction–life satisfaction relationship. *Journal of Applied Psychology*, 78, 939–948.  
 Leong, F. T. L. (Ed.). (1995). *Career development and vocational behavior of racial and ethnic minorities*. Hillsdale, NJ: Lawrence Erlbaum.  
 Meir, E. I., Melamed, S., & Dinur, C. (1995). The benefits of congruence. *Career Development Quarterly*, 43, 257–266.  
 Parsons, F. (1909). *Choosing a vocation*. Boston: Houghton Mifflin.  
 Sharf, R. S. (2002). *Applying career development theory to counseling* (3rd ed.). Pacific Grove, CA: Brooks/Cole.  
 Strong, E. K., Jr. (1943). *Vocational interests of men and women*. Stanford, CA: Stanford University Press.  
 Ton, M., & Hansen, J. C. (2001). Using a person–environment fit framework to predict satisfaction and motivation in work and marital roles. *Journal of Career Assessment*, 9, 315–331.



# Occupational Psychology, Overview

Neil Anderson

University of Amsterdam, Amsterdam, The Netherlands

1. Introduction
2. Theoretical Perspectives
3. Key Topics in Occupational Psychology  
Further Reading

## GLOSSARY

**contracts** Employment and psychological contracts form the basis of the work relationship between employers and employees. Employment contracts tend to be in written form, stipulating legally binding terms and conditions of work. Psychological contracts are sets of shared understandings, obligations, and expectations of the work relationship.

**job stress and burnout** The negative psychological effects of workplace demands relative to the personal resources of the individual to cope with environmental stimuli. Stress is the interaction between job environment stimuli and individual responses in situations where the individual does not possess sufficient resources to cope. Burnout is characterized by feelings of emotional exhaustion, depersonalization, a lack of personal accomplishment, and a manifest inability to cope with even basic, ongoing demands in the work environment.

**occupational psychology** The scientific sub-discipline concerned with individual, work team, and organizational performance, psychological well-being, and human integration in the workplace.

**occupation and gender** Topics concerned with occupational choice, socialization, job performance, career choice, and

occupational choice as they relate to gender differences between males and females in the work environment.

**personal initiative and innovation** Topics concerned with idea generation, applied creativity, and innovation implementation in the workplace. Personal initiative involves work role changes that are self-initiated, that are proactive in orientation, and that demand persistence until change is implemented. Creativity is the generation of new ideas, whether eventually implemented or not. Innovation is the generation and implementation of new ideas, products, or processes at the individual, workgroup, or organizational levels of analysis.

**personnel selection** The attraction, recruitment, psychological assessment, and decision-making process involved in selecting newcomers for job vacancies in an organization in a valid, reliable, fair (i.e., legally defensible), and cost-effective manner.

**work team** A collective of two or more individuals that performs organizationally relevant tasks, shares one or more common goals, interacts, exhibits task interdependencies, manages boundaries, and is embedded in a broader organizational context.

Occupational psychology is a discipline related to industrial-organizational psychology that is concerned with individual, work team, and organizational performance, psychological well-being, and human integration in the workplace. Used especially in the United Kingdom, the term occupational psychology is used synonymously with that of industrial-organizational (I/O) psychology in the United States and Australia, and

with that of work and organizational (W/O) psychology in continental Europe. Due to its historic background in the United Kingdom, the subfield of occupational psychology developed somewhat specific foci, methods, and research concerns. Here, however, the term is used in its more general sense with acknowledgement to these historic precedents. Occupational psychology concerns work performance outcomes and health and well-being outcomes at three interrelated levels of analysis: (1) individual, (2) work group or team, and (3) organizational. Additionally, a few authors suggest a fourth level of analysis, that of interorganizational outcomes, although this has been more the concern of management scientists and organizational sociologists.

## 1. INTRODUCTION

### 1.1. Domain-Relevant Fields

The British Psychological Society (BPS) identifies eight major areas of knowledge and professional practice as constituting occupational psychology: (1) human-machine interaction, (2) design of environments and of work, (3) personnel selection and assessment, (4) performance appraisal and career development, (5) counseling and personal development, (6) training, (7) employee relations and motivation, and (8) organizational development. For individuals to become a chartered occupational psychologist (COP) under the BPS framework (essentially equivalent to licensing in other countries), they must have knowledge of all eight areas, supervised practice in five areas, and higher level independent practice over several years in at least two areas. Ongoing continuing professional development (CPD) workshops and seminars ensure that practitioners are kept updated on recent developments, but tensions between practice and research have risen under this framework. Researchers pursuing a Ph.D. and subsequent scientific career have found it difficult to gain the breadth of practical experience necessary to become COPs, as their work is usually specialized within one of the eight major areas. Recently, however, the Occupational Division of the BPS has issued tenders for project work to re-examine and possibly update this framework.

Arnold *et al.* stipulated 12 domains of occupational psychology, with some overlaps with the BPS professional society listing: (1) selection and assessment, (2) training, (3) performance appraisal, (4) organizational change and development, (5) ergonomics and equipment design, (6) career choice, development

and counseling, (7) interpersonal skills, (8) equal opportunities, (9) occupational safety and health, (10) work design, (11) attitude surveys, and (12) well-being and work. Again, the overlaps with the domain knowledge frameworks used to train I/O psychologists in the United States and Australia, and with European syllabi for W/O psychologists stipulated by the European Association of Work and Organizational Psychology (EAWOP) and the European Federation of Psychologists' Associations (EFPA), are self-evident. Occupational psychology, nevertheless, has several areas of interest and active research that are not widely represented in other countries. These include occupational choice and career guidance, gender studies, health and safety in the workplace, including accidents and errors, and historically, unemployment research, especially the negative psychological effects upon individuals of being jobless.

### 1.2. Brief History

Within the context of the United Kingdom, occupational psychology has a history of some 50–60 years. Industrial psychology, as it was called then, began during World Wars I and II as a highly focused effort to support the armed services and government departments. The inter-war years also saw the creation of the National Institute for Industrial Psychology (NIIP) and the Medical Research Council, which subsequently provided funding to several research programs in the United Kingdom, most notably the applied psychology unit at Cambridge University and the social and applied psychology unit at Sheffield University under Peter Warr (later, under Toby Wall). These all developed from war-time work and research studies into such topics as soldier selection and “shell shock” among soldier casualties, and demobilization studies into how best to return soldiers and commissioned officers to civilian and commercial careers in peacetime. This period also witnessed formative visits to Britain by U.S. and European applied psychologists, including, most influentially, Viteles in 1922.

The inter-war years brought with them troubles for occupational and applied psychologists. The economic depression and the heavy reliance of the NIIP upon industrial funding caused Elton Mayo to decline the position of its directorship upon C.S. Myers' retirement in 1938. His rationale for this decision is fascinating, describing the reliance for research funding upon industrial sources as being “made at the request of troubled firms” and resulting in “planlessness” of such

an unprogrammatic approach to research. Exactly the same ills bedeviled occupational psychology and I/O psychology internationally in later periods of economic depression, when research funding was exclusively reliant upon business sources and commercially applied studies. Yet the NIIP was highly influential; the titles of some of its published reports closely resembled the domain areas noted above, which were to become the main foci of occupational psychology (e.g., *Industrial Accidents, Vocational Guidance and Selection, Methods of Work, Psychological Effects of Noise, Labour Turnover, and Hours of Work*). The NIIP and, later, the Tavistock Clinic formed the foundation of much commercial and applied psychology in Britain, but additionally the three armed services (Army, Navy, and Air Force) committed substantial resources, during World War II and after, toward research on occupational issues, including the development of soldier selection tests and exercises. Work by F.C. Bartlett at the Cambridge Psychological Laboratory (later named the Applied Psychology Unit) also contributed to these themes, again with a distinguishing feature being the synthesis between research on selection and psychological health outcomes. Post-war applications included the development of assessment centers for the British government (the so-called Civil Service Selection Board, or CSSB, which had close parallels to U.S. assessment center work at that time), pilot selection and error, naval officer selection and training, and pilot ergonomic studies.

The 30-year period between 1960 and 1990 witnessed substantial growth in both the numbers of registered occupational psychologists and the range of activities they were engaged in. The establishment of the Careers Service to guide school leavers in labor market entry, research work conducted at the Tavistock Institute (the so-called human relations movement), quality of working life (QWL) studies, and industrial training studies in the 1960s and 1970s culminated in the Socialist-Labour government of the time creating the Steering Group on Job Satisfaction, with representatives from trades unions and employers, and chaired personally by the Minister of State. The economic boom years of the 1980s led to a concomitant growth in occupational psychology, with COP numbers rising substantially. Consultancy firm growth was also a marked feature of this period, with firms such as SHL, Oxford Psychologists Press, and the Psychological Corporation expanding rapidly. International diversification by consultancies led several large U.S. firms to establish European subsidiaries, and also led SHL to set up offices in the United States.

Growth spurred a step-shift in the numbers of accredited occupational psychology masters degree programs in the United Kingdom from a mere handful at the start of the 1980s to almost 30 at the end of the 1990s; in fact, the accreditation procedure itself was introduced by the BPS Training Committee during this period. Health and human resource management (HRM) issues continued to be high on the agenda for research and consultancy, stimulated still further at the end of the century by legal precedent that allowed employees to sue their organizations for claimed negative psychological effects caused by allegedly negligent HRM practices.

### 1.3. Defining Occupational Psychology

Modern-day use of the term occupational psychology has come to be synonymous in the United Kingdom with the terms I/O psychology in the United States and W/O psychology in continental European countries. Several handbooks and textbooks offer useful overview definitions, for example, “The application of psychological principles, theory, and research to the work setting” (Landy and Conte, 2004, p. 6), and “Work psychology is about people’s behavior, thoughts and emotions related to their work. It can be used to improve our understanding and management of people” (Arnold *et al.*, 2004, p. 10).

### 1.4. Applications of Occupational Psychology

As psychology applied to the world of work, occupational psychology has generated a whole host of practical applications. This science–practice blend is both the essence of the discipline and cause for some periodic tensions. Porteous noted 13 main areas of application: (1) individual differences, (2) personality and work, (3) work design and measurement, (4) performance appraisal, (5) behavior and work performance, (6) humanistic values, (7) cross-cultural values, (8) counseling workers, (9) stress, (10) leadership, (11) human–computer interaction, (12) health and safety, and (13) new work. Occupational psychologists are employed in industry, government, and consultancies if functioning as practitioners, and by universities or research institutes if working as research scientists.

## 2. THEORETICAL PERSPECTIVES

As with other areas of applied psychology, occupational psychology has adopted a range of theoretical perspectives and approaches as it has developed over the decades. Distinct from other areas of psychology, however, occupational psychology has also been heavily influenced by theories originating from the management sciences and from the discipline of organization theory and behavior. It is possible to identify at least four main theoretical perspectives that have had a discernible impact upon occupational psychology: (1) individual differences, (2) applied social psychology, (3) applied cognitive psychology, and (4) management science. Table I summarizes these approaches and historic influences.

### 2.1. Individual Differences

The individual differences approach to personality traits, cognitive abilities, values and attitudes to work,

and workplace behavior has a long and illustrious history in occupational psychology. Much of the seminal work in trait differences was undertaken using correlational and factor analytic statistical techniques (including the work of Eysenck, Cattell, and later by commercial personality test developers). The structure of human intelligence was similarly informed by early, seminal research mostly in laboratory experimental studies by scholars such as Galton and Vernon. Debate still continues as to whether tests of GMA or measures of specific abilities best predict subsequent job performance. Current meta-analytic research internationally converges on the finding that GMA tests are highly valid stand-alone predictors of job performance and that measures of specific abilities do not add significant incremental validity. Other individual differences have also been investigated, including attitudes toward work, leadership beliefs, work-related values, and interpersonal attitudes and style. All of these strands have tended to adopt a trait perspective, presuming that these

TABLE I  
Underlying Theoretical Approaches to Occupational Psychology

	<i>Individual differences</i>	<i>Applied social psychology</i>	<i>Cognitive psychology</i>	<i>Management theory</i>
Basic tenets	Individuals differ along stable, measurable dimensions that affect observable behaviors	Interpersonal and group processes affect group productivity and individual well-being	Individual cognitions, information processing capacities, and decision heuristics affect individual task performance	Organizing processes and leader behavior are critical determinants of organization success
Primary level(s) of analysis	Intra-individual Inter-individual	Inter-individual Intra-group Inter-group	Intra-individual	Organizational Inter-group
Illustrative topics of influence	Selection Personality at work Stress proneness	Group processes Attribution theory Innovation and creativity Conflict	Decision making Human-machine interface	Organization structure Authority relations Organization success
Key models and constructs	Five Factor model (FFM) of personality Type A, Type B General mental ability (GMA)	Group development models West's Four Factor model of team innovation	Information overload Heuristic biases Information Integration Theory (IIT)	Contingency theory HRM and human capital Organizational culture and climate

dimensions are stable, replicable over time, and causal of behavior across workplace situations, a perspective for which there has indeed been considerable research support.

## 2.2. Applied Social Psychology

As can be seen in Table I, the primary level of analysis for which applied social psychology research has been influential in occupational psychology is that of the workgroup or team in organizations. This has been a far more recent theoretical influence than individual differences research, however, and so there has been considerably less attention in the field paid to group-level issues than to individual-level issues. In more recent years, organizations internationally have moved more toward team-based structures of work organization, spurring considerable growth in research interest at this level of analysis. Much social psychological research into groups has been conducted in laboratory experimental conditions, leaving open to doubt whether their findings can be applied to real-life workplaces and business organizations. Recent occupational psychology research has been grounded upon earlier findings, models, and theories in applied social psychology but has examined phenomena in the workplace. Phenomena such as workgroup innovation, team performance and productivity, team decision-making among boards of directors, process losses in work teams, distributed cognition and decision-making, and conflict in workgroups have come to the fore. This research has irrevocably shifted the focus of occupational psychologists away from an exclusively individual level of analysis toward issues of interpersonal, group, and inter-group processes in organizations. Multilevel designs, that is, research measuring variables and effects across two or more levels of analysis (individual, workgroup, organizational), are becoming more common as a result.

## 2.3. Cognitive Psychology

Historic influences from cognitive psychology upon certain issues in occupational psychology are evident, as illustrated in Table I. Many of these influences concern individual-level issues of ergonomics, human-machine interface, and individual error in task functions. In certain countries (the United States and Germany, for instance), such applications were traditionally termed industrial psychology, with large-scale funding for often being provided by military research institutes or

in industries where error consequences are of particular concern (e.g., nuclear power generation).

## 2.4. Management Theory

Distinct from other areas of applied psychology, occupational and I/O psychology has been heavily influenced by more macro-analytic theories, approaches, and issues addressed generally in the management sciences. Table 1 suggests some key examples, but the field has been impacted by research from several sub-disciplines in management, principally organization behavior (OB), organization theory, HRM, and organization development (OD). Most of these approaches have adopted an organizational level of analysis in considering performance and processes across entire organizations or multiple subunits as part of a work organization. Thus, management research in occupational psychology has been highly pragmatic in orientation, almost always involving quantitative field studies or even qualitative case studies, and has examined practical problems and processes as they occur in work organizations. In essence, this fourth approach has considered the managerial implications of occupational psychology through major topics such as leadership, HRM processes and organizational performance, theories and models of organizational structure, organizational change and development, culture and climate as determinants of success, and so forth.

## 3. KEY TOPICS IN OCCUPATIONAL PSYCHOLOGY

This section of the encyclopedia is composed of seven separate articles covering key topics in occupational psychology. They are as follows:

1. Contracts
2. Job stress and burnout
3. Occupation and gender
4. Personal initiative and innovation
5. Personnel selection
6. Safety and risks
7. Work teams

These purposely concentrate on the individual and team levels of analysis, given the historic background and theoretical approaches to occupational psychology

delineated above. An overview of each of these topics is provided in the following sections.

### 3.1. Contracts

Employment contracts and psychological contracts form the basis of the employer–employee relationship. The exchange nature of what each gives relative to what is expected to be received from the other is crucial, and can be termed the “mutuality” of expectations and obligations. While employment contracts are usually written, formal, and legally enforceable, psychological contracts are sets of mutual expectations that may be differentially construed by individual employees and managers in the workplace.

Psychological contracts especially are changeable in nature, open to violation by either party, and may be open to negotiation and renegotiation over time. The relative power of each party to negotiate a favorable contract is therefore important. Perceived violations of the psychological contract by employees have been found to be frequent, but research into perceived violations on the employer’s side by underperforming employees is sparse.

### 3.2. Job Stress and Burnout

Stress and burnout at work represent major themes of research attention and concern within occupational and I/O psychology internationally. Stress is defined generally as the interaction between job environment stimuli and individual responses when the individual does not possess sufficient resources to cope, thus resulting in strain. So-called “interactional definitions” of stress emphasize the interaction between stimuli and individual responses as opposed to earlier, more limited, non-interactional conceptualizations of job stress. “Transactional definitions” imply that stress is a result of the transaction between the individual and his or her work environment.

Job stress and strain has been modeled in various ways, each guided by the definition of stress adopted by the researcher. Interactional and moderator models, which are more complex in their design and their inclusion of different types of variables, have become more prominent over recent years. A range of causal variables resulting in job stress has been identified, including work role demands, role conflict, role ambiguity, role overload, work complexity and multi-tasking, the home–work interface, the work–family balance, and financial and other non-work problems.

Wider causes include business travel demands, relocation, expatriate working, bullying and mobbing at work, harassment, job insecurity, and redundancy. Increasing attention has thus been devoted to stress coping mechanisms and individual differences in coping with stressful events and work environments. Organizations, concerned by the potential costs of absence caused by stress and burnout, have invested in Employee Assistance Programs (EAPs), which are designed to support individual employees in a confidential and constructive manner. “Burnout” has been defined as “a special form of strain” characterized by feelings of emotional exhaustion, depersonalization, a lack of personal accomplishment, and a manifest inability to cope with even basic, ongoing demands in the work environment.

### 3.3. Occupation and Gender

Gender differences in career prospects, earnings, seniority of job roles achieved, and occupational choice have concerned occupational psychologists for many years. Women have been found to earn less than men, be concentrated into certain industries, and be more likely to work in certain lower authority job roles, and may have imposed upon them a “glass ceiling” of how high up the organization they can progress in terms of promotion to higher grade managerial job roles. A greater proportion of women work in part-time jobs and in service industry job roles, largely due, it has been suggested, to family commitments and traditional stereotypes of the mothering role and the housewife. These differences, still present in several developed economies cross-nationally today, have been fundamentally influenced by historic trends and precedents.

The socialization of boys and girls into potential work roles in organizations is also thought to have an impact. Gendered job roles, that is, traditionally male- or female-dominated occupations, may be held up as desirable during child development. For boys these may include military roles, emergency services, senior management, technical and scientific careers, and executive functions. For girls, in addition to family being emphasized, helping roles such as nursing, the caring professions, and service roles may be held up as stereotypical examples. Occupational values and choice may thus differ between the genders, and this spills over into differences in leadership role choice and style between males and females. Anti-discriminatory legislation in several countries that makes it unlawful

to discriminate on the grounds of gender does not appear to have eliminated these problems. Organizational efforts to enhance and manage diversity have shown positive outcomes, but many inequalities remain, and further research and practical intervention is called for.

### 3.4. Personal Initiative and Innovation

As organizations have moved away from job specialization and Taylorian scientific management as methods of work organization, the importance of creativity, innovation, and personal initiative in the workplace has grown. The need for product or service flexibility, adaptability to changing business environments, matrix- and team-based organization structures, and more flexible forms of working have all heightened the importance of work role creativity and team-level innovation. Creativity has been defined as the generation of new ideas, whereas innovation involves both idea generation and solution implementation in the work setting. Personal initiative is a related construct, involving self-starting, proactivity, and persistence, the outcome of which may be deemed to be innovative by observers.

Innovations develop over time not in a neat, linear fashion or through discrete step-by-step phases, but generally as iterative, complex cycles of development–regression–further development. Factors associated with innovativeness at the three interrelated levels of analysis in occupational psychology have been identified. At the individual level, these include personality, motivation, cognitive ability, knowledge, and job characteristics. At the work group level, these incorporate team dynamics and structure, climate, processes, and leadership style. Organizational level variables include structure, climate and culture, size, resources, work processes, and physical environment. Linear, curvilinear, threshold effect, and indirect relationships between these variables and innovation have been suggested or found. More recently, multilevel designs, possibly using more robust meta-analytic techniques, have been called for to clarify this mass of individual study findings.

Applied occupational psychology findings have countered a predominant but fallacious pro-innovation bias common in management texts that suggests that all innovations are positive and that the more innovation, the better for organizational performance. Rather,

applied psychology research suggests that some types of innovations can be stressful for the initiator, problematic for the workgroup to assimilate, or conflict-provoking in the wider organization where other departments or subunits are affected.

### 3.5. Personnel Selection

Similar to stress and burnout, personnel selection has long been a major area of research and professional practice in occupational psychology. Based historically in individual differences theories and tenets, selection more recently has expanded its concentration of levels of analysis toward the team level and issues of person–work group fit. Applicant as well as organizational decision-making has also received considerably more research attention, but the traditional model of personnel selection is based upon fitting the person to the job through the administration of various prescreening and candidate assessment predictor methods.

Seminal usage and development of meta-analysis methods in selection have generated definitive research findings over the criterion-related validity of different predictors. The best stand-alone predictor of subsequent job performance across all complexities of jobs, and in different countries worldwide, is general mental ability. Other methods add to the prediction of the wider criterion space, including work sample tests, some dimensions of the “big five” personality framework, and integrity tests. Historically much maligned, interviews, particularly structured interviews, have been found in meta-analytical research to have far better predictive validities than previously thought. Other methods that demonstrate near-zero predictive validity include age, graphology, and other so-called alternative methods.

The dilemma that the most valid and reliable predictors are not necessarily the most popular with practicing recruiters continues to concern researchers in personnel selection. Other important research has investigated fairness issues (adverse impact or unfair discrimination) in selection in different countries. Driven by differing national legal frameworks, the stringent minority rights legislation in the United States has led to research there being far more developed than in other countries. The minority groups protected under legislation differ across countries, but usually depend upon the pattern and origins of immigrant workers to a host country.

Increasing research attention is being devoted to applicant reactions, so-called “banding” of minority

group scores to mitigate the effects of underprediction, how best to combine the information gleaned from different predictors at different stages in a selection process, expatriate selection, selection for team-working, and new technology in the recruitment process.

### 3.6. Safety and Risks

Fay and Tissington proposed that unsafe behavior and errors lead to higher risks and therefore may cause accidents to happen on the work floor. Unsafe behavior is defined as “behaviors that are not in accordance with safety rules and procedures.” Error can be seen as “an intentional action which does not result in the attainment of the goal [whereby] prevention of the error must have been under the control of the person(s) [involved].” Research has adopted two different approaches toward safety and risks: the person-centered approach, focusing on the role of individual difference variables, and the engineering control solutions approach, seeking to design safety into the work system by acknowledging human factor variables. Three subsequent levels of prevention are central to this approach: the design of an error-proof system, engineering safeguards, and finally, installation of warning systems. Interventions striving to enhance safety aim at changing characteristics of the workforce, changing the work itself, and/or altering elements of the work context.

### 3.7. Work Teams

A work team is defined by Kozlowski and Bell as “a collective of two or more individuals that performs organizationally relevant tasks, shares one or more common goals, interacts, exhibits task interdependencies, manages boundaries, and is embedded in a broader organizational context.” As organizations have become flatter and delayed, team-working has become the predominant form of work organization in many post-industrial economies. In the past, Taylorism and the specialization of work dominated many large corporations’ mode of work organization, but organizations have become increasingly dependent upon team-level co-operative efforts for their success. Team effectiveness models of an input–process–output nature guided classical research into workgroup effectiveness, yet have become increasingly complicated by researchers’ attention to moderator and mediator variables and relationships.

Contingency models of team composition and structure emphasize the need to base team formation upon environmental and task demands, including such design issues as team size and diversity of members (i.e., homogeneity versus heterogeneity). Certain outcomes warrant a more diverse composition (e.g., idea generation, adaptability), whereas others suggest a more homogenous member composition (e.g., routine task performance, innovation implementation). Models of team development have also moved from a more prescriptive-descriptive stance to becoming more situation-dependent (i.e., contingent) as research has become more sophisticated. Fixed stage models have been supplanted by more recent punctuated equilibrium models.

Emerging research centers on cognitive and affective models of team motivation, the efficacy of team-building interventions, task interdependence and performance including knowledge structures, and highly pragmatically important issues of team creation, team-building, and performance improvement attempts. Little research appears to support the commercially driven consultancy market in this regard despite repeated calls by researchers for such evidence. Future research in occupational psychology into team-working in organizations is likely to reflect its increasing importance for organizational success, and the growth in research in this area is thus likely to continue apace.

### See Also the Following Articles

Contracts ■ Industrial/Organizational Psychology, Overview ■ Job Stress ■ Job Stress and Burnout ■ Occupation and Gender ■ Personal Initiative and Innovation ■ Person–Environment Fit ■ Personnel Selection ■ Safety and Risks: Errors and Accidents in Different Occupations ■ Vocational Interests ■ Work Environments ■ Work Motivation ■ Work Role, Values Sought in the ■ Work Safety ■ Work Teams

### Further Reading

- Anderson, N., & Prutton, K. (1993). Occupational psychology in business: Strategic resource or purveyor of tests? *The Occupational Psychologist*, 20, 3–10.
- Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner-researcher divide in Industrial, Work, and Organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, 74, 391–411.
- Anderson, N., De Dreu, C. K. W., & Nijstad, B. A. (2004). The routinization of innovation research: A constructively

- critical review of the state-of-the-science. *Journal of Organizational Behavior*, 25, 147–173.
- Arnold, J., Cooper, C. L., & Robertson, I. T. (1998). *Work Psychology: Understanding Human Behaviour in the Workplace*. London: Financial Times Publishing.
- Arnold, J., Silvester, J., Patterson, F., Robertson, I. T., Cooper, C. L., & Burnes, B. (2004). *Work Psychology*, 4th ed. Harlow, UK: Pearson Education.
- Cleveland, J. N., Murphy, K. R., & Stockdale, M. (2000). *Women and Men in Organizations*. Hillsdale, NJ: Lawrence Erlbaum.
- Cooper, C. L., Dewe, P. J., & O'Driscoll, M. P. (2001). *Organizational Stress: A Review and Critique of Theory, Research, and Applications*. Thousand Oaks, CA: Sage.
- Hodgkinson, G. P., Herriot, P., & Anderson, N. (2001). Re-aligning the stakeholders in management research: Lessons from industrial, work and organizational psychology. *British Journal of Management*, 12, 41–48.
- Klein, K. J., & Kozlowski, S. W. J. (Eds.). (2000). *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions and New Directions*. San Francisco, CA: Jossey Bass.
- Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. Klimoski (Eds.), *Handbook of Psychology: Industrial and Organizational Psychology* (Vol. 12, pp. 333–375). Chichester: Wiley.
- Landy, F. J., & Conte, J. M. (2004). *Work in the 21st Century: An Introduction to Industrial and Organizational Psychology*. New York: McGraw-Hill.
- Myers, C. S. (1926). *Industrial Psychology in Great Britain*. London: Jonathon Cape.
- Myers, C. S. (1936). Autobiography. In C. Murchison (Ed.), *A History of Psychology in Autobiography*. Worcester, MA: Clark University Press.
- Porteous, M. (1997). *Occupational Psychology*. London: Prentice Hall.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European meta-analysis. *Personnel Psychology*, 56, 573–605.
- Schalk, R., & Rousseau, D. M. (2001). Psychological contracts in employment. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology* (Vol. 2, pp. 133–142). London: Sage.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Shimmin, S., & Wallis, D. (1994). *Fifty Years of Occupational Psychology in Britain*. Leicester, UK: British Psychological Society.
- Sinangil, H. K., & Avallone, F. (2001). Organizational development and change. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology* (Vol. 2, pp. 332–345). London: Sage.
- Vernon, P. E., & Parry, J. B. (1949). *Personnel Selection in the British Forces*. London: University Press.
- West, M. A. (2004). *Effective Teamwork*, 2nd ed. Oxford: British Psychological Society/Blackwell.



# Occupation and Gender

Eva Bamberg

University of Hamburg, Hamburg, Germany

1. The Gender-Specific Employment Market
  2. History of I/O Psychology
  3. Concepts and Results
  4. Explanations
  5. Career Advancement for Women and Managing Diversity
  6. Practical and Scientific Perspectives
- Further Reading

## GLOSSARY

*diversity management* Planned steps to increase the diversity of the workforce and to integrate minorities in organizations.

*gender* Social construction, based on sex.

*gender-related interventions* Interventions designed to equalize men and women with regard to the workplace.

*gender-specific division of labor* Different types of work for men and women.

*gender stereotypes* Generalized beliefs about men and women.

The gender-specific division of labor in society implies that men and women practice different occupations, and consequently differ with regard to income, status, power, influence, and career perspectives. This gender-specific division of labor is supported and kept strong by processes of socialization and assignments. In the past few years, a number of interventions have been designed to advance the developmental potential of both sexes.

## 1. THE GENDER-SPECIFIC EMPLOYMENT MARKET

In most industrialized nations, the percentage of working women has risen continuously during the past few years, whereas the percentage of working men has remained almost the same. Approximately 80% of all men and 60% of all women practice an occupation. Although women and men share an equal basis of knowledge, acquired in school, career perspectives differentiate in horizontal as well as vertical aspects of occupation. The horizontal aspects of occupation concern the type of occupation and trade; the vertical aspects of occupation concern inner structural hierarchies. More men than women are suitably employed after educational training. Women more often practice work characterized by low demands and low decision latitude. Employment uncertainty affects women more than men; women more often work part-time.

Even when men and women hold similar jobs, men are provided with higher salaries. Female managers earn less than their male counterparts. Women in Europe earn approximately 15% less than men, and women in the United States earn approximately 25% less than men. These differences in salaries affect not only the sexes but also families. Female-headed families are one of the most economically vulnerable segments of the population.

Men and women differ with regard to their career perspectives. In Europe, for example, men are twice as likely to be in managerial positions. Only 3% of top

management positions are occupied by women. Worldwide, men dominate executive functions.

Gender-specific working conditions are found within the psychology community as well. Male psychologists are more likely to be provided with higher salaries, secured positions, and more executive functions than their female colleagues. Another phenomenon of the employment market can be observed for the psychology job market as well: When an organization or profession becomes more “feminine”—female employees increasingly predominate—the public value and influence of its members decrease considerably.

The gender-specific division of labor is not limited to paid work: Beginning in early childhood, girls perform more household chores than boys. Numerous empirical studies show that women do chores more than men. These findings are not limited to housewives and part-time working women; they also hold for full-time working women.

Work tasks are gender specific as well. Twice as many women as men do volunteer work in the social field, and nearly three times as many men volunteer in the area of politics. Men nevertheless mainly hold executive functions and prominent posts in the volunteer segment.

Work is divided in a gender-specific way. Type of occupation and gender are therefore important. Industrial/organizational (I/O) psychology focuses on premises, conditions, and consequences of work and division of labor within society. Gender and occupation should also be integrated.

## 2. HISTORY OF I/O PSYCHOLOGY

In the early years of I/O psychology, researchers reported sex-related characteristics of work and noted gender differences. The following is an example:

### Early Observations of Gender and Occupational Behavior

“Sex differences can be directly identified methodically, based on any profession tasking both sexes in a parallel manner. Female and male tailors are differing widely concerning the manual aspects of their occupation. A man treats and processes fabrics even on a subjective point of view, applying strength, energy, and a dominance that should be called systematic. A tailoress (female tailor) treats her fabrics somehow almost intuitional, playful, delicate, without any systematic, totally according to her own individual fittings.” (p. 922)

From Griese, F. (1928). *Psychologie der Arbeitshand*. Berlin: Urban & Schwarzenberg. Translated by E. B.

Differences between the sexes based on physiological aspects (e.g., muscular strength and operation of the cardiopulmonary system) and psychological aspects (e.g., intelligence, motivation, and commitment) were stated; specific sex-related aspects were noted.

The thesis of monotony resistance of women, developed in approximately 1920, stated that women engaged in repetitive work feel less bored and less monotony than men. Because repetitive work prevents struggling with constant decisions and allows daydreaming, women, who seek no satisfaction in their jobs, regard this kind of work to be convenient and desirable.

These statements are still made even today, but they have never been scientifically proven. A number of early studies by I/O psychologists integrated women (e.g., the Hawthorne study). However, no scientific discussion concerning the sexes of the test subjects and related consequences for the studies took place (and still do not take place). Other studies involved women only marginally in their samples. This has to be kept in mind especially regarding studies examining the consequences of working conditions and tasks on performance, health, and well-being. Until the 1980s, the issue of occupation and gender was almost totally ignored. Reacting to criticism about the neglect of women within industrial psychology research, a number of studies were performed that increasingly considered women and related results to gender-related topics. Kohn and Schooler, for instance, in their widely quoted study on occupation and personality, questioned men first and later their female partners, applying a longitudinal design.

Since the 1980s, an increasing number of studies have focused on women, resulting in increased research on gender-specific differences. Applying this new framework, research has focused on identifying and evaluating gender-specific differences (e.g., the degree of distinction in gender-specific values and interests and the degree of distinction in leadership behavior between men and women) and evaluating the differences in gender-related aspects within certain theoretical models (e.g., the connection between working conditions and personality).

## 3. CONCEPTS AND RESULTS

Common knowledge holds that gender-specific division of work is based on the differing work potential

of women and men. Applying this point of view, boys and men are mainly focused on career aspects, whereas women tend toward family aspects. Men are mainly interested in scientific/technical professions, which provide better career opportunities, whereas women focus on caring and nursing professions; men show materialistic interests, and women show social interests. Men are technically oriented and they impose their will; therefore, men succeed easier and are especially qualified for prominent positions of executive functions. Women are more cooperative. The following concept is regularly pointed out: Men and women differ according to the values related to the importance of work; therefore, consequences differ as well. For instance, for women unemployment is stated to be a minor problem because they have alternative roles.

How does common knowledge compare with scientific results? The following sections discuss three concepts of I/O psychology, focusing on occupation and gender.

### 3.1. Occupational Values and Goals

A number of studies have focused on whether occupational values and goals of men and women differ. Intrinsic values (social relations and job characteristics) and extrinsic values (salary, promotions, and occupational safety) are often separated. With regard to male and female frameworks of occupational values, there are only minor differences. Accordingly, intrinsic values are more important to women, and extrinsic values are more important to men.

The most prominent differences occur for young, married employees. The older the workers, the less prominent the differences become. This is due to the fact that men's focus on extrinsic values diminishes and their focus on intrinsic values increases with age.

A very striking connection was made by research examining the career orientation of students. The main sample showed only few differences between men and women. Focusing on the students' progression, the intrinsic career motivation of women is less prominent at the end of their scientific qualification period than at the beginning, whereas that of men is more prominent at the end of their scientific qualification period than at the beginning. Accordingly, there is a more prominent orientation on career aspects for men at the end of their scientific qualification period than for women.

Although the previously discussed results on occupational values show only few gender-related differ-

ences, these are found to be prominent by examining specific occupational plans and goals.

For many decades, research has elicited women's plans to reduce or discontinue occupational practice in favor of household matters. Correspondingly, especially women include housework in their future plans. This can be explained by the different social roles of men and women. Congruent to the gender-specific division of labor, men consider their future occupational practices to involve paid work. Their traditional responsibility is securing the family income. This injects special importance in career-related aspects, especially if entering the occupational market while simultaneously starting a family. Women view their working lives as involving paid employment as well as the family. Women have multiple roles. They are less likely to accept negative aspects or disadvantages of paid work than are men.

### 3.2. Leadership

Based on the previously mentioned gender-specific difference in employment within the management segment, leadership aspects have to be considered especially important. Here, the focus is on the occupational history and promotion of managers and on leadership behavior.

Studies have shown no significant difference between occupational values and motives of male and female managers. Compared to their male counterparts, female managers are more often single and have fewer offspring. Their occupational history can be compared to a "chimney," leaving space only for the paid work.

Mainly qualitative research indicates that men and women differ with regard to their identification of factors that enabled them to enter their career and gain promotion. Women refer more often to prejudices and systemic barriers—hindrances. Women in prominent positions with executive functions are more closely watched and they cannot use internal networks as effectively as men.

The prototype of the successful manager matches mostly male gender stereotypes and less often female gender stereotypes. The often validated phenomenon of think manager/think male is closely related to gender stereotypes, being the basis of judgments concerning leadership behavior. Despite major correspondences, these cause the difference in judgment of behavior of male and female managers. The position of female manager is considered as a token status. Women token workers experience a high degree

of visibility, isolation, and gender-stereotyped roles, causing discrimination.

Gender-related differences in leadership behavior have been the focus of many research projects in recent years. A more participative or personal-oriented style of leadership of women was noted especially by qualitative research. However, applied meta-analysis was unable to verify these findings. Research on leadership behavior only produced gender-related differences by applying laboratory-based methodologies. Field studies indicated only marginal gender-related differences. This could be due to the occupational socialization of the manager, corresponding with the development of specific skills. These skills build the central foundation of executive functions and are developed independently of being male or female.

### **3.3. Occupation and Health Issues**

The average life expectancy of women is approximately 7 years longer than that of men. At first, this is not remarkable. Compared to women, men engage in behavior with significantly higher risks: Men smoke and consume alcohol more often than women, they abuse drugs more often, they less often have a healthy diet, they take advantage of health care programs less regularly, they are less sensitive toward signs of exhaustion, and they are involved in accidents at work more often. However, men are not more prone to sickness compared to women. Based on studies relying on self-reports, women suffer in a more intense way from psychological distress.

A number of explanations have been given for this finding. First is the more prominent emotional openness of women, which places women more at risk of psychological distress. This is more consistent with the female gender stereotype than the male, especially with regard to psychological distress and sickness. Another explanation is based on the differing vulnerability of men and women. The genetic, physiologic, and hormonal constitution of women is responsible for the inconsistency concerning stress, demonstrated by morbidity rates. The final explanation is the difference in working and living conditions between men and women, including stress related to paid work and housework.

The results provided by a meta-analysis on the significance of gender-related differences with regard to the relation of work and health are inconsistent with the second explanation presented previously. The second explanation states that there is a moderating function of gender within the relation of work and health. Due to the

contrariness of the presented findings related to working conditions and health, determining the most plausible explanation seems to be impossible. Such a determination would require significantly more information about gender-specific working and living conditions. Information on stress and resources is needed to gain deeper insight into working and living conditions. Information is needed that is unlimited to aspects of the practiced occupational matters alone. There is a lack of adaptable measuring instruments capable of analyzing the overall stress impact and resources.

## **4. EXPLANATIONS**

In the past, gender research was primarily based on biological models. This frames the previously mentioned statements about I/O psychology and the differing vulnerability of men and women. Biological models are based on the assumption that genetic, physical, and hormonal factors establish (gender-specific) behavior. Current approaches are based on the assumption that in addition to biological aspects (sex), social aspects (gender) also have to be considered. According to socialization models, the difference between men and women and gender identity is established throughout human socialization based on processes of learning, such as model learning, imitation, differential enhancement, and gaining cognitive concepts.

Social structural-cultural models of gender highlight the specific division of labor applied by certain institutions (e.g., schools, businesses, and the military) in order to maintain their specific function. In addition to division of labor, especially gender-related division of labor, a distribution of status and power takes place. This sets the foundation for social structures, norms, rules, and behavioral standards accepted and executed by the participants. Current concepts of occupation and gender integrate the previously mentioned approaches.

Biologically determined sex implies socially determined gender: Men and women perform different tasks and different types of activities. This has effects for all types of work—housework, freelance work, voluntary work, and paid work—forming a horizontal and vertical division in the employment market that affects not only types of activities but also power and status.

By working within specific areas of work, men and women adapt different attitudes, actions, competencies, and abilities. Gender-specific attitudes or competencies are found whenever the types of activities differ between men and women.

Gender-typical conditions, the types of activities of men and women, affect gender-specific differences and gender-related characteristics and processes. This influences the common knowledge base about men and women and their occupational attributes considered typically male or female.

Regarding psychology, research has focused on gender stereotypes. Gender stereotypes are structured assumptions creating a connection between personality traits and social categorizations—constructions of manliness/womanliness.

As shown by research on gender stereotypes, adaptive-instrumental attribution (e.g., achievement oriented, logically oriented, and independent) particularly describes men, and integrative-expressive attribution (e.g., sensitive, warm-hearted, and emotional) particularly describes women. Gender stereotypes include more than attributes; they also include role-consistent behavior, occupational praxes, etc. Gender stereotypes contain statements about the basic way men and women are (sex trait stereotypes) and how they have to be (sex role stereotypes).

The gender-specific division of labor affects the way in which men and women are described, and it affects gender stereotypes (see *Three Schoolboy Essays*). Research has shown that gender stereotypes are more prominent in people whose professions can be classified as traditionally male/female.

#### *Three Schoolboy Essays*

1934: "I have often wondered what profession I would choose if I was a girl and was at the end of my school years. Since girls have not got so many options. Most of them become sales assistants. Since girls are frail other professions cannot be chosen for them" (p. 111).

1946: "If I was a girl, I would choose becoming a sales assistant for a bakery. I would like to do so because there would always be food for me in a bakery. Nowadays, girls are looking for apprenticeships at a grocery. The reason for this being, food can really be scarce" (p. 112).

1986: "Girls have fewer job perspectives than boys because there are mainly technical jobs left. Being a girl, I would take a job dealing with plants because I think that girls know more about plants. The job of a hairdresser is good for girls as well because they have more delicate hands than boys do. Myself being a girl, I would become a florist because boys are not tender enough to seed and care for plants" (p. 115).

From Bamberg, E. (1996). *Wenn ich ein Junge wär*. Göttingen, Germany: Hogrefe-Verlag. Translated by E. B.

Gender-related division of labor affects not only the way men and women are described but also the way

they describe themselves. A number of studies indicate that girls and women state they have expressive attributes, whereas boys and men state they have instrumental attributes. This can be mostly found in cases of similar experiences. In other words, whenever the job is considered typically male or female, generalized attributes of the job are adapted in the framework of the personal self. Research has shown that jobs considered typically female support the self-acquisition of female attributes, whereas jobs considered typically male support the self-acquisition of male attributes.

Based on the concepts presented here, a correlation between gender and occupation can be shown by applying two different perspectives. The individual or person-centered perspective focuses on acquisition of competencies, attitudes, and self-descriptions within socialization, including occupational socialization. Sex-related division of labor leads to the development of gender-specific attributes, which again support division of labor. Women who describe themselves as having expressive attributes will choose professions in which they can realize these expressive attributes.

The situational or context-related perspective focuses on valuation, furtherance, attachment, and hindrances set up by the actors and their frame of action. This again supports the gender-specific division of labor. Women working in typically male professions are often exposed to hindrances or systemic barriers that are not found in professions considered to be typically female.

## 5. CAREER ADVANCEMENT FOR WOMEN AND MANAGING DIVERSITY

In order to optimize manpower and minimize employee turnover, in the 1970s a number of businesses started to employ plans for career advancement for women. These plans employed a number of different starting points or target areas:

1. Ensuring personal safety: This includes guidelines to prevent discrimination and the creation of special safety areas for women, such as break rooms or reserved parking places.
2. Reduction of horizontal gender-specific segregation: This includes plans aimed at the facilitation of access to specific occupations, such as scientific/technical-oriented occupations.

3. Interventions focusing on changing vertical segregation: This includes mainly incentives aimed at increasing the percentage of female managers (e.g., mentoring).

4. Incentives aimed at increasing the compatibility between occupation and family life: This includes guidelines concerning work hours, maternity leave (the German *erziehungsurlaub* is an extended version of maternity leave for educational purposes), and implementation of part-time work.

As demonstrated by the latter target area, career advancement for women and the Equal Rights Amendment as well as family-support programs share overlapping components.

As mentioned previously, occupation and gender can be portrayed by two different perspectives: the individual or person-centered perspective and the condition-related perspective. Gender-related interventions are aimed at the implementation of individually oriented ways of development and at the creation of conditions necessary to advance development potential for both sexes. Therefore, both perspectives must be considered.

Nearly all interventions and measures for personal and organizational development can be implemented for gender-related interventions and measures as well. These include placing women in prominent positions, considering gender-related issues with regard to employment matters and recruiting, designing career advancement paths for women (e.g., offering training programs, mentoring models, and coaching), and implementing flexibility concerning workplace and working hours. Recently, the influence of gender-related measures on public opinion and ranking of organizations has been observed. Campaigns such as the family-friendly index, family audit, or the election of "house husband" of the year are part of this trend. In the United States, and in Europe as well, career advancement for women and the implementation of the Equal Rights Amendment are being replaced by the concept of managing diversity. The most prominent difference is that the concept of managing diversity does not just take women into consideration and does not focus only on forms of discrimination. The new approach centers on advancement of the developmental potential of all social groups within an organization (e.g., same-sex groups, age groups, and ethnic groups). Regarding multicultural organizations, employees are integrated not only structurally on all positional and hierarchical layers but also on the informal layer of networks. This decreases tension and discrimination based on employees belonging

to minority groups. Staff politics is based on the premise of equal opportunity. Norms, criteria, procedures, and methods are tested with regard to discrimination and equality and changed accordingly if necessary.

## 6. PRACTICAL AND SCIENTIFIC PERSPECTIVES

The central concept of occupation and gender is based on the type of work. Men and women acquire attitudes, competencies, and abilities based on their work activities. The processing of gender stereotypes and gender-typical self-adjustment depends on the type of work. Consequently, there are implications for the areas of practice and science. Practical measures aimed at changing attitudes or increasing competencies have to begin with job design. Classical interventions within organizations and firms, such as job design or organizational development, did not focus on gender-related issues explicitly. This implies the nonobservance of women-specific working conditions. Most measures concerning job design, for instance, are based on jobs held by male employees and not on jobs mostly held by women.

Therefore, future approaches should not focus on equal rights politics and management diversity alone. The development of methods and measures designed to implement conditions considered typically male or female, with regard to job design and personal and organizational development, is needed. Gender mainstreaming is becoming an increasingly more relevant intervention within organizations.

It is necessary for concrete working conditions to be considered in research. This approach implies the existence of theoretical models that enable work to be defined in diverse fields. It further implies the existence of measuring instruments that enable work to be analyzed in different contexts. Both the theoretical models and the instruments of measurement do not exist. As mentioned previously, I/O psychology's focus on paid work (and on specific forms of paid work) indicates the limitation of applied models and instruments of measurement. Further developmental work has to be undertaken.

### See Also the Following Articles

Gender and Culture ■ Gender and Education ■ Industrial/Organizational Psychology, Overview ■ Occupational Psychology, Overview

### Further Reading

- Cleveland, J. N., Murphy, U. R., & Stockdale, M. (2000). *Women and men in organizations*. Hillsdale, NJ: Lawrence Erlbaum.
- Fredriksen-Goldsen, K. I., & Scharlach, A. E. (2001). *Families and work. New directions in the twenty-first century*. Oxford, UK: Oxford University Press.
- Gilbert, L. A., & Rader, J. (2001). Current perspectives on women's adult roles: Work, family, and life. In R. K. Unger (Ed.), *Handbook of the psychology of women and gender*. New York: Wiley.



# Openness to Experience

Robert R. McCrae

National Institutes of Health, Baltimore, Maryland, USA

1. Conceptualizations of Openness
  2. Openness and Creativity
  3. Openness and Vocational Behavior
- Further Reading

## GLOSSARY

**absorption** A trait characterized by intense concentration on an activity or experience.

**divergent thinking** The ability to generate many possible solutions instead of the single correct answer.

**tolerance of ambiguity** A willingness to deal with situations in which there is no clear-cut answer.

**transformational leadership** A leadership style required to make major changes in the direction or operation of an organization.

Openness to Experience is one of the dimensions of individual differences in the Five-Factor Model (FFM) of personality. Open individuals are characterized by an intrinsic interest in experience in a variety of areas and by a particularly fluid style of consciousness. Openness is strongly heritable and generally stable during adulthood. Open individuals are creative and adaptable, whereas closed people are pragmatic and down-to-earth. Recent research on Openness and vocational behavior shows that this factor is relevant to vocational interests, organizational change, and transformational leadership.

## 1. CONCEPTUALIZATIONS OF OPENNESS

Openness to Experience is the broadest of the five basic factors that comprise the Five-Factor Model (FFM), covering a wide range of loosely related traits. Among the relevant traits are tolerance of ambiguity, low dogmatism, need for variety, aesthetic sensitivity, absorption, unconventionality, intellectual curiosity, and intuition as measured by the Myers–Briggs Type Indicator. Perhaps because the links among such traits are not obvious, Openness is the least well understood of the five factors.

In studies of trait-descriptive adjectives, the fifth factor is usually called Intellect and is defined by terms such as perceptive, analytical, and reflective. Such words suggest intellectual ability, and Openness is sometimes confused with intelligence. In fact, intelligence (as measured by ability tests such as the Wechsler Adult Intelligence Scale) shows small correlations with Openness but forms a separate factor. Years of education is also modestly related to Openness; however, it is not clear whether education promotes Openness or whether open individuals pursue more education.

Perhaps the most widely used measure of this factor is the Openness scale of the Revised NEO Personality Inventory. In that instrument, Costa and McCrae provided subscales for six aspects of life to which individuals may be relatively more open or closed: fantasy, aesthetics, feelings, actions, ideas, and values. Thus, high scorers on the overall scale are imaginative,

responsive to art and beauty, attentive to their own feelings, willing to try new activities, intellectually curious, and unconventional. These traits appear to have both a motivational and a structural component. Motivationally, open people are receptive to new experience and to a deeper examination of their own thoughts, feelings, and values. Structurally, they have a permeable style of consciousness in which remote associations are easily made. In contrast, closed people are more comfortable with the world they know and tend to compartmentalize their ideas and feelings.

Although Openness appears to be an experiential style, it has pervasive social consequences. Open individuals tend to join liberal political parties and religious denominations. They seek friends who share their interests in the arts and tend to marry spouses who have similar open interests and values. Openness is also related to vocational interests and occupational behavior. Because Openness is so clearly expressed in attitudes and behaviors, there is generally strong agreement between self-reports and observer ratings on this factor.

In general, psychologists tend to be more open than closed, and many assume that Openness is a desirable trait. However, closed individuals prefer their own conservative and down-to-earth approach to life and would rate the desirability of Openness differently. Openness is associated with some negative outcomes, including nightmares, depression, and drug experimentation. Closedness is associated with authoritarianism and difficulty in adapting to change.

## **2. OPENNESS AND CREATIVITY**

It is generally conceded that creativity is a complex phenomenon, depending on cognitive ability, training and experience, and the social context. However, creativity has also repeatedly been associated with personality traits and motives, and most of these (e.g., tolerance of ambiguity, preference for novelty, independence of judgment) can be considered aspects of Openness. Several features of open people predispose them to creative work. The fluid nature of their consciousness means that they can make unusual and remote (and sometimes creative) associations. The fact that they are low on the need for closure means that they can suspend judgment and seek alternative solutions instead of accepting the first that presents itself. The fact that they enjoy novelty means that they are motivated to seek new ways in which to accomplish tasks. Their love of beauty stimulates artistic creativity.

The aspect of intelligence thought to be most closely related to creativity is divergent thinking, assessed by tasks that require multiple answers to a single question. For example, a respondent may be asked to write as many sentences as possible with words beginning with a fixed set of letters or to list all of the possible uses for a brick. Tests of divergent thinking are moderately related to Openness, even after controlling for age, vocabulary, and years of education. Open individuals may be more interested in these tasks, or they may in fact possess this specific cognitive ability to a greater extent.

Openness has also been linked to creative performance in tasks such as writing stories, and it has been shown to predict real creative achievements. In 1999, Helson's pioneering studies of women's creativity over the life span demonstrated that those women who were to make substantial creative achievements had been characterized in college by traits related to Openness. Indeed, one easy way in which to understand the construct of Openness is by considering as case studies creative geniuses such as Jean-Jacques Rousseau and Leonardo da Vinci.

## **3. OPENNESS AND VOCATIONAL BEHAVIOR**

During recent years, applied psychologists have begun to appreciate the importance of Openness for understanding vocational behavior. Openness is a major determinant of vocational interests. Open people prefer investigative and artistic occupations such as author, anthropologist, and research scientist, and they also endorse a wider range of interests. Closed people prefer conventional occupations such as banking and bookkeeping.

Open people welcome change, so it is not surprising that they are more likely than closed people to make mid-career shifts. Recent studies have also shown that they adapt more easily to organizational change. Part of that adaptability is tied to their willingness to seek information. Among new employees, more open individuals are more likely to seek out feedback about their performance from supervisors and coworkers, and they are more likely to regard learning a new job as a challenge. Meta-analyses have shown that open employees respond more favorably to training programs.

Because open men and women are creative and willing to seek new directions, they can become effective leaders of change, and studies have shown that Openness is associated with transformational leadership. Leaders, of course, also require other traits

(e.g., Conscientiousness) and skills, but for some leadership tasks, Openness may be a great asset.

It must be recalled that Openness is not inevitably superior to closedness. For many vocations and many roles within every organization, there are advantages to being conservative and conventional. Maintaining tradition can be as important as is creating innovation. Assessments of Openness can assist managers in assigning tasks and in understanding how employees are likely to perceive and react to organizational change.

### See Also the Following Articles

Holland's Theory (Vocational Personality Types) ■ Leadership and Culture ■ Personality Assessment ■ Power, Authority, and Leadership ■ Psychometric Tests ■ Vocational Interests ■ Vocational Psychology, Overview

### Further Reading

Helson, R. (1999). A longitudinal study of creative personality in women. *Creativity Research Journal*, 12, 89–101.

Judge, T. A., & Bono, J. E. (2000). Five-Factor Model of personality and transformational leadership. *Journal of Applied Psychology*, 85, 751–765.

McCrae, R. R. (1994). Openness to Experience: Expanding the boundaries of Factor V. *European Journal of Personality*, 8, 251–272.

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin*, 120, 323–337.

McCrae, R. R., & Costa, P. T., Jr. (1997). Conceptions and correlates of Openness to Experience. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 269–290). Orlando, FL: Academic Press.

Van Hiel, A., Kossowska, M., & Mervielde, I. (2000). The relationship between Openness to Experience and political ideology. *Personality and Individual Differences*, 28, 741–751.

Wanberg, C. R., & Kammeyer-Mueller, J. D. (2000). Predictors and outcomes of proactivity in the socialization process. *Journal of Applied Psychology*, 85, 373–385.

Wolfenstein, M., & Trull, T. J. (1998). Depression and Openness to Experience. *Journal of Personality Assessment*, 69, 614–632.



# Optimism

Christopher Peterson

University of Michigan, Ann Arbor, Michigan, USA

Nansook Park

University of Rhode Island, Kingston, Rhode Island, USA

1. Introduction
2. Contemporary Approaches to Optimism
3. Measures of Optimism
4. Outcomes
5. Development of Optimism
6. Deliberate Interventions
7. Conclusion

Further Reading

have approached it under the rubrics of dispositional optimism, explanatory style, and hope. Regardless of how optimism is measured, researchers usually find that it is associated with desirable outcomes, with one notable exception: Optimistic individuals underestimate likelihood of future risks. More is known about the development of pessimism than of optimism, although deliberate interventions can nurture optimism among youth as well as adults.

## GLOSSARY

**agency** Determination that goals can be achieved.

**dispositional optimism** Global expectation of optimism versus pessimism.

**explanatory style** Habitual style of explaining the causes of bad events, either with external, unstable, and specific causes (optimism) or with internal, stable, and global causes (pessimism).

**hope** Global belief that goals can be achieved (agency) by generating successful plans (pathways).

**learned helplessness** Passivity produced by learning that responses and outcomes are independent.

**optimism** Belief that the future holds more good events than bad events.

**pathways** Beliefs that successful plans can be generated to reach goals.

Optimism refers to the belief that the future holds more good events than bad events. Optimism is a popular topic of study among contemporary psychologists, who

## 1. INTRODUCTION

Optimism represents a cognitive, emotional, and motivational stance toward the future. Thinking about the future, expecting that desired events and outcomes will occur, acting in ways believed to make them more likely, and feeling confident that these will ensue given appropriate efforts sustain good cheer in the here-and-now and galvanize goal-directed actions.

The term optimism is a relatively recent arrival on the historical scene, as is its ostensible cousin pessimism. In the 1700s, Leibniz characterized optimism as a mode of thinking, and Voltaire popularized the term in his 1759 novel *Candide*, which was highly critical of the shallowness of an optimistic perspective. Interestingly, pessimism appeared fully a century later, when it was independently introduced by Schopenhauer and Coleridge.

At least in their original forms, optimism and pessimism were not symmetric. Optimism as discussed by Leibniz was cognitive in its emphasis, reflecting a

reasoned judgment that good would predominate over evil, even if goodness were sometimes associated with suffering. In contrast, pessimism as discussed by Schopenhauer had an emotional meaning: The pessimistic individual was one for whom suffering would outweigh happiness. Someone can be optimistic in the cognitive Leibniz sense yet pessimistic in the emotional Schopenhauer sense.

As these notions evolved in the language, optimism and pessimism became juxtaposed, and the connotations of each spilled over into one another. The resulting composite is what psychologists usually mean by these terms, a dimension anchored by optimism at one end and pessimism at the other. Accordingly, optimism refers to a belief—perhaps wish would be a better term or even motive—that in the future good events and associated positive feelings will outweigh or be more likely than bad events and associated negative feelings.

## 2. CONTEMPORARY APPROACHES TO OPTIMISM

Contemporary writers approach this family of strengths in two ways. Some—usually social philosophers—treat these as features of general human nature to be praised or decried depending on the writer's own biases, whereas others—usually research-minded psychologists in the personality or clinical tradition—regard optimism and pessimism as individual differences, characteristics that people possess to varying degrees.

There are three well-known approaches within psychology to this strength as an individual difference. Each line of work has an associated self-report measure, each usually has focused on the consequences of the individual difference as opposed to the antecedents, and each has spawned a large empirical literature demonstrating that hope and optimism (or at least the absence of their opposites) are associated with desirable outcomes—positive mood and good morale; perseverance and effective problem solving; academic, athletic, military, occupational, and political success; popularity; good health; and even long life and freedom from trauma.

### 2.1. Dispositional Optimism

Charles Carver and Michael Scheier pioneered the study of a personality variable they identify as dispositional optimism, which is the global expectation that good things will be plentiful in the future and bad

things scarce. Their overriding perspective is in terms of how people pursue goals, defined as desirable values. To them, virtually all realms of human activity can be cast in goal terms, and people's behavior entails the identification and adoption of goals and the regulation of actions vis-à-vis these goals. They therefore refer to their approach as a self-regulatory model. Optimism enters into self-regulation when people ask themselves about impediments to the achievement of the goals they have adopted. In the face of difficulties, do people nonetheless believe that goals will be achieved? If so, they are optimistic; if not, they are pessimistic. Optimism leads to continued efforts to attain the goal, whereas pessimism leads to giving up.

### 2.2. Explanatory Style

Martin Seligman and colleagues have approached optimism in terms of an individual's characteristic explanatory style. Those who explain bad events in a circumscribed way, with external, unstable, and specific causes, are described as optimistic, whereas those who favor internal, stable, and global causes are described as pessimistic. The notion of explanatory style emerged from the attributional reformulation of the learned helplessness model. Briefly, the original helplessness model proposed that following experience with uncontrollable aversive events, animals and people become helpless—passive and unresponsive—presumably because they have “learned” that there is no contingency between actions and outcomes. This learning is represented as a generalized expectancy that future outcomes will be unrelated to outcomes. It is this generalized expectation of response–outcome independence that produces later helplessness. However, the helplessness/explanatory style research tradition has infrequently examined expectations. Rather, researchers measure explanatory style and correlate it with outcomes thought to revolve around helplessness: depression, illness, and failure in academic, athletic, and vocational realms. In contrast to dispositional optimism, explanatory style is a construct concerning agency.

### 2.3. Hope

Rick Snyder's studies of hope in effect integrate these two visions of optimism—expectation and agency. Snyder traced the origins of his thinking to earlier work on expectations and goal achievement. According to Snyder's view, goal-directed expectations are composed of two separable components. The first is agency,

and it reflects someone's determination that goals can be achieved. The second is identified as pathways, which represent the individual's belief that successful plans can be generated to reach goals.

### 3. MEASURES OF OPTIMISM

Optimism (and/or pessimism) and hope (and/or hopelessness) have over the years inspired many measures. Most have been face-valid self-report questionnaires; other measures have relied on interviews, observer reports, or content analyses of written or spoken material.

Measures of optimism are usually based on what an individual says about what he or she usually believes. This strategy follows from the central role accorded expectations in the various definitions of these constructs; expectations are sensibly measured by self-report of beliefs. Each of the well-known contemporary research traditions has been facilitated by the existence of reliable and valid measures, such as the Life Orientation Test (LOT) to measure dispositional optimism, the Attributional Style Questionnaire to measure explanatory style, and the Hope Scale to measure agency and pathways. Each of these measures has been refined and revised several times to improve reliability and validity, and they are respected tools in the arsenal of research psychologists.

In the case of explanatory style and hope, there are separate self-report measures available for children and adults, as well as content analysis strategies for scoring these characteristics from written or spoken material. Ways to assess hope by observer report and from interviews also exist. Hope measures, and to some degree explanatory style measures, have spawned various domain-specific versions (e.g., hope concerning family life and explanatory style for academics). Dispositional optimism and pessimism have not been assessed among children and youth, although the straightforwardness of the LOT would presumably allow its use with early teens or even younger individuals.

A problem with respect to explanatory style is that questionnaire and content analysis measures do not impressively converge, although their correlates are usually identical. More generally, convergence among the different measures of optimism and hope has not been the subject of much research, although the typical study finds a modicum of agreement.

It matters whether measured expectations are about good events versus bad events. Therefore, the LOT has separate optimism and pessimism items, and these are

somewhat independent and occasionally have different correlates. Explanatory style for good events is independent of explanatory style for bad events; the correlates are usually opposite for "good" versus "bad" explanatory styles and more robust for the "bad" style. This latter finding may be an artifact of the tendency in the explanatory style tradition to study negative outcomes (depression, disease, and failure).

### 4. OUTCOMES

As noted previously, optimism and hope predict many desirable outcomes: achievement in all sorts of domains (academic, athletic, military, political, and vocational), freedom from anxiety and depression, good social relationships, and physical well-being. More specific correlates include active problem solving and attention to problem-relevant sources of information, as well as conscientiousness, diligence, and the ability to delay gratification. People who are optimistic make "to do" lists, use day planners, and wear wristwatches; they also balance their checkbooks. All of these activities imply an active orientation to the future.

The only well-documented downside of optimism is its link with biases in risk perception. Optimistic people exaggerate the widespread tendency of most individuals to see themselves as below average for such dire occurrences as cancer and heart disease, failure and heartbreak, and so on. This bias can be a problem when reality matters and optimism leads one to neglect preventive or remedial actions. However, if people acknowledge the possibility of risk, the more optimistic among them take the most appropriate coping steps.

### 5. DEVELOPMENT OF OPTIMISM

Most of the relevant research has examined the environmental influences on explanatory style, and here it is clear what makes explanatory style more pessimistic: trauma and failure of all sorts. Less is known about what fosters an optimistic style, although an inconsistent literature finds that the explanatory styles of children and parents sometimes converge. Perhaps modeling is implicated, but so too might be shared family environment or even indirect genetic influence, as already discussed. If explanatory style is modeled, one needs to consider not just parents but also teachers, coaches, peers, and the media. Along these lines, a neglected but

likely important influence on the development of optimism is adherence to orthodox religion.

## 6. DELIBERATE INTERVENTIONS

The Penn Prevention Program (the Penn Resiliency Project) is a deliberate intervention to foster optimism. Schoolchildren are taught basic cognitive-behavioral strategies, especially disputation and decatastrophizing, for thinking about the causes of events in more optimistic fashion. This intervention makes explanatory style more optimistic and prevents the later development of depression and anxiety. It has also been demonstrated that both hope and explanatory style change in the course of individual psychotherapy with adults, especially of the cognitive ilk.

Missing in any of these intervention projects is an explicit focus on the fulfillments that presumably accompany optimism and hope. Thus, we know that the Penn Prevention Program makes emotional distress and even physical illness less likely, but it is unknown whether it makes youth happy and healthy.

## 7. CONCLUSION

Optimism and hope have had a checkered reputation since their introduction centuries ago. To call someone a Pollyanna is an insult, and to speak of cockeyed optimism or false hope is to underscore the foolishness of an endeavor that is based on an unrealistic appraisal of future risks. However, the overall benefits of optimism are not diminished by these occasional excesses. The most practical implication of optimism and hope

research is to answer the question “Should one be optimistic?” by saying, “Yes, unless there is reason not to.” In other words, optimism and hope should be regarded as cognitive strategies under the person’s control—often useful but sometimes not—as opposed to immutable dispositions.

## See Also the Following Articles

Personality Assessment

## Further Reading

- Buchanan, G. M., & Seligman, M. E. P. (Eds.). (1995). *Explanatory style*. Hillsdale, NJ: Lawrence Erlbaum.
- Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, *97*, 19–35.
- Peterson, C. (2000). The future of optimism. *American Psychologist*, *55*, 44–55.
- Peterson, C., Maier, S. F., & Seligman, M. E. P. (1993). *Learned helplessness: A theory for the age of personal control*. New York: Oxford University Press.
- Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco: Freeman.
- Seligman, M. E. P., Reivich, K., Jaycox, L., & Gillham, J. (1995). *The optimistic child*. Boston: Houghton Mifflin.
- Sicinski, A. (1972). Optimism versus pessimism: Tentative concepts and their consequences for future research. *Polish Sociological Bulletin*, *25/26*, 47–62.
- Snyder, C. R. (Ed.). (2000). *Handbook of hope: Theory, measures, and applications*. San Diego: Academic Press.
- Tiger, L. (1979). *Optimism: The biology of hope*. New York: Simon & Schuster.
- Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science*, *246*, 1232–1233.



# Organizational Culture and Climate

Mark F. Peterson

Florida Atlantic University, Boca Raton, Florida, USA

Ronald Fischer

Victoria University, Wellington, New Zealand

1. Basic Concepts
2. History of Organizational Climate Analysis
3. History of Organizational Culture Analysis
4. Organizational Climate Change and Consultancy
5. Organizational Culture Change and Consultancy
6. The Questionable Virtues of Strong Organizational Cultures
7. Are Organizational Climate and Organizational Culture Analogous to Characteristics of Larger Societies and Nations?
8. Institutional Analysis as an Alternative to Climate and Culture
9. Recent Literature Reviews and Compilations  
Further Reading

## GLOSSARY

**institution** In classical institutional theory, explicit structures (e.g., nations, organizations) created by legal means that give the institutions explicitly delineated authority to control their members and to enter into agreements with other parties; in neo-institutional theory, institutions are sets of understandings, actions, and interactions that characterize an identifiable, but not necessarily sharply or explicitly bounded, entity.

**Likert scale** A method for measuring attitudes, beliefs, or values in which a respondent answers a short question using a quantitative response scale, typically ranging from 1 to 5 or from 1 to 7, with each quantitative response linked to a verbal anchor reflecting a continuum (e.g., from *strongly agree* to *strongly disagree*).

**organizational climate** A term used to refer to how sets of individuals perceive and come to describe their work environment, their job, and/or organizational practices, procedures, and behaviors that are rewarded and supported, often in relation to the implicit goal priorities of the organization such as a climate for service or a climate for trust.

**organizational culture** A term used to refer to the configuration of values, implicit assumptions, and understandings that are expressed in overt statements, patterns of behavior, and interpretations of symbols that distinguish one organization from others.

**strong organizational culture** An organizational culture that is understood to have particularly strong effects on the ways in which organization members think and behave; it is usually contrasted with competing influences on organization members other than culture, including direct supervisory oversight, rules such as job descriptions and budgets, and explicit contracts.

**values** Desired states of affairs variously used to reflect implicit or explicit preferences, sometimes ordered into means-ends chains such that some values are viewed as preferred only as a means to achieving other outcomes, applied at various social levels from individuals across societies.

Organizational climate and organizational culture are terms that are used to comprehensively represent the social quality of organizations. The term “organizational climate” reflects psychological and sociological thinking about organizations, whereas the term “organizational culture” includes an anthropological element. Organizational climate is typically analyzed by using work attitude questionnaires. The scholarly literature

about climate often has been devoted to technical survey design and data analysis problems. These include how to reliably and validly measure individuals' attitudes and combine the answers to represent characteristics of whole organizations. Organizational culture is sometimes represented by surveys about values but is more typically analyzed by using interviews to gather stories or accounts of organizations from their members or by directly observing organizations either as an outsider or as a participant. Both terms are applied by managers and consultants who are interested in promoting or evaluating large-scale organization change.

## 1. BASIC CONCEPTS

Organizational climate and organizational culture both began as terms for comprehensive overviews of what an organization as a whole is like. The term "organizational culture" has retained its meaning as a comprehensive representation, whereas some uses of the term "organizational climate" have drifted from the original use. Organizational culture is most often used for referring to values and understandings, broadly shared in an organization or in some part of an organization, that influence the thoughts and actions of organization members. The term "climate" is now used for three purposes: (a) as psychological climate to describe questionnaire surveys that provide a broad profile of attitudes and perceptions of an organization's members covering more specific topics (e.g., job satisfaction, leadership, group dynamics), (b) as organizational climate to describe shared perceptions of organization members about their work environment and organizational policies and procedures that are likely to affect their well-being, and (c) as facet-specific climate to represent the distinctive value focus or profile of value priorities in an organization such as a climate for service or a climate for innovation. Although the concepts of organizational climate and organizational culture have come to be differentiated to some degree, each also has a heritage that reflects an interest in a holistic view of organizations.

## 2. HISTORY OF ORGANIZATIONAL CLIMATE ANALYSIS

The concept of organizational climate has a longer history than does the concept of organizational culture. The former derives from analyses during the 1930s of

the qualities of a variety of social entities ranging in scope from a nation down to a group. Social climate was evoked to discuss the rise of fascism in Germany before World War II and the response of children's groups to adult leaders. The application of the social climate idea to organizations coincided with the development of the fields of organizational psychology and organizational behavior. Organizational climate was most often invoked when consultants and scholars working in these fields sought to provide a broad overview of the social side of an organization.

During the 1960s, climate surveys began to be used to ask organization members about what they experienced their organizational surroundings as being like. Employee responses were then combined to paint an overall picture of an organization that managers could use to evaluate the need for organization change and to plan and assess the effects of changes. Climate surveys included topics such as group dynamics, leadership, employee motivation, job characteristics, and management propensities to be supportive and fair. Many of these topics have become part of separate lines of research only loosely informed by climate research and are now most often encountered as distinctive literatures.

Likert, a social scientist and consultant with interests in attitude measurement, sociology, psychology, and organization change, inspired a great deal of the early organizational climate work. Most climate surveys use the now familiar Likert-type scale questions that ask people to rate work attitudes and perceptions. The typical scale for a climate item ranges from 5 to 7 points and includes verbal anchors ranging from "strongly agree" to "strongly disagree" (an agreement scale), from "to a very great extent" to "to a very small extent" (an extent scale), or from "very often" to "seldom or never" (a frequency scale). Climate surveys were one application to organizations of the questionnaire methods that began being designed for sociological and psychological research during the 1930s. Organizational climate surveys depended on emerging data analysis methods such as factor analysis during the early postwar years. The analysis of climate data became both more feasible and more complex as the field took advantage of developments in computer technology beginning in the 1970s. These methods made it increasingly practical to analyze data sets based on hundreds or thousands of respondents answering hundreds of questions.

The concept of organizational climate has evolved, in part, to sharpen its distinction from the concept of organizational culture. Three main perspectives have

been taken in the scholarly literature to distinguish it from organizational culture. First, climate has been conceptualized as individuals' representations or cognitive schema of their work environment. Individuals are thought to respond to situational variables in a manner that is psychologically meaningful to them. Therefore, this perspective focuses on the individual and is concerned with the psychological description of organizational conditions.

The second perspective focuses on shared perceptions of climate that can be used to describe either an entire organization or a work team. If individuals agree with each other about organizational policies, practices, or procedures that affect their well-being, these perceptions can be aggregated to the organizational level and used to describe the organization. Proponents of this approach have faced several methodological challenges. One is the difficulty of agreeing about appropriate ways in which to conceptualize and measure climate perceptions. Another is the establishment of statistical methods and cutoff criteria to determine that there is sufficient agreement or consensus among organizational members to believe that perceptions are sufficiently shared among organizational members to be used to describe the larger organization. This perspective is based on the argument that for shared perceptions to arise, it is necessary that (a) individuals interact at work, (b) individuals share a common goal or work toward an attainable outcome that predisposes them to collective action, and (c) individuals have sufficient task interdependence so that they develop a shared understanding of behavior. Because these conditions are more likely to be found in work teams than in the overall organization, proponents of this perspective argue that it is more likely that a shared climate develops at a group or team level rather than at an organizational level, especially if the organization is large, divisionalized, and multilayered. Consequently, this perspective on organizational climate has recently renewed the original social climate theme of group or team climates. Another offshoot of work treating climate as agreement or consensus has been the study of climate strength conceptualized as the level of agreement or consensus among organization members. Some argue that the effect of organizational climate on other variables, such as customer satisfaction, depends on how much members agree with each other, indicating a strong organizational climate.

Finally, climate has been used to represent a particular focus—a "climate for" something. Advocates of such facet-specific climates argue that using the term

"climate" requires a specific referent such as a climate for service, a climate for change, or a climate for innovation. This perspective has revived the interest in climate research and has provided a new perspective for studying organizational phenomena with a strong applied focus. This facet-specific approach to climate also has been argued to sharpen the distinction of climate from the broader construct of organizational culture.

### 3. HISTORY OF ORGANIZATIONAL CULTURE ANALYSIS

Organizational culture is the newer of the two terms in organizational analysis, although (paradoxically) the use of the agricultural idea of culture as a social metaphor has quite a long history. Organizational culture is linked to analyses of societal culture having their origins during late 19th-century cultural anthropology. The basis of the metaphor is that just as societies engage in the culture of animals and plants, they also actively engage in social culture.

The relevance of anthropological perspectives on culture for managing organizations came to the worldwide attention of managers during the 1980s, although occasional references to the culture of organizations appeared several decades earlier. During the 1980s, Japanese management techniques such as quality circles were being widely imitated throughout the world, and Japanese companies were making many acquisitions throughout the United States, Europe, Asia, and Latin America. These business activities initiated debates about whether the success of Japanese organizations was due more to management practices that could readily be transferred or to distinctive qualities of Japanese culture that were either hard or impossible to imitate. Scholars were also showing increased dissatisfaction with limitations in climate research, notably with whether climate surveys were really more suited to representing the attitudes of individuals than to representing the attitudes of whole organizations. In addition, an even more basic set of debates was fomenting in academic management theory about how adequate quantitative techniques were for representing the social aspects of organizations. Although explicit discussion of Japan now has fallen out of use in the organizational culture literature, many characteristics of culture analysis can be traced to issues in Japan's effects on the business world during the 1980s.

The organizational culture literature has drawn attention to a number of aspects of organizations that had been overlooked previously. For example, in 1985, Schein drew directly from a functional cultural anthropology theory to suggest that all organizations need to deal with basic problems such as whether people are basically good, evil, or some combination of the two. The culture of an organization, notably its approach to control, was understood to be linked to the often unstated assumptions about such matters that tended to prevail. Other culture authors drew attention to the symbolism of objects such as the physical facility and the ways in which employees dressed. The anthropological tradition of interpreting stories as culture indicators was reflected in interest in organization stories and myths and in the distinctive roles that influential organization members played in transmitting these stories.

The organizational culture literature has also sought to use the methods developed for analyzing organizational climate and to adapt them to the new themes brought out by organizational culture analysts. In particular, climate surveys have focused on employees' attitudes and perceptions, whereas culture surveys have focused on values. This quantitative side of the organizational culture literature has encountered controversies about whether it is "really" culture analysis or is instead a form of climate analysis. Value surveys had in fact existed and been used frequently in psychology since the 1970s, but until the appearance of organizational culture, they had rarely been part of organization analysis.

#### **4. ORGANIZATIONAL CLIMATE CHANGE AND CONSULTANCY**

Organizational climate occasionally appears as a topic in an academic journal article, but it is more significant now as a routine part of organization assessments by corporate human resources managers in large organizations and by management consultants. Beginning in the 1960s, organizational climate surveys became part of the process of diagnosing organizations to identify how they need to change and providing feedback to promote change. For example, first a climate consultant would conduct a questionnaire survey for people throughout an organization. Afterward, the consultant would provide summaries of the survey results to various groups in the organization so that they could

discuss what has produced the attitudes typical of their groups and what changes in the work system they might like to make. Alternatively, a consultant would provide managers with a summary report to inform their decisions about leadership training programs, improving the pay administration system, and/or implementing some other human resources change. Climate surveys are also used to assess the effects of either changes in an organization's competitive environment or new internal procedures initiated by management.

Following Likert, the types of organization improvements typically considered when climate surveys were first used were to increase employee participation in an organization's decision-making processes and to promote employees' satisfaction. These applications began during what is sometimes called the human relations period in organization analysis, that is, when management scholars and consultants typically had an egalitarian bias and accepted a view that employee satisfaction promotes employee performance. Although subsequent research has indicated that employee participation and satisfaction can contribute to organization performance under some circumstances, the consensus is that early advocates exaggerated their significance. More recent uses of climate surveys have retained the survey administration and feedback approach but have focused on identifying impediments that an organization is encountering toward achieving some particular goal such as service, trust, or innovation. Climate surveys remain one of the best ways in which to involve all organization members, regardless of their level, in diagnosing and assessing organizational changes.

#### **5. ORGANIZATIONAL CULTURE CHANGE AND CONSULTANCY**

The idea of organizational culture tended to come from consultancy-oriented approaches to organizations that appeared during the late 1970s and early 1980s and then were followed by scholarly assessments. Even more than organizational climate, organizational culture is now more firmly rooted in the management consultancy literature than in the academic management literature. One line of organizational culture consultancy advocates the virtues of establishing strong cultures. A second describes ways in which corporate leaders can show transformational leadership to effect cultural change. A third focuses more on diagnosing cultures using questionnaires in a way

that is reminiscent of organizational climate research. As detailed previously, organizational climate surveys focus on employee attitudes, beliefs, and perceptions about the people and groups around them, whereas organizational culture surveys focus on values. Some applications of culture surveys have been very similar to organizational climate applications, identifying values to the end of deciding whether and how to change them. Other applications have been to understanding and trying to promote effective work between groups that need to work together. These might be departments that traditionally have different values such as production and marketing, notably production efficiency versus customer responsiveness. They might also be parties in a merger or joint venture situation that are seeking to either interact more effectively or to anticipate and solve problems when combining into a single entity.

## **6. THE QUESTIONABLE VIRTUES OF STRONG ORGANIZATIONAL CULTURES**

Many organizational culture consultants have recommended that organizations with strong cultures had advantages over those with weaker cultures. Specific views of what culture strength means vary, but most views have some common elements. These are that a strong culture includes understandings and assumptions that are broadly shared, as well as norms and values that are broadly accepted, throughout an organization.

Originally, there were two bases for advocating strong organizational cultures, both linked to presumed characteristics of Japanese society and the Japanese organizations that rose to world preeminence during the 1980s. One argument linked the advocacy of strong cultures to the idea of employee commitment. Theory and research related to organization commitment was developing at the same time as was culture research. It provided part of the explanation as to why Japanese culture promoted a generally more committed workforce than did most national cultures. Inspired by this belief, many scholars and consultants expected that a consistent link between commitment and performance could be documented despite the unsuccessful prior attempts to link satisfaction to performance. This hope was contradicted by research results that linked high commitment to low turnover but not to high levels of typically measured performance criteria.

A second argument was based on the idea of transaction cost economics that was being introduced to organization studies at about the same time. One category of transaction consists of those that occur between an organization and employees. Transaction cost thinking drew attention to the costs that organizations incurred in supervising people. One of the themes that linked anthropological understandings of culture to the economics of transaction costs was that cultural systems of norms and understandings could reduce the expense of monitoring employees. Japanese organizations were thought to economize on monitoring because employees knew how to behave appropriately due to a combination of general societal culture understandings and their socialization into an organization's culture. Some have argued that this understanding reduces the necessity of expensive ongoing performance evaluation and other monitoring systems. Based on both the expectation that employee commitment would have economic benefits and the view that the reduced need for monitoring would reduce monitoring costs, strong cultures came to be advocated. The virtue of strong organizational cultures continues to have its advocates, but the research base supporting the strong culture position has provided equivocal results.

## **7. ARE ORGANIZATIONAL CLIMATE AND ORGANIZATIONAL CULTURE ANALOGOUS TO CHARACTERISTICS OF LARGER SOCIETIES AND NATIONS?**

The histories of climate and culture both are linked to analyses of larger societies. Culture theory was stimulated by questions about the business success of Japan, and climate research was influenced by attempts to understand how the social climate of Germany was transformed by fascism prior to World War II. The question of the relationship between nation characteristics and organization characteristics has continued. Part of the debate is about whether socialization into organization values has the same quality as does socialization into values while growing up as a child. Some would argue that the process is identical, whereas others would argue that organizations develop distinctive values either by selecting (or encouraging self-selection for) a particular type of value or by continually seeking to make salient a subset of those values developed during childhood.

## 8. INSTITUTIONAL ANALYSIS AS AN ALTERNATIVE TO CLIMATE AND CULTURE

The ideas of climate and culture have similarities to the idea of institution that is increasingly used in organization analysis. Institutions, like cultures, are viewed as patterns of understandings and actions that maintain stability in social systems and distinguish one system from another. Like organizational culture, institutions that appear as typical organizational practices, assumptions, and norms in one nation can be transferred to other nations, but the transfer requires adaptation to preexisting institutions. The term “culture” tends to be used in analyses that center on values, whereas institutions cover a broader range of phenomena that include established practices. The theory of institutions comes out of political science and is that field’s way of dealing with the type of issues and phenomena that climate covers in sociology and psychology and that culture covers in anthropology.

## 9. RECENT LITERATURE REVIEWS AND COMPILATIONS

Numerous books offer managers guidance about particular ways in which to understand and change organizational climates and cultures. The seminal book *Organizational Climate and Culture*, edited by Schneider in 1990, was the first to focus on organizational climate and culture simultaneously and still marks a milestone in this area of research. Four more recent books provide scholarly reviews that expand and deepen these discussions and provide careful evaluations of many perspectives on climate and culture change. Trice and Beyer, in their 1993 book, thoroughly reviewed how the organizational culture concept developed in a way that highlights its anthropological roots and its distinctness from organizational climate. Academic articles by Denison in 1996 and by Verbeke and colleagues in 1998 discussed the historical development and conceptual overlap between the two constructs. Handbooks edited by Ashkanasy and colleagues in 2000 and by Cooper and colleagues in 2000 compiled an eclectic mix of perspectives on topics related to organizational climate and culture. These recent reviews and integrations suggest a number of directions being taken in organizational climate and culture research and consultancy in

addition to those noted previously. In addition to covering many of the basic issues in climate and culture noted in the current review, some of the chapters in these handbooks linked organizational culture to other topics such as symbols, time, meaning, careers, individual attachment to organizations, and roles. In 2002, Martin provided a book-length review that frames the fields of climate and culture by using them as examples of struggles among competing basic conceptual perspectives on organization studies. Although organizational climate and culture are now less often the main focus of scholarly journal articles in management than they are in scholarly and applied books, some recent articles in the *Journal of Applied Psychology* and the *Journal of Organizational Behavior* show that academic interest continues. Furthermore, many topics that either first appeared or gained momentum through climate surveys and culture analyses are now studied separately.

### See Also the Following Articles

Industrial/Organizational Psychology, Overview ■ Organization Development

### Further Reading

- Anderson, N. R., & West, M. A. (1998). Measuring climate for work group innovation: Development and validation of the team climate inventory. *Journal of Organizational Behavior, 19*, 235–258.
- Ashkanasy, N., Wilderom, C., & Peterson, M. F. (Eds.). (2000). *Handbook of organizational culture and climate*. Thousand Oaks, CA: Sage.
- Cooper, C. L., Cartwright, S., & Earley, P. C. (Eds.). (2000). *The international handbook of organizational culture and climate*. New York: John Wiley.
- Denison, D. R. (1996). What is the difference between organizational culture and organizational climate? A native’s point of view on a decade of paradigm wars. *Academy of Management Review, 21*, 619–654.
- Hofstede, G. (2001). *Culture’s consequences* (2nd ed.). Thousand Oaks, CA: Sage.
- James, L. R., James, L. A., & Ashe, D. K. (1990). The meaning of organizations: The role of cognitions and values. In B. Schneider (Ed.), *Organizational climate and culture* (pp. 40–84). San Francisco: Jossey-Bass.
- James, L. R., Joyce, W. F., & Slocum, J. W. (1988). Comment: Organizations do not cognize. *Academy of Management Review, 13*, 129–132.
- Likert, R. (1961). *New patterns in management*. New York: McGraw-Hill.

- Lincoln, J. R., & Kalleberg, A. L. (1990). *Culture, control, and commitment*. Cambridge, UK: Cambridge University Press.
- Martin, J. (2002). *Organizational culture: Mapping the terrain*. Thousand Oaks, CA: Sage.
- Ouchi, W. G. (1980). Markets, bureaucracies, and clans. *Administrative Science Quarterly*, 25, 129–141.
- Schein, E. (1985). *Organizational culture and leadership*. San Francisco: Jossey-Bass.
- Schneider, B. (Ed.). (1990). *Organizational climate and culture*. San Francisco: Jossey-Bass.
- Schneider, B., Salvaggio, A. N., & Subirats, M. (2002). Climate strength: A new direction for climate research. *Journal of Applied Psychology*, 87, 220–229.
- Schneider, B., White, S. S., & Paul, M. C. (1998). Linking service climate and customer perceptions of service quality: Tests of a causal model. *Journal of Applied Psychology*, 83, 150–163.
- Scott, W. R. (1995). *Institutions and organizations*. Thousand Oaks, CA: Sage.
- Trice, H. M., & Beyer, J. M. (1993). *The cultures of work organizations*. Englewood Cliffs, NJ: Prentice Hall.
- Verbeke, W., Volgering, M., & Hessels, M. (1998). Exploring the conceptual expansion within the field of organizational behavior: Organizational climate and organizational culture. *Journal of Management Studies*, 35, 303–329.



# Organizational Diagnosis

André Büssing<sup>†</sup>

Technical University of Munich, Munich, Germany

1. The Concept of Organizational Diagnosis
  2. Goals and Tasks
  3. Methods
  4. An Example: Organizational Diagnosis of Climate and Culture
  5. Conclusion
- Further Reading

## GLOSSARY

*climate and culture* Important nonstructural organizational coordination mechanisms, representing an exemplary area for organizational diagnosis.

*organizational diagnosis* The systematic and scientifically based assessment, description, explanation, and prediction of regularly occurring experience, behavior, and performance of members of organizations and their interaction.

Organizational diagnosis is the systematic and scientifically based assessment, description, explanation, and prediction of regularly occurring experience, behavior, and performance of members of organizations and their interaction.

## 1. THE CONCEPT OF ORGANIZATIONAL DIAGNOSIS

Organizational diagnosis (OD) is a multistage process of assessment, analysis, description, explanation, and

<sup>†</sup>Deceased.

prediction. Features to be assessed are identified on the basis of theoretical considerations of relevant organizational characteristics and their relations (systematic). The diagnostic instruments have to comply with the standards of empirical research (scientific basis). The experience, behavior, and performance of organizational members, and also their interaction, are examined, setting aside individual differences in personality. The intention is not merely data gathering but also explanation and prediction.

Therefore, OD is connected to the understanding of organization and organizational psychology. Despite their diversity, organizational theories and models are based on the assumption that orders and rules exist in organizations and influence experience and behavior as well as performance of organizational members. The function of organizational psychology can be defined as the description, explanation, prediction, and modification of psychological phenomena in organizations. Its subject remains the organization with its multiple levels—individual and work task, group, unit, and organization—with the focus depending on the organizational theory being used. Consequently, the definition, goals, and tasks of OD are connected to organizational theories and models, and OD is always conducted from a certain position in organizational psychology.

There are several other definitions of OD; the majority date to the 1980s and do not originate from applied psychology but, rather, from management and organizational sociology. For example, Van de Ven and Ferry introduced a comprehensive approach to OD. Their framework of OD, which they called

organizational assessment, draws on psychology, sociology, and management and incorporates a high degree of integration of levels, perspectives, and interdisciplinarity. Harrison and Shirom present a pragmatic approach to OD, which they define as investigations, using models and methods of the behavioral sciences, to examine the state of organizations, solve problems, or enhance organizational effectiveness. Whereas the term OD in this article is reserved for a predominantly psychological approach, and approaches from business administration, management, and organizational sociology are referred to as organizational analysis, Harrison and Shirom's approach is in between. Addressing all organizational levels, including individuals and jobs, with a focus on the application and implementation of OD, they represent a management-oriented position, primarily responsive to the needs of managers and consultants.

## 2. GOALS AND TASKS

OD represents a methodological framework for assessing, describing, explaining, and predicting experience, behavior, performance, and interaction of organizational members. This general goal of OD can be realized concretely in a variety of specific topics. The diagnosis of climate and culture and the diagnosis of quality in organizations are examples of the wide range of practical applications of OD. Consequently, the tasks of OD comprise a broad spectrum. Lawler *et al.* identified three groups of stakeholders interested in data and results of OD: members of the organization (employees and management), external stakeholders (e.g., government, shareholders, and responsible public bodies), and researchers.

Accordingly, they distinguish three kinds of tasks and applications for data and results of OD. For decision makers (management) in organizations, these comprise

- Organizational diagnostic data to improve human resource management
- Organizational diagnostic data to support staffing decisions about internal transfers and job changes
- Organizational diagnoses to prepare for and support organizational changes (i.e., conducting systematic and continuous analyses of deficiencies) pertinent to implementing strategic changes
- Organizational diagnoses to evaluate organizational program activities—structural (e.g., decentralization via

team structures), psychological (e.g., quality circles), or clinical (e.g., addiction and rehabilitation programs)

- Organizational diagnoses to support decisions on the allocation of organizational resources (e.g., personnel, technology, and time)

The use of OD to provide decision makers outside the organization with important information is relatively rare. The following functions are the focus of OD for external decision makers:

- Organizational diagnostic data as the basis for investment decisions: There are no binding obligations for businesses or nonprofit organizations to provide external investors or controlling bodies—shareholders, governments, public advisory councils, or regulatory agencies—with decision-making information on the organization's state with regard to work and organizational psychological issues. However, findings of OD, such as information on qualification and knowledge, mobility, indicators of the quality of work life (e.g., satisfaction, commitment, and trust), can be highly relevant for decisions on technology, mergers, joint ventures, and strategic alliances.

- OD as a basis of a reporting system monitoring working conditions, health and safety, quality of work life, job attitudes, etc. to evaluate compliance with the criteria of humane work design: Similar to data on working accidents, absenteeism, or layoffs, these data could be considered in policy decisions regarding laws and regulations (e.g., tax, work, and social laws) and research programs. This application of OD will become more important as globalization and unification among countries (e.g., the European Union) increase.

The functions of OD for external stakeholders impose special requirements on the quality of data. This necessitates the authorization of independent external institutions, similar to auditing firms, to conduct OD because conflicting intraorganizational political interests make the validity of self-assessments questionable. Additionally, agreement on the applicable instruments is needed to allow for interorganizational comparability. These prerequisites illustrate that, to date, the application of OD in this context is more theory than practice.

The third domain for tasks and applications of OD is organizational research. OD can contribute to the development and testing of organizational theories. One criticism is the methodological one-sidedness of many organizational diagnostic approaches, primarily regarding measurement, which limits the comparability

of results and inhibits the advancement of organizational theory. To account for the diversity of organizational reality, Lawler *et al.* demand a renunciation of single methods in favor of multiple methods.

### 3. METHODS

OD incorporates methods of data gathering, data processing and analysis, formation of diagnostic judgment, and appraisal (i.e., the evaluation and presentation of results). In the following sections, the emphasis is on practical applications.

#### 3.1. Instruments and Quality Criteria

Instruments from different disciplines, predominantly work and organizational psychology, economics, management, and organizational sociology, are used in OD. The choice of instruments largely depends on which aspects of experience, behavior, performance, and interaction of organizational members are relevant with regard to the focus and specific goals of the OD. Whereas, for example, for the diagnosis of organizational culture, the assessment of values and normative beliefs of organizational members is of particular relevance, for an OD focusing on quality, the perception and evaluation of internal and external customers are of vital importance. There are no comprehensive reviews on the instruments from the different disciplines; instruments are available from individual publications or books on certain groups of instruments. Table I lists seven ways of gathering data, for most of which a large number of instruments exists.

Frequently used are the analysis of documents and statistics; inquiries of organizational key persons, experts, and staff members; group discussions; and observations on the job of meetings, etc. Less common is the analysis of interactions in organizations by sociometry, network analysis, Bales' Interaction Process Analysis, or Kelly's relatively work-intensive Repertory Grid Technique. The choice of instruments from the variety of sometimes competing measures should be based on three aspects: focus (mainly if structures or processes are assessed), goal and purpose of the OD, and methodological strengths and weaknesses of the measures.

In practice, different methods and instruments are frequently combined, and often multiple methods are a response to specific strengths and weaknesses of the

TABLE I  
Data Sources in Organizational Diagnosis

---

Analysis of documents: organization charts, records, regulations, job instructions, job descriptions, organizational handbooks, special handbooks (e.g., for malfunctions and irregularities)
Organizational and economic statistics: general human resource data, absenteeism, working accidents, fluctuation, and production and investment data
Inquiries of key persons and experts: oral or written questioning of persons who are assumed to have implemented and/or control organizational regulations or who are involved in the implementation and/or control in certain subareas
Inquiries of staff members: oral or, more often, written questioning of organizational members on different levels
Observations on the job: meetings, etc.
Group discussions
Analysis of interactions: sociometry, network analysis, and interaction-process analysis

---

instruments. A step-by-step organizational diagnostic approach includes the use of nonstandardized and standardized methods in different stages. Initially, observations and interviews are conducted, mainly in unfamiliar, complex organizational areas. Next, processes and relationships are assessed primarily through standardized methods, such as questionnaires, guided observation, and archival analyses.

The quality criteria of psychological test theory represent a fundamental issue regarding the strengths and weaknesses of organizational diagnostic instruments. Because perfect measurement typically is not possible, different kinds of measurement errors, depending on the design of the instrument, are caused by insufficient objectivity (objectivity in use, analysis, and interpretation), reliability (estimated as test-retest, parallel, split-half reliability, or internal consistency), or validity (estimated as construct or criterion validity).

In addition to the essential criteria of objectivity, reliability, and validity and their necessary standards for conducting empirical research, practical and economic requirements have to be considered in the choice of instruments. Organizational diagnoses necessitate various kinds of investments. Demands for time and manpower, structure, semantic and visual comprehensibility of manuals and instructions, and simplicity of use and interpretation are important determinants of investments and have to be taken into account in the decision to use an instrument.

Despite methodological concerns, a goal conflict, which is inherent to an applied discipline such as OD and consistently has to be solved anew, remains between methodological–scientific demands and the pragmatic–technical requirements of practice. If organizational goals are achieved, despite a weak foundation of diagnostic measures and interventions, theoretical and methodological considerations sometimes may appear secondary.

### 3.2. Data Analysis

The analysis of data is an integral part of the organizational diagnostic process. Analytic techniques of empirical social research and applied statistics and specific techniques of individual psychometric diagnostics or process diagnostics are used. The former include descriptive statistics and statistical inference, such as analysis of variance to conduct comparisons of organizational groups, cluster analysis to generate diagnostic types, and discriminant analysis to generate and apply multivariate statistical selection strategies. For process diagnostic assessments of changes, analyses of variance with multiple measurement points or, in the case of many measurement points, more appropriately time series analyses can be conducted. Although a wide range of statistical methods exist for analyzing data from standardized OD, there is a lack of work on strategies for analyzing qualitative organizational diagnostic data, be it content analysis or interpretative methods. In this case, the diagnostician is referred to the literature on qualitative empirical social research.

### 3.3. Conducting Organizational Diagnosis

Conducting OD involves the phases of introduction, exploration, planning, main investigation, data processing and analysis, interpretation, summary, and presentation of results. In the introduction phase, diagnosticians develop a preliminary understanding of the relevant variables and dependencies, usually through talks and negotiations with clients, management, and employee representatives. Important variables are goals, potential benefits of results, available resources, and possibilities for participation of organizational members. Additional information on methods and anticipated problems of the OD is exchanged. A clarification of interests in this phase is essential for the success of the

OD. An independent diagnostician has to take great care to avoid becoming biased toward one of the organizational groups (employee representatives and management). Both groups have to be completely informed during all stages. This also serves to prevent rumors and prejudices, which can result in misinterpretations of the OD as being part of a restructuring or downsizing program, which in turn can lead to low response rates, systematic biases, and the failure of measures.

In the exploratory phase, reconnaissance studies are conducted using largely nonstandardized interviewing methods, group discussions, and observations to consider and question issues and approaches and to prevent imbalances or restrictions in the diagnostic perspective. An evaluation of the sampling criteria for organizational units may also be indicated in this phase. Theoretical, methodological, and practical training may be necessary to align existing and required competencies and qualifications of investigators who are not experienced in conducting OD (e.g., freelancers and students).

The planning phase comprises the compilation, adaptation, and preliminary testing of the instruments, based on a model of the object of investigation, and considerations on their appropriateness regarding data processing and analysis. Additionally, various administrative issues of the investigation should be resolved in this phase (e.g., time, sampling, duration, location, information of participants, administration of instruments, and confidentiality of data).

In addition to these methodological, administrative, and concrete practical preparations, this phase can contain two higher level aspects of planning. First, cost–benefit analyses can be conducted for the purpose of the investigation and to help determine the diagnostic method (e.g., regarding requirements of time, personnel, materials, and sample size). Second, a meta-analysis of the scientific literature on methods and data collection can be performed if the reviews on certain issues of OD do not suffice for a reliable evaluation of the state of method development. The economy of the investigation should also be considered in the planning phase. Internet and corporate intranet offer good opportunities for its improvement. Currently, several sophisticated software packages for the social sciences facilitate online surveys and automated data processing (e.g., optical data scanning).

The planning phase takes on a central role in the OD. Not only does it set the course for the choice of methods but also a multitude of decisions are made during this phase that determine the organizational diagnostic process. The more accurate the planning,

the less fundamental problems will occur during the actual investigation. Despite careful planning, however, problems arise, at least in longer term organizational diagnostic projects. The four main problem areas involve the investigators conducting the OD, project controlling, changes in the examined organization, and unpredictable problems.

The lack of qualified personnel can cause staffing problems to become a crucial burden for any OD. Good leadership and human resource management are therefore imperative for success. Larger, longer term organizational diagnostic projects cannot be realized without regular feedback on compliance with schedule, budget, and quality standards. A systematic controlling of these aspects has to be incorporated into any project from the outset. Controlling time is usually less problematic than controlling costs and quality. This is caused, for instance, by the incalculable financial reserves needed to cover staffing gaps due to sickness or quitting of diagnosticians or the difficulties of controlling the quality of interviews, observations, etc. Changes in the diagnosed organization constitute a third problem area. Changes that interfere with the OD occur frequently in organizations but are difficult or impossible to predict (e.g., discontinuing of units due to market changes, changes in management, or unforeseen reactions of organizational members to the OD). Finally, there is always a residual category of unpredictable problems, ranging from unexpected method effects to the loss of data due to technical problems.

In the data-processing phase, compliance with predefined procedures for sampling and investigating and quality of data have to be assessed. This includes problems with methods (e.g., systematic nonresponse), data processing (e.g., errors in coding and screen forms for data input), or data management (e.g., confounding of variables and wrong recoding of items). Ensuring confidentiality of data is a major issue in the data-processing phase.

The actual phase of data analysis, carried out with statistical computer applications, should be preplanned, well specified, and hypotheses based, especially for larger data sets with numerous variables.

In the subsequent interpretation phase, results are evaluated with respect to the goal of the OD. Depending on the goal, results are used for inter- or intraorganizational comparisons (e.g., comparisons between units or divisions). Furthermore, results are interpreted with regard to hypotheses on relations (e.g., the perception of job insecurity is related to staff changes and a lack of cooperation/communication

with the supervisor) and assumptions about causes and consequences of symptoms and interrelations (e.g., the perception of job insecurity results in decreased flexibility and increased absenteeism).

The OD concludes with the summary, report, and oral presentation of results for clients and interested organizational groups. Summary, report, and presentation significantly determine the organization's ability and willingness to accept and implement the results. Considering their importance for the success of the OD, these aspects often do not receive sufficient attention.

Active involvement of organizational members should be considered for every stage of OD, not just a subsequent phase of organizational design or development. This participation of affected individuals refers to a direct involvement in decision-making processes, including participation in the introduction, planning, and interpretation phases. Participation not only realizes aspects of democratizing and humanizing work life but also can represent a valuable support for the ecological validity of the OD. After all, who could give a better account of practices, processes, etc. than the persons involved, especially at the unit and work group level?

#### **4. AN EXAMPLE: ORGANIZATIONAL DIAGNOSIS OF CLIMATE AND CULTURE**

As nonstructural coordination mechanisms, organizational culture and organizational climate play a special role in organizational psychology. Both culture and climate are theoretical constructs that possess observable correlatives but are not directly observable. Observable cultural correlates are the use of language (communication culture), work and production means (technological culture), goals and mission statements (progress culture), rituals (traditional culture), safety standards (safety culture), etc. As a fundamental part of the organizational identity, organizational culture forms the character of the organization and its members both by attracting and assimilating members and by excluding or selecting them.

Although knowledge on the cultural and climatic situation of an organization attains strategic importance, culture and climate as constructs are not easily accessible for OD. Additional complications arise because a variety of models and definition of culture and climate in organizations exist and because culture and climate are not unequivocally distinguishable. Culture and

climate have in common the fact that they both refer to organizational characteristics that are based on communication and cooperation. However, whereas organizational culture refers to characteristics that are taken for granted, unquestioned, implicit, and customary (e.g., norms, symbols, and rituals), aspects of organizational climate are more consciously perceived, reflected, and judged in their connection with changing conditions.

For the diagnosis of organizational culture, standardized questionnaires assessing traits of organizations and organizational units (group and team) exist. According to Neuberger and Kompa, instruments for the diagnosis of organizational culture can be classified by method (quantitative vs qualitative) and by the relevant aspects of culture (sociocultural configuration vs latent factors). Consequently, four classes of instruments can be distinguished: interviews with key persons using, for example, questionnaires on leadership behavior (quantitative, sociocultural configuration); structured questionnaires assessing norms, values, etc. (quantitative, latent factors); structured questionnaires assessing stories, jokes, slogans, etc. on customs, rituals, and traditions in organizations (qualitative, sociocultural configuration); and spontaneous self-interpretations of meaning and function of cultural practices (qualitative, latent factors). This classification illustrates the different methodological approaches of OD for the example of organizational culture.

In 1990, Rousseau reviewed several quantitative measures for assessing organizational culture, including the Organizational Culture Inventory by Cooke and Lafferty, which is a widely used instrument.

The Organizational Culture Inventory assesses normative beliefs regarding the extent to which specific behaviors help members fit in and meet the expectations of the organization. Twelve cultural styles, each assessed by 10 self-report items using a 5-point Likert scale, are distinguished and arrayed in the underlying circumplex model by their orientation on task vs people and security vs satisfaction:

#### Cultural styles and sample items

1. Humanistic–helpful (e.g., helping others to grow and develop)
2. Affiliative (e.g., sharing feelings and thoughts)
3. Approval (e.g., “going along” with others)
4. Conventional (e.g., always following policies and practices)
5. Dependent (e.g., pleasing those in positions of authority)
6. Avoidance (e.g., waiting for others to act first)
7. Oppositional (e.g., pointing out flaws)
8. Power (e.g., motivating others any way necessary)
9. Competitive (e.g., turning the job into a contest)
10. Competence/perfectionistic (e.g., doing things perfectly)
11. Achievement (e.g., pursuing a standard of excellence)
12. Self-actualization (e.g., thinking in unique and independent ways)

The instrument has been applied in a variety of settings and cultural contexts. In a study of fund-raising organizations, it was shown that team-oriented cultural styles (achievement, self-expression, humanistic–helpful, and affiliative) were positively related to desirable attitudes (e.g., role clarity, satisfaction, and propensity to stay), whereas security-oriented beliefs (approval, conventional, dependent, avoidance, oppositional, power competitive, and perfectionistic) were negatively related to these attitudes and to organizational fund-raising success.

OD is not an end in itself, and this holds true for the diagnosis of culture and climate. The interest in reliable knowledge on culture and climate is reinforced by the expectation that a positive climate and culture support organizational goals, that their implications can be actively shaped, and that their benefits are significant. Attitudes, motivation, and values of organizational members are affected by culture and climate. For example, the relation of a positive communication culture and a climate of trust with high job satisfaction has been empirically confirmed.

## 5. CONCLUSION

Due to rapid sociotechnical changes in organizations, assessing the state before an innovation, controlling its implementation, and evaluating its success comprise an important function of OD. Changes in experience and behavior of individuals, due to technological innovations, impose demands on OD that are not consistently met.

The diversity of sociotechnical processes involves the extent to which different organizational levels are affected by a change in the system (from the macro-organization to the individual) and the nature and quality of changes (from changes in organizational structure and technological processes to changes in the possibilities for interaction and communication on the job). This diversity represents challenges for

OD and necessitates the development of integrative approaches, incorporating multilevel theory and methods, and broadening the perspectives adopted in the investigation of sociotechnical processes.

The considerable practical benefits of OD have not received sufficient attention. With the notable exception of the seminal work of Van de Ven and Ferry, examples of methodologically sound and comprehensive OD projects are rare. This, however, may change if the practitioners in organizations and consulting firms become familiar with the possibilities and the cost-benefit relation of OD. To achieve this, the comprehensiveness of OD and its necessity and position in the process of organizational design and development have to be emphasized and illustrated by examples. The misperception of OD as an interesting methodological approach that is dispensable for the solution of time-sensitive practical problems can be easily disproven with concrete practical results.

In addition to the various general and specific benefits derived from information and knowledge on the state of organizations, an important argument for the value of a psychologically based OD for businesses, managers, and consultants is the financial aspect. Similar to other areas with regard to work and organizations (e.g., worker safety and health protection), cost reductions attributable to OD are hardly calculable in detail. The profitability of OD cannot be appropriately evaluated by the traditional economic calculation of the costs of isolated measures but, rather, includes the optimization and integration of the whole system with regard to functionality, safety, etc. Only a comprehensive economic resource accounting, including a human capital assessment and using data from controlling and comparisons of costs and benefits, has the potential to generate a sustainable effect by constantly providing feedback on the results of OD and thus stimulating and reinforcing the motivation of decision makers, managers, consultants, and other involved groups (e.g., work councils). A comprehensive economic evaluation can be provided by extended accounting systems, which incorporate both monetary and nonmonetary goals (i.e., goals of the company and goals of the employees) and capital value and human capital reflected, for example, in the level of qualification and motivation.

### Acknowledgments

This article is adapted from Büssing (2003). I thank Severin Hornung for help preparing this article.

### See Also the Following Articles

Industrial/Organizational Psychology across Cultures  
 ■ Industrial/Organizational Psychology, Overview  
 ■ Organizational Culture and Climate ■ Organizational Structure

### Further Reading

- Anastasi, A., & Urbina, S. (1996). *Psychological testing*. Englewood Cliffs, NJ: Prentice Hall.
- Büssing, A. (2003). Organisationsdiagnose [Organizational diagnosis]. In H. Schuler (Ed.), *Lehrbuch der Organisationspsychologie* [Textbook of organizational psychology] (3rd ed., pp. 445–480). Bern: Huber.
- Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1993). *The experience of work. A compendium and review of 249 measures and their use*. London: Academic Press.
- Cooke, R. A., & Lafferty, J. C. (1989). *Organizational culture inventory*. Plymouth, MI: Human Synergistics.
- Cooke, R. A., & Rousseau, D. M. (1988). Behavioral norms and expectations: A quantitative approach to the assessment of culture. *Group and Organizational Studies*, 13, 245–273.
- DiPofi, J. A. (2002). Organizational diagnostics: Integrating qualitative and quantitative methodology. *Journal of Organizational Change Management*, 15, 156–168.
- Harrison, M. I., & Shirom, A. (1999). *Organizational diagnosis and assessment: Bridging theory and practice*. Thousand Oaks, CA: Sage.
- Hays, W. L. (1994). *Statistics*. Cambridge, MA: Wadsworth.
- Howard, A., et al. (Ed.). (1994). *Diagnosing for organizational change: Methods and models*. New York: Guilford.
- Klein, K. J., & Kozlowski, S. W. J. (Eds.). (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco: Jossey-Bass.
- Lawler, E. E., Nadler, D. A., & Cammann, C. (1980). Uses of organizational assessment data. In E. E. Lawler, D. A. Nadler, & C. Cammann (Eds.), *Organizational assessment: Perspectives on the measurement of organizational behavior and the quality of work life* (pp. 25–118). New York: Wiley.
- Miles, M. B., & Huberman, M. A. (1994). *Qualitative data analysis—An expanded sourcebook*. London: Sage.
- Neuberger, O., & Kompa, A. (1993). *Wir, die Firma. Der Kult um die Unternehmens-kultur* [We, the company. The cult about the organizational culture]. Munich: Heyne.
- Rousseau, D. M. (1990). Assessing organizational culture: The case for multiple methods. In B. Schneider (Ed.), *Frontiers of industrial and organizational psychology* (Vol. 3, pp. 153–192). San Francisco: Jossey-Bass.
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Van de Ven, A. H., & Ferry, D. L. (1980). *Measuring and assessing organizations*. New York: Wiley.



# Organizational Justice

Paul E. Levy and Christina Norris-Watts

University of Akron, Akron, Ohio, USA

1. What Is Organizational Justice?
  2. History of Justice in Organizations
  3. Theoretical Conceptualizations of Justice
  4. Antecedents of Organizational Justice
  5. Consequences of Organizational Justice
  6. Conclusion
- Further Reading

## GLOSSARY

*distributive justice* The perception that what one receives in terms of outcomes is fair.

*fairness heuristic theory* An approach to making justice judgments that relies on cognitive shortcuts as guides in decision-making behavior.

*fairness theory* An approach to making justice judgments that considers whether other alternatives could and should have been used as well as the extent to which they would have led to more favorable outcomes (an extension of the earlier referent cognitions theory).

*instrumental model* A perspective arguing that justice is important to individuals because the voice provided them allows them some control over outcomes.

*interactional justice* The perception that the interpersonal treatment that is received from one's coworker, supervisor, etc. is fair.

*organizational justice* Individual's perceptions of the general level of fairness in a specific organizational context.

*procedural justice* The perception that the procedures or processes used in a particular situation such as the workplace are fair.

*relational model* A perspective arguing that justice is important to individuals because being treated fairly suggests that one is an important individual to others.

Organizational justice refers to perceptions of fairness in an organizational context. Justice research focuses on theoretical conceptualizations of justice, how justice decisions are made, and the dimensionality of justice, as well as its antecedents and consequences. Fair treatment is highly valued in the 21st century workplace, where justice perceptions tend to impact on employee attitudes and behaviors as well as the effectiveness of organizations.

## 1. WHAT IS ORGANIZATIONAL JUSTICE?

Organizational justice is the study of people's perceptions of fairness in organizational contexts. The term was coined by Jerry Greenberg in the 1980s to describe individuals' interest in and concern with fairness-related activities taking place in various organizations such as one's workplace. Organizational justice can be viewed as a class of motivated behavior that is engaged in by different individual and contextual characteristics. Perceptions of organizational justice then result in cognitive and affective responses that guide ensuing behaviors. It did not take long for the concept to capture the attention of both researchers and practitioners, and this interest has resulted in numerous books,

journal articles, and conferences that have been devoted to organizational justice. In addition, when consultants talk to managers, employees, and CEOs about the “hot” issues confronting their organizations and the people of those organizations, the consultants are deluged with examples, anecdotes, and stories that revolve around organizational justice and its impact on the organization and its members.

## 2. HISTORY OF JUSTICE IN ORGANIZATIONS

Adams' work on equity theory is the foundation for psychological work on fairness. The focus of this work was on the fairness of outcomes, or what has become known as distributive justice. Organizational researchers recognized the potential effect of distributive justice on organizational functioning and, thus, it has been a major focus of organization-based research since the 1980s. However, as it became clearer that employees' reactions to organizational events were not based solely on their perceived notions of outcome fairness or distributive justice, the direction of research followed from work in social psychology and process-related issues began to be examined. Thus, procedural justice, or the fairness of the processes used to determine outcomes, became a major research focus and this has continued into the 21st century. Procedural justice is enhanced by providing employees with opportunities for voice or input into the process and by following fair and consistent rules. In fact, Leventhal suggested a set of rules to follow to ensure procedural justice that included accuracy, ethicality, and representativeness.

In the mid- to late 1980s, Bies suggested a third dimension of organizational justice that he called interactional justice. Interactional justice is believed to be related to procedural justice, but in particular it describes how decisions are communicated to employees and focuses on interpersonal treatment. That is, whereas procedural justice focuses on the fairness of organizational procedures involved in decisions, interactional justice deals more with how these decisions are conveyed to employees.

There continues to be disagreements among researchers as to the number of independent dimensions of justice. Some believe only in the traditional distinctions between procedural and distributive, where interactional justice is subsumed under procedural justice. Others maintain interactional justice as a third independent dimension of

justice. It has also been suggested that interactional justice is composed of two dimensions: (i) interpersonal justice, which focuses on treating individuals with respect, dignity, and courtesy, and (ii) informational justice, which focuses on the explanations provided to individuals about the decision-making process.

## 3. THEORETICAL CONCEPTUALIZATIONS OF JUSTICE

Justice research has been driven by two main questions. First, why is justice important? Research has investigated the extent to which justice is critical to employees and why they value justice. Second, how do employees decide what is just? Researchers have developed theoretical models for examining and understanding this judgment process.

### 3.1. The Importance of Justice

Research has demonstrated that individuals value justice and behave, in part, based on their perceptions of fair treatment. One approach explaining why individuals value justice takes an instrumental perspective, arguing that individuals desire control over events and situations. Individuals perceive that they have control when they are provided the opportunity to voice their opinions, beliefs, and ideas. Ultimately, this instrumental model suggests that voice is used to exercise some control over situations, increasing the likelihood of favorable outcomes. A slightly different perspective is the relational model (formerly called the group-value model), which argues that individuals desire favorable relationships with others. Being treated fairly provides individuals with evidence or proof that they are valued by others, increasing their own sense of self-worth.

A recent perspective takes the position that justice is its own reward. The moral virtues model argues that justice is important to individuals based on an innate sense of right or wrong. In other words, individuals have a fundamental respect for the worth of human beings and, therefore, expect fair behaviors that are consistent with this respect. Finally, Cropanzano and colleagues attempted to integrate these various perspectives into what they call the multiple needs model, in which basic desires for control, belonging, positive self-regard, and meaning motivate a desire for justice.

### 3.2. Deciding What Is Just

Two major theories have been developed that purport to explain how individuals make justice decisions. Applied to organizations, these theories describe two different processes through which employees might evaluate work situations and draw conclusions about the fairness of treatment received. Referent cognitions theory, developed by Folger in the 1980s, posits that situations are perceived to be unfair when individuals can think of more favorable outcomes that *would* have occurred from alternative procedures that *could* have been used. If individuals cannot generate alternative procedures that would have led to more favorable outcomes, the situation is perceived as fair. In Folger's revision of this theory, fairness theory, a third requirement was added. Along with the *would* and *could* elements, he proposed that situations are only perceived as unfair if an alternative procedure *should* have been used according to a moral imperative. In sum, injustice is perceived when (i) unfavorable outcomes are experienced, (ii) alternative procedures could have been used, (iii) the alternative procedures would have led to more favorable outcomes, and (iv) these alternative procedures should have been used based on some ethical rationale.

Lind developed another perspective on the justice decision-making process in the 1990s that he termed fairness heuristic theory. In general, heuristics are shortcuts used to make decisions that free up cognitive resources for other purposes. With respect to justice decisions, Lind argues that individuals make fairness judgments and then use those judgments to regulate behavior. Research has shown that individuals' behaviors are more driven by early fairness information than later fairness information. In other words, once individuals make a fairness judgment based on available information, that judgment demonstrates stable properties and guides ensuing behavior.

## 4. ANTECEDENTS OF ORGANIZATIONAL JUSTICE

Now that justice has been defined and the prominent theories reviewed, it is necessary to consider variables that influence justice perceptions. These antecedents can be classified into three groups: individual differences, situational characteristics, and supervisory factors.

### 4.1. Individual Differences

Individual differences such as demographic variables, affectivity, and self-esteem have been investigated as potential antecedents of organizational justice perceptions. First, demographic variables such as age, gender, race, education, and organizational tenure seem largely unrelated to justice perceptions. This suggests that no particular category of individuals are more inclined to perceive injustice than any other category. Second, studies have found negative affectivity to be related to justice perceptions. Negative affectivity can be thought of as a personality disposition descriptive of individuals who have a generally pessimistic worldview. The findings suggest that those who have more negative affectivity tend to perceive more injustice in their organizations. In particular, this relationship is stronger for procedural and interactional justice than it is for distributive justice. Third, employees who have high self-esteem, manifested by self-confidence and positive perceptions of self-worth, tend to perceive more procedural justice in their organizations than do those who have low self-esteem.

### 4.2. Situational Characteristics

Organizational policies, procedures, and practices have been examined as predictors of justice perceptions. For instance, strong relationships have been demonstrated between the extent to which employees are provided with opportunities for voice and perceptions of procedural and distributive justice. This is consistent with the previous discussion about why justice is important to employees: Having input leads to perceptions of control, which is valued by the employee and results in impressions of justice. The degree to which organizations provide support and high-quality communication to their employees has also been linked to justice perceptions. For instance, employees who perceive more organizational support also view their organizations as being more procedurally just. Finally, one of the most consistent findings in the literature is that individuals react more strongly to unfavorable outcomes if those outcomes are accompanied by unfair procedures. In other words, fair procedures can partially overcome the negative effects that are usually associated with unfavorable outcomes. Individuals who experience unfavorable outcomes but who view the procedures as fair react more positively than do those who receive similar outcomes and view the procedures as unfair.

### 4.3. Supervisory Factors

The social context plays an intricate part in organizational functioning and it has been examined in terms of justice perceptions as well. In particular, leadership variables have been extensively studied with respect to justice perceptions. The relationship between leaders and subordinates has been the focus of a great deal of research in industrial/organizational psychology. This relationship is often operationalized as an exchange between leaders and subordinates. Leader-member exchange (LMX) theory has been developed to describe this social interaction. Research has found strong relationships between leader-member exchange and justice. In particular, more favorable leader-subordinate relationships are associated with higher perceptions of justice. Interestingly, consistent with the previous discussion of the interpersonal nature of interactional justice, the relationship between LMX and justice is strongest when justice is operationalized as interactional in nature.

In addition, transformational leadership—a leadership style in which the leader uses consideration and charisma to inspire employees—has been positively linked to procedural justice perceptions. Furthermore, transactional leadership—a leadership style in which the relationship between leader and supervisor is based on simple economic exchanges (i.e., meet these goals and receive this reward)—has been positively linked to distributive justice perceptions, as the theories would predict.

## 5. CONSEQUENCES OF ORGANIZATIONAL JUSTICE

One way to distinguish among the many consequences of organizational justice is to categorize them as attitudinal or behavioral in nature. When applied to organizational justice, this results in a variety of outcomes in each category.

### 5.1. Attitudes

One of the most studied variables in organizational psychology is job satisfaction or the emotional state resulting from the cognitive evaluation of what one's job provides to the employee in relation to one's expectations. It has been extensively studied with respect to justice, and strong consistent relationships have been demonstrated. The three major dimensions

of justice have been consistently linked, in a positive way, to job satisfaction such that employees who experience organizational justice also tend to report greater job satisfaction than those who do not experience organizational injustice. Furthermore, satisfaction with one's pay level has been strongly related to procedural justice, but as predicted by the theory, it has been even more strongly related to distributive justice. The most well-established conceptualization of organizational commitment is termed affective commitment, defined as the psychological attachment to the organization based on common goals and objectives. Research has found that the three dimensions of organizational justice are uniformly related to levels of affective commitment. Employees who perceive more fair experiences at work also tend to be more committed to the organization. Justice has also been found to be related in predictable ways to other attitudes, such as trust in the supervisor, trust in the organization, supervisor satisfaction, and management satisfaction.

Other attitudinal variables of interest focus on employee reactions to various organizational systems. First, considerable research has demonstrated that reactions to the performance appraisal process are affected by perceptions of organizational justice. Individuals perceiving the performance appraisal process as fair tend to be more satisfied with the appraisal system and session, view the appraisal as more accurate, view the appraisal as more useful, and are more motivated to use it for developmental purposes. Second, considerable research has investigated the relationship between justice perceptions about the employee selection system and reactions to that system. These studies have found that both rejected and selected applicants respond more favorably to selection systems that they perceive as fair and job relevant. These favorable applicant reactions include the tendency to recommend the company to other potential employees, satisfaction with the application experience, and work self-efficacy.

### 5.2. Behaviors

Research suggests that justice perceptions predict employee performance. In particular, procedural justice perceptions have been strongly related to work performance such that employees experiencing fair work procedures are better performers than those experiencing unfair work procedures. Performance has been expanded to include contextual performance, which consists of those behaviors focused on maintaining the social core of the work environment rather than

those task-related behaviors that make up the technical core of one's job. In other words, contextual performance maintains and supports the work environment but is not part of the main tasks of the job. Research has found consistent, significant relationships between justice perceptions and contextual performance, such as altruism, good sportsmanship, loyalty, and courtesy, but these relationships tend to be small.

Counterproductive behaviors in organizations include employee deviance, such as stealing, sabotage, and both physical and psychological aggression. In general, both distributive and procedural injustice have been shown to predict these behaviors, but interactional justice has yet to receive much research attention with regard to counterproductive behaviors. Employees experiencing unjust outcomes or procedures have been more apt to exhibit counterproductive behaviors than have those experiencing fair outcomes or procedures.

## 6. CONCLUSION

A plethora of organizational justice research has been conducted, mostly in the past 15 years. Much of what goes on in organizations is affected by perceptions of organizational justice. Employees and supervisors alike make judgments about what is fair and unfair in their organizations. These judgments, in part, affect

individuals' attitudes and behaviors and the interactions among individuals. Organizational functioning is indirectly affected via the processes involved in justice judgments and the ensuing behaviors. Be it in performance appraisal, selection, compensation systems, organizational change interventions, or the legal realities of organizational life, justice perceptions will continue to play an important role.

## See Also the Following Articles

Leadership and Culture ■ Power, Authority, and Leadership

## Further Reading

- Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 86(2), 278–321.
- Colquitt, J., Conlon, D. E., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86(3), 425–445.
- Greenberg, J., & Cropanzano, R. (Eds.). (2001). *Advances in organizational justice*. Stanford, CA: Stanford University Press.
- Subich, L. M., & Levy, P. E. (Eds.). (2001). Organizational justice: Implications for theory, research, and practice [Special Issue]. *Journal of Vocational Behavior*, 58, 163–278.



# Organizational Participation

Frank Heller

Tavistock Institute, London, United Kingdom

1. Introduction
  2. Conceptual Schemes
  3. Background
  4. Examination of the Evidence
  5. Conclusion
- Further Reading

paper. This results in joint consensual responsibility. The process is slow, but implementation is quick.

**semiautonomy** Derived from research in the Tavistock Institute, it represents a delegation of a degree of influence and responsibility to an individual or group for a defined task. The definition of the prefix “semi” is critical.

## GLOSSARY

**codetermination** Describes the German system, started by law in 1951 in the coal, iron, and steel industry, that gives workers 50% representation at the boardroom level. This was later extended, with lower worker representation, to other industries. Variations of this boardroom representation have been adopted in other European countries.

**human resource management (HRM)** A broader assessment of the role of employees than personnel management. The term indicates that employees are a resource (rather than a cost) in the way finance and capital are resources.

**quality circle** Japanese method of delegating quality responsibility to groups. It usually includes an intensive process of training. It has been adapted in various ways in other countries.

**participative dialogue** Derived from ideas developed by the German philosopher Jurgen Habermas and used particularly in Sweden to involve as many people as possible in organizational problem-solving processes.

**ringo seido** Japanese bottom-up decision making. The “ring” process starts at the lowest appropriate level and gradually moves up, with each person putting his or her “seal” on the

Organizational participation (OP) is about sharing influence and power in work organizations. This area of research has become important since World War II and has been called a number of names, such as employee involvement, high involvement management, participative decision making, partnership, joint consultation, industrial democracy, codetermination, and empowerment. Participation is still the most frequently used term and is often treated as a continuum. For instance, the research program at the University of Michigan used the terms system 1 (exploitative authoritative), system 2 (benevolent authoritative), system 3 (consultative), and system 4 (participative). In later research, particularly in Europe, the continuum was extended to include semiautonomy and delegation or complete control. In this extension, the term influence–power–continuum is sometimes used based on the assumption that participation can lead to degrees of autonomy or control and delegation of power in specified circumstances. There is a link with leadership and the decision-making process from the lowest level of organization to the boardroom and beyond to the national legal framework within which OP operates.

There is also a relationship between OP and the literature on industrial relations and organizational change. Collective bargaining, although part of the influence and power-sharing approach, is usually excluded from OP, as are cooperatives and share ownership; this article adheres to this tradition.

## 1. INTRODUCTION

Organizational participation (OP) can be treated from the micro perspective of the individual, the meso level of the organization, and the macro level of national policy. There are substantial differences between countries. The United States and Britain have voluntary schemes based on companies and sometimes on sections or divisions of companies. Most continental European countries have a legal underpinning for OP. For instance, the German codetermination scheme gives employees up to 50% representation at the supervisory board of their two-tier board structure. There are also elected representatives on works councils with legal rights to be consulted in defined circumstances. Similar legislation exists in most continental and Scandinavian countries.

However, formal representative participation does not require legislation. In Britain and the United States, such schemes are set up voluntarily by many companies sometimes related to collective bargaining. Formal schemes, sometimes called joint consultation or employee involvement, become part of company policy and frequently have a written constitution and elected representatives. Informal participation is quite different and can take many forms. It may be encouraged by top management, but it is not usually accepted boardroom policy. It is normally left to individual managers, departments, or divisions and can vary from ad hoc meetings to regular forums. Informal participation is an aspect of managerial leadership and changes as managers leave or retire. A particular department may acquire a tradition of OP, which then lasts beyond a particular manager's style of operating. In voluntary formal and informal OP, the content of deliberation and the amount of influence available to different groups vary considerably.

In Japan, tradition and culture, rather than legislation, have produced formally sanctioned schemes of bottom-

up management (*ringi seido*), the extensive use of quality circles, and the custom of lifelong employment in large companies. Some of these Japanese OP and human resources schemes have begun to weaken.

## 2. CONCEPTUAL SCHEMES

The goal of OP, whether formal or informal, is to achieve some organizational benefit, such as a more content workforce, more flexible work arrangements, less resistance to change, and increased productivity. Two theoretical models have emerged. First, the human relations model postulates a causal relationship between participation, work satisfaction and lower resistance to change, and increased organizational performance (productivity) (Fig. 1). As discussed later, evidence for this model is mixed. This model also lends itself to inauthenticity or pseudo-participation, often mentioned in the literature, since it may produce "a feeling of participation" through job satisfaction, without any real influence.

The second model is based on evidence that the positive outcome of OP seems to be due to the way in which participation allows existing experience and skill to be utilized. It is called the human resources model (Fig. 2). The assumed causality is participation → improved human resource utilization → improved organizational performance. Performance can take different forms, such as an increased propensity to innovate. The model does not exclude job satisfaction, although the direction of causality is not clear. The importance of this model is its emphasis on the utilization of competence, without which there can be no improvement of output. It follows that OP is not appropriate when employees do not have relevant experience and skill. It also virtually excludes inauthentic or pseudo-participation because the utilization of competence cannot be manipulated. The model suggests that any improvement in organizational performance is due to the utilization of existing experience and skill rather than to changes in attitudes.

Both models can be used in cross-sectional research with a single organizational level or department and therefore isolated from the total organization. Alternatively, they can be conceptualized as part of a systemic relationship within an integrated human

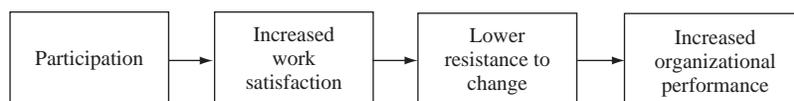
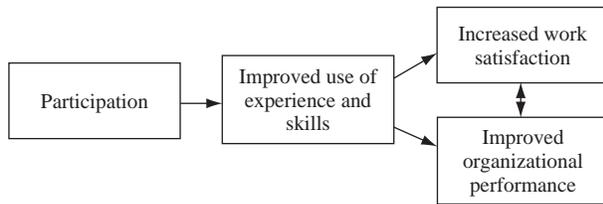


FIGURE 1 The human relations model. Possible feedback loops are not indicated.

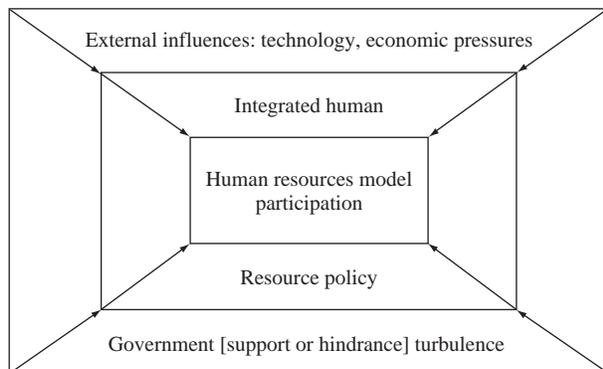


**FIGURE 2** The human resources model. Possible feedback loops are not indicated.

resources policy. Furthermore, it may be necessary to recognize the impact of the external environment within an open systems framework. The effectiveness of OP may be influenced by new technologies, external uncertainties, national policy, legal provisions, and macroeconomic events. An open systems framework may therefore be appropriate (Fig. 3).

### 3. BACKGROUND

As economic competition increased in the decades after World War II, OP was frequently advocated as a way to improve organizational efficiency. This justification for OP has continued. In the 1980s, when Japan had a very competitive economy, European and U.S. managers began using Japanese methods such as quality circles and similar employee involvement techniques. Toward the end of the 20th century, economic conditions changed and Japan started to dismantle some of its human resource policy, particularly lifetime employment security. Nevertheless, during these competitive changes in Japan and the rest of the world, some aspect of OP has remained an organizational objective in most countries.



**FIGURE 3** Open-system human resources model.

## 4. EXAMINATION OF THE EVIDENCE

### 4.1. Traditional Fieldwork

A considerable amount of the research literature concentrates on the lower levels of organization using a variety of quantitative methods and statistical designs. Much of this work, including meta-analyses, has been reviewed, for instance, by Locke and Schweiger, Cotton, and Strauss. The results do not present a uniform picture. OP frequently correlates with job satisfaction but much less often with improvements in organizational performance. Wagner reviewed 11 U.S. studies and concluded that even when there is a reliable statistical relationship between OP and performance, the correlations are so small that it is doubtful they have any practical significance and will tempt managers to adopt them on economic grounds alone. Other research, particularly if based on consultancy experience, such as that by Pfeffer, has found much more positive results; nevertheless, Pfeffer and others have been disappointed and surprised by the low uptake of OP work designs.

### 4.2. Cross-National Studies and Surveys

A number of large-scale cross-national studies on participative practices have been performed, especially in the United States and most European countries. Several studies cover the whole organization and use judgments from more than one level. Participative practices vary substantially between countries. One of the major reasons for this is the legal framework in which they are embedded (Fig. 3). When aspects of OP are mandatory, de facto influence sharing reaches the lowest levels of an organization.

The other consistent finding from large-scale national and cross-national studies is that the amount of distributed influence is very small. In a 12-country Industrial Democracy in Europe study, one of the scales asked people to describe their influence on a 5-point continuum from no influence (1) to very much influence (5). With regard to 14 decisions, the average low-level employee scored 2 (little influence). In relation to tactical and long-term decisions, their influence was much lower than 2.

### 4.3. Various Adaptations

The organizational function called personnel management has transmuted into human resource management

(HRM), which, as the name implies, gives greater weight to the role of human capital without diminishing the importance of economic and physical capital. Giving employees more voice plays a part in various HRM developments; for instance, semiautonomous work groups or teamwork have become popular and are often credited with improved efficiency as well as greater job satisfaction. As with other schemes, there are varying degrees of semiautonomy within teams. Sometimes, the power given to teams is minimal. In other cases, the team is able to reassign work sequences, layout, and task priorities and may be responsible for quality. Occasionally, more extensive power and responsibility are devolved to teams, such as choosing their own leader, setting targets and devising or distributing incentive schemes, agreeing on holiday rosters, job rotation, and job design, and communicating with management at a higher level. When teamwork is based on trust between management and employees and the group is assigned many of the tasks described previously, this is characteristic of a holistic system, which is discussed in the next section.

Quality circles have not always been successfully adapted to Western companies. They require skilled members, motivation, and incentives. In Japan, they were embedded in a nonadversarial industrial relations culture and could often call on the advice of outside experts.

The terms used to describe various HRM policies change frequently. Partnership is a term used to emphasize formal or informal influence-sharing arrangements between constituencies inside an organization or between an organization and an outside body (e.g., a private company and a public body). The development of virtual problem-solving organizations with people geographically dispersed, sometimes globally, has given rise to a form of interpersonal participation that can be said to create social capital.

As organizations and the environment in which they operate become increasingly complex, more emphasis is placed on a flexible workforce and the need for constant updating and learning. Among the characteristics required for an effective learning organization is a high level of motivation supported by a dispersal of influence to facilitate the acquisition of experience and skill. Equally important, but less frequently recognized, is the role of participation in reducing the underutilization of existing competence.

#### 4.4. Systems-Oriented Research

Systems-oriented organization theory has been advocated since at least the 1960s through the writing of

Katz and Kahn and the Tavistock team, among others. However, until recently, it has had little impact on HRM policy. Most research on OP is designed and assessed within one part of an organization. It might be favored by a particular manager or consultant and does not survive when the manager leaves the organization. Even if OP applies to a whole organization, it is usually isolated vis-à-vis other HRM schemes.

Given the inconclusive results from traditional isolated projects, based on the human relations model (Fig. 1), more holistic OP designs have received attention. They are more like a human resources model operating within a systems perspective (Figs. 2 and 3). Several facets of an organization are clearly interdependent and are treated as a system, such as accounting. Managers at all levels work within the accepted set of interrelated accounting procedures. There is evidence that a systems design of HRM that includes OP can yield positive results. There are packages of HRM schemes, sometimes called high-involvement management (HIM), that include participation and a bundle of measures, such as teamwork, compensation-incentives, training, quality circles, broad job description, and similar work practices. Research suggests that such interrelated HIM-OP measures create synergy, leading to increased effectiveness, efficiency, and/or productivity.

Another relevant experience derives from Norway and Sweden. A number of individual sociotechnically designed OP projects were started in the 1960s but failed to grow and did not spread to other organizations. Gradually, later projects became more holistically designed. Several large-scale regional projects in a variety of industries were started in the 1980s and continue today. They are supported by employers and unions and oriented toward participative dialogue among as many employees as possible. The aim is to achieve a participative-communicative infrastructure to facilitate a continuous process of adaptive changes and improvements in organizational effectiveness. This systemic Scandinavian approach has been compared to the Japanese total quality management, and both appear to have achieved a measure of success.

There are substantial opportunities for further research. For instance, one can hypothesize that holistically organized participative organizations would increase work motivation, reduce boredom, facilitate conflict resolution, and increase a sense of responsibility as well as compliance with safety measures. It is reasonable to predict that OP will continue to play a useful role in applied psychology and related social science areas.

## 5. CONCLUSION

Organizational participation has several names and operates in a variety of contexts. It can be formal (i.e., representative) or informal and ad hoc. It can be isolated from other managerial and HRM schemes, or it can be coupled with them in a systemic arrangement. Improvement of organizational performance is usually the objective. In the human resources model (Fig. 2), this is achieved because OP, where appropriate, uses experience and skills that in the absence of OP would remain dormant. Although the case for OP has been widely advocated in management programs and the literature, employment of OP has been moderate or low. There are several explanations for this contrast between espousal and implementation. It has been argued that OP is not independent of positive or negative environmental influence (Fig. 3), but research design rarely allows for this. Another view suggests that when OP is seen as empowerment of lower levels of organization, it can be assumed to be a threat to the “managerial prerogative.” This interpretation is contested by those who argue that participation in the sense of contributing experience and skill does not abrogate the final responsibility of senior management to integrate the liberated competence in an effective decision-making process.

## See Also the Following Articles

Competence at Work ■ Cooperation at Work ■ Decision Making ■ Industrial/Organizational Psychology, Overview ■ Organizational Culture and Climate ■ Organization Development ■ Organizational Socialization ■ Organizational Structure ■ Power, Authority, and Leadership

## Further Reading

- Cotton, J. (1993). *Employee involvement: Methods for improving performance and work attitudes*. London: Sage.
- Heller, F. (2003). Participation and power: A critical assessment. *Applied Psychology: An International Journal*, 52(1), 144–163.
- Heller, F., Pusic, E., Strauss, G., & Wilpert, B. (1998). *Organizational participation: Myth and reality*. Oxford, UK: Oxford University Press.
- Ichniowski, C., Kochan, T., Levine, D., Olson, C., & Strauss, G. (1996). What works at work: Overview and assessment. *Industrial Relations*, 35(3), 299–333.
- Industrial Democracy in Europe Group. (1993). *Industrial Democracy in Europe revisited*. Oxford, UK: Oxford University Press.
- Naschold, F., Cole, R., Gustavsen, B., & van Beinum, H. (1993). *Constructing the new industrial society* (Vol. 3). Assen/Maastricht, The Netherlands: Van Gorcum.
- Pfeffer, J., & Veiga, J. (1999). Putting people first for organizational success. *Academy of Management Executive*, 13(2), 37–48.



# Organizational Socialization

Talya N. Bauer

Portland State University, Portland, Oregon, USA

1. Introduction
  2. Stages of Socialization
  3. Common Outcomes of Socialization
  4. What Do Organizations Do to Facilitate Socialization?
  5. How Do Insiders Aid in Socialization?
  6. Newcomer Proaction
  7. Socialization Research Methods
  8. Conclusion
- Further Reading

*socialization tactics* The ways in which organizations train and orient new employees.

This article reviews what socialization is, why it matters, and thoughts on how to ensure its success. Relevant research findings are reported.

## GLOSSARY

*accommodation* The second stage of socialization, when newcomers learn how to fit into their new organization socially and task relatedly.

*anticipatory socialization* The first stage of socialization, when newcomers are thinking about joining the organization.

*entry* The point at which newcomers officially join the organization.

*information seeking* The process of new employees obtaining information, either overtly or covertly, about organizational life.

*insiders* Organizational members who are already established in the organization when a new employee joins.

*longitudinal research* Research that includes data collection at multiple points in time; this has typically occurred at 3, 6, 9, and/or 12 months postentry.

*newcomers* New employees who are undergoing the socialization process.

*role management* The third stage of socialization, when new employees become fully integrated into the organization.

## 1. INTRODUCTION

Organizational socialization refers to the “learning” of what it is to be an organizational insider. It has been referred to as the process of “learning the ropes” of organizational life. By definition, organizational socialization deals with changes for newcomers. Unsuccessful socialization means that newcomers have not learned to fit into the organization either socially or with regard to the work of the job. Lack of socialization can result in high turnover rates and employee dissatisfaction. On the other hand, successful socialization can help organizations to retain valuable employees and may result in considerable cost savings. When socialization is effective, newcomers understand and adopt the organization’s central values and norms. This guarantees better coordination among employees. Learning is key when looking at socialization. This article summarizes some of the important elements of organizational socialization and gives some practical implications for people in organizations.

## 2. STAGES OF SOCIALIZATION

Several scholars have proposed that socialization takes place in distinct stages. For example, the first stage is one of “anticipatory socialization.” This stage includes all of the ways in which newcomers prepare themselves for entry into the organization as well as interaction they have with the organization such as recruitment, interviews, the selection process, and paperwork. This is followed by the “accommodation” stage. During this second stage, organizationally entry begins and newcomers tackle the challenges associated with learning to deal with the people as well as the actual work of doing the job. The final stage of socialization, which has been referred to in a variety of ways such as “role management,” is when newcomers have mastered both the social and task-oriented portions of the job.

## 3. COMMON OUTCOMES OF SOCIALIZATION

Research on socialization has focused on newcomer adjustment and outcome variables. Adjustment variables include how much clarity newcomers have with their role, how accepted they feel, and how well they are adjusting to the organizational or career culture. Culture refers to the norms and operating guidelines within an organization, so these are not usually explicitly defined for new employees. Rather, they must discern them through interactions and questions. Other outcomes that have been studied include job attitudes, such as job satisfaction and organizational commitment, as well as performance and turnover. Research has shown that socialization is positively related to adjustment, job attitudes, job performance, and organizational culture and is negatively related to turnover.

## 4. WHAT DO ORGANIZATIONS DO TO FACILITATE SOCIALIZATION?

In Van Maanen and Schein’s seminal article on socialization in 1979, these authors delineated six tactics that vary from individualist to institutional: the level of interaction with other newcomers (formal vs informal), the number of newcomers within a given cohort (collective vs individual), the order in which socialization takes place (sequential vs random), whether or not there

is a specific time frame for socialization (fixed vs variable), how newcomers are trained (serial vs disjunctive), and whether or not newcomers are stripped of their old identity (investiture vs divestiture). All of these dimensions can be boiled down to a continuum with highly institutionalized approaches, where newcomers are socialized in a formal setting as a group and given clear information about the sequence and timing of the socialization process, and highly individualized approaches, where newcomers are socialized informally and given little explicit information about sequence or timing. Organizations vary in terms of these tactics. For example, in an extreme case, the military has very institutionalized formal tactics where individuals are brought in as large groups, the individuals go through formal training that is tied to specific goals and time lines, and individuality is not encouraged regarding how tasks are done.

## 5. HOW DO INSIDERS AID IN SOCIALIZATION?

Although a key focus on socialization research is newcomers, organizational insiders are important as well. Peers, supervisors, and mentors are often referred to as “agents” of socialization and are seen as playing an integral role in facilitating newcomer sense making. Insiders can facilitate socialization by providing newcomers with additional information, giving advice, clarifying job instructions, and giving social support.

Research has shown that the strength and nature of the newcomer–manager relationship relates to the success or failure of newcomer socialization. In general, research has also shown that job attitudes are positively related to the quality of newcomers’ relationships with peers and the frequency of their communication with peers and that performance is related to social support from coworkers.

A significant discovery in organizational socialization research is how important mentors are for successful socialization. Mentors can provide many benefits such as teaching newcomers about the organization, offering advice, helping with job instructions, and providing social support. Mentors are also important because they can be a sounding board for questions that employees are too intimidated to ask their managers, and mentors can also provide the support structure for new employees to help them fit in both socially and politically. It is important for an organization to

recognize the importance of mentors when bringing new members into the organization. This can help with many things such as meeting and managing expectations better. Mentors can help new employees to learn about the organization more efficiently. Events and practices of the organization can be relayed with mentors simply through interaction.

## 6. NEWCOMER PROACTION

It has been argued that newcomers' personalities can influence the ways in which they see their roles and the challenges facing them. Information seeking also seems to affect organizational socialization. Studies have also demonstrated predictable patterns in the relative frequency with which newcomers seek various types of information. For example, newcomers tend to seek more technical information than social information, believe that there are costs associated with seeking information, and believe that seeking technical and referent information is the safest form of information seeking.

A key form of learning that newcomers need to gain is how to fit into the organization. Research has shown that a variety of factors influence how well newcomers learn to fit over time.

## 7. SOCIALIZATION RESEARCH METHODS

Socialization researchers tend to conduct longitudinal studies, use actual newcomers to organizations, and (often) use 3-month intervals (i.e., 3, 6, 9, and 12 months) to measure socialization-related variables, especially outcome measures. However, little empirical evidence has been offered to support the selection of these measurement periods beyond the fact that other studies have used them and found significant results using them. Often, the samples studied are college graduates or specific professionals such as nurses or accountants. Thus, what we know about socialization must be inferred from a relatively narrow set of occupations or job types studied.

## 8. CONCLUSION

The overall conclusions include the need for organizations to be proactive when socializing newcomers and the need for them to be aware of the surrounding conditions. Learning efficiently comes from training programs that may include mentoring, internships, and other ongoing training programs that are extremely important for socialization given that learning is key during the early stages. There is a need to understand how to effectively manage the changing workforce and how it affects the socialization process because the workforce is a key factor in determining employee commitment, performance, and retention and affects how people view the organization's culture. It is also necessary to have different socialization processes for different types of employees, such as permanent versus temporary (e.g., contingent, internships), and to examine where the particular organization is with regard to the changing workforce.

### See Also the Following Articles

Employment Interviewing ■ Organizational Culture and Climate ■ Recruitment

### Further Reading

- Bauer, T., Morrison, E. W., & Callister, R. (1998). Organizational socialization: A review and directions for future research. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 16, pp. 149–214). Greenwich, CT: JAI.
- Feldman, D. C. (1976). A contingency theory of socialization. *Administrative Science Quarterly*, 21, 433–452.
- Fisher, C. D. (1986). Organizational socialization: An integrative review. In K. M. Rowland, & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 4, pp. 101–145). Greenwich, CT: JAI.
- Louis, M. R. (1980). Surprise and sensemaking: What newcomers experience in entering unfamiliar organizational settings. *Administrative Science Quarterly*, 25, 226–251.
- Morrison, E. W. (1993). Newcomer information seeking: Exploring types, modes, sources, and outcomes. *Academy of Management Journal*, 36, 557–589.
- Van Maanen, J., & Schein, E. H. (1979). Toward a theory of organizational socialization. In B. M. Staw (Ed.), *Research in organizational behavior* (Vol. 1, pp. 209–263). Greenwich, CT: JAI.



# Organizational Structure

Lois E. Tetrick and Michael K. Camburn

George Mason University, Fairfax, Virginia, USA

1. Introduction
  2. Traditional Dimensions
  3. Modern Approach
  4. Conclusions
- Further Reading

## GLOSSARY

**alternative employment relationships** The use of different employment contracting to vary the number of hours worked, the length of employment relationship, and the combination of the two rather than the “traditional” full-time, long-term employee–employer relationship.

**centralization** The extent to which power is concentrated in the upper levels of management.

**formalization** The extent to which rules, regulations, and policies are written and enforced.

**globalization** The extent to which organizations cross national boundaries in producing and marketing their goods and services.

**matrix structure** An organizational structure characterized by the use of project teams in which employees may be members of multiple project teams; teams include expertise from multiple functional areas and exist only until the projects are completed.

**organizational learning** The process of acquiring new knowledge and skills that can be applied in novel ways.

**organizational structure** Mechanisms that determine the interactions among organizational members within the social system.

**specialization** The diversity of operations within an organization as well as the number of departments and positions within departments in an organization.

**work teams** Groups of employees that work to complete a specific task within the organization.

Organizational structure refers to the processes by which individuals within an organization interact with one another and with individuals outside the organization. The design of the organization, based on various organizing principles such as centralization, formalization, specialization, team structures, matrix structures, organizational learning, globalization, and alternative employment relationships, shapes individuals’ behaviors and organizational functioning. These characteristics of the organization are interdependent, and contextual factors may moderate the effectiveness of organizational structure components in various environments both at the individual and organizational levels.

## 1. INTRODUCTION

The structure of an organization refers to mechanisms that control the interaction of organizational members in achieving organizational goals. The effects of organizing mechanisms on individual employees’ behaviors and attitudes have received considerable attention, as have the effects of organizing mechanisms on

organizations' stability, growth, and adaptability. This article provides a summary of this research. First, it focuses on three traditional characteristics of organizational structure: formalization, centralization, and specialization. Then, it discusses modern approaches to organizing with an emphasis on teams, matrix organizing, organizational learning, globalization, and alternative employment relationships.

## 2. TRADITIONAL DIMENSIONS

In 1961, Burns and Stalker suggested that organizations were structured on a continuum ranging from mechanistic to organic. Mechanistic organizations reflected those in which power was centralized and control was achieved through rigid hierarchical communication channels and formal procedures. Mechanistic organizations are considered hierarchical, or tall, with multiple levels of management. Decision-making power is concentrated at the higher levels. Organic organizations, on the other hand, are more flexible and decentralized with open communication among organizational members across levels and minimal use of formal procedures. Organic organizations are heterarchical, or flat, with relatively few levels of management. Decision-making power generally is distributed across all levels in the organizations.

### 2.1. Centralization

Centralization refers to the concentration of power within an organization to the upper levels of the hierarchy, suggesting that the organization does not trust the employees to make decisions that are consistent with organizational goals. This concentration of power also has the potential to make employees more dependent on upper level managers for resources and information. Thus, in the centralized organization, control over resources and information is retained by the top management team and power resides in formal authority. Lower level personnel generally do not participate in decisions, even those regarding their own tasks.

#### 2.1.1. Individual Consequences of Centralization

Employees working in more centralized organizations experience less autonomy as a result of the increased control over resources and information by individuals at the upper levels in the organization. According to

Hackman and Oldham's Job Characteristics Model, reduced autonomy would be expected to result in lower motivational levels and ultimately in lower job performance. Also, the reduction in control resulting from lower levels of participation in the decision-making process can result in increased levels of stress. It would appear that decentralization is a more positive organizing mechanism from an individual perspective, although decentralization is not without some negative consequences. For example, in organizations where personnel decisions are decentralized, greater power is typically given to individuals' immediate supervisors in determining rewards and promotions. Because supervisors have greater power, there is an increased possibility that employees might feel as though they are being treated unfairly and might reduce their efforts. Depending on the quality of the relationship between an individual employee and his or her immediate supervisor, there might actually be a decline in job performance, especially discretionary aspects of job performance such as contextual performance and organizational citizenship behavior.

#### 2.1.2. Organizational Consequences of Centralization

Centralization allows an organization, or at least the upper levels of the organization, to control lower level employees' behaviors so that they are consistent with organizational goals. Also, high degrees of centralization reduce the need for high levels of formal procedures because managers already directly monitor and advocate appropriate behaviors without needing written standards. However, as the organization increases its members and functions, the core group of top managers is less able to monitor all of the activities within the organization and more authority is shifted to departmental managers. Furthermore, centralization may constrain adoption of innovative ideas because lower levels in the organization have little decision-making power.

### 2.2. Formalization

Formalization refers to the extent to which there are formal rules, regulations, and policies in an organization. These formal rules and regulations control the behavior of employees, institutionalizing practices and procedures that have been successful in the past. If these procedures are followed, they provide consistency in decision making throughout the organization.

### **2.2.1. Individual Consequences of Formalization**

The consistency of decision making provided by formal rules and procedures results in greater perceptions of organizational fairness and increased predictability within the work environment. This could result in increased job performance (especially contextual performance) and reduced job stress. In 2000, Schminke and colleagues suggested that the key to enhanced perceptions of fairness and predictability is not the extent of formalization but rather the extent to which the rules fit the organization's environment. For example, if the organization has a relatively stable and predictable environment, formalization matches the environment and is expected to have positive effects. However, if the environment is more fluid and dynamic, formalization might actually have negative consequences if the policies are perceived as unfair and/or too rigid. The gain in predictability from formalization can reduce job stress; however it can also lead to a sense of unnecessary red tape, a lack of control over the decision-making process, and a decreased sense of control, all of which can increase job stress.

### **2.3. Organizational Consequences of Formalization**

Formalization results in standardization of processes and outputs, allowing an organization to maintain high levels of efficiency and productivity. For certain practices, such as selection and promotion, formalization helps the organization to avoid disparate treatment of applicants and candidates by standardizing procedures across all departments. This standardization of practices reduces the likelihood of legal actions against the organization. However, as Sitkin and Bies noted in 1993, promotion of a legalistic environment in an organization may actually undermine the social trust between managers and employees. Strict adherence to formal policies and procedures can limit the organization's adaptability and innovation much as in the case of high levels of centralization.

### **2.4. Specialization**

Specialization refers to the diversity of operations performed within an organization. The organization may group similar areas of specialization into functional groups such as production, sales, and finance. Division of labor refers to the number of different job titles in the organization. In general, the greater the

number of departments within the organization and the more positions within each department, the higher the degree of specialization. A high degree of specialization creates a more complex organization with multiple types of expertise.

### **2.5. Individual Consequences of Specialization**

Higher levels of specialization result in employees perceiving a greater level of autonomy in an organization. This autonomy reflects the increased power of possessing unique skills and knowledge. Performing work that is not duplicated by others in the organization also enhances individuals' control over their own work.

#### **2.5.1. Organizational Consequences of Specialization**

Greater specialization within an organization indicates more divergence among managerial perspectives and interests. This divergence can lead to greater innovation, flexibility, and adaptability if there is not too high a level of formalization and centralization. However, upper level managers will often attempt to enforce similar standards across specialized subunits. This generally results in an increase, rather than a decrease, in formalization.

## **3. MODERN APPROACH**

Modern organizations are becoming more decentralized and specialized. There appears to have been a transition to what Moch called more pluralistic political systems rather than hierarchically coordinated social systems. Organizations have become more diversified and heterarchical. They are concerned with the process of organizing and directing personnel. This organizing approach is characterized by growth, flexibility, and adaptability. Organized power hierarchies have weakened, resulting in greater decentralization and increased specialization. There also has been a shift from control through formalization to greater self-management, with modern organizations using a combination of formalization and flexibility to obtain reliable results while enhancing opportunities for innovation. Organizational structures are still characterized by the traditional dimensions, but a shift toward organizing has resulted in the increasing significance of a variety of new structural components. This section discusses several structural dimensions that

have emerged with the modernization of organizations: work teams, matrix structure, organizational learning, globalization, and alternative employment relationships.

### **3.1. Work Teams**

One of the major factors associated with organizations shifting from the traditional hierarchical organization to a more diversified heterarchical organizational structure is the increased use of work teams. The increasing complexity of the business environment makes the use of work teams increasingly appropriate where diverse skills can be applied to strategic decision making. The use of self-directed teams makes fewer hierarchical levels necessary.

#### **3.1.1. Individual Consequences of Work Teams**

In the traditional approach to organizations, larger organizations increase the likelihood that employees will feel less valued. Use of work teams can functionally reduce the sense of largeness by embedding individual employees within smaller organizational units or teams. It also is likely that individuals will have a greater sense of accomplishment in completion of these team tasks rather than the smaller tasks that individuals might have in a non-team environment. However, if a work team has more members than is necessary, members might reduce their individual efforts. If the team is understaffed, on the other hand, members might experience low group potency, reduced motivation, and greater job stress.

#### **3.1.2. Organizational Consequences of Work Teams**

Organizing based on a work team structure allows organizations to undertake more complex projects. However, it is more difficult for upper level management to monitor and control performance in a team-based structure. Therefore, a specialized manager for each team is more effective and economical than a hierarchy of executives. The use of work teams also requires an organizational culture that trusts teams to work toward organizational goals relatively independently. This suggests that implementation of a team-based structure requires investment in training and development of teams and individual team members.

### **3.2. Matrix Structure**

A matrix structure is a complex organizational structure that is based on the use of project-driven teams,

generally requiring individuals with diverse functional expertise to be on the same teams. Project teams include multiple specialists, and each specialist may be part of multiple project teams. Members need to be highly skilled within their own discipline with strategic decision-making abilities and an ability to collaborate with others. Work teams, as described earlier, may work together indefinitely, but project teams in a matrix organization typically disband when the projects are completed. Matrix structures are intended to result in innovation, high task performance, and a collaborative shaping of organizational goals. Some automobile companies use a matrix structure during the development of new automobiles. Experts from a variety of functions, such as engineering, accounting, production, and marketing, all play a role in the success of new automobiles. Issues such as development, production, and sales are discussed, debated, and decided on by the team of experts, with each member bringing a unique and important perspective to the team.

#### **3.2.1. Individual Consequences of Matrix Organization**

The matrix organization involves considerable fluidity among organizational members, their assignments, and their social networks. Project team membership provides high levels of the core job characteristics. Thus, individuals would be expected to experience increased motivation and enhanced job performance. Furthermore, the social interaction among project team members has the potential to increase employees' social capital. However, working on multiple project teams with multiple managers can lead to role conflict and role ambiguity, both of which can result in job stress. Also, the relatively short duration of team membership and multiple team membership may reduce team members' identification with and commitment to each team.

#### **3.2.2. Organizational Consequences of Matrix Organization**

Project teams unite members with diverse skills and backgrounds that are expected to increase the quality of the projects and allow for innovative solutions if all members contribute. To some extent, they are the antithesis of stability in that they result in considerable movement within the organization by project team members. Matrix structures promote coordination across disciplines and functional areas within the organization, allowing a wide range of participation in

decision making and sharing of knowledge. However, communication across project teams is often not supported by the organization, and there may be power struggles among managers from different functional departments.

### **3.3. Organizational Learning**

Adaptability requires learning, and as organizations are faced with competitive pressures, it is important that individual and organizational learning occur at a continual and rapid pace. Organizational learning is an active process in which knowledge is shared among organizational members, with information being shared across organizational subunits.

#### **3.3.1. Individual Consequences of Organizational Learning**

A structure that promotes individual learning promotes individual growth. If individuals are given opportunities to learn through training, they are likely to believe that the organization is supportive and committed to them as individuals and will reciprocate with increased organizational commitment and contextual performance. Also, when the organization provides an environment in which learning is valued and mistakes are tolerated, individuals will experience an increased sense of control over their work environment, thereby decreasing job stress and increasing well-being.

#### **3.3.2. Organizational Consequences of Organizational Learning**

The intended consequences of organizational learning are for the organization to be more innovative and adaptive to its environment. By acquiring and using new knowledge and skills, the organization is able to develop novel and more effective strategies for competing within the business environment. The effectiveness of organizational learning interventions varies across groups within an organization, depending on the interpersonal climate.

### **3.4. Globalization**

Organizations have extended their domain from within a single country or region of the world to encompass multiple nations to reach a global market. There is no single standardized structure for multinational corporations. Their structures seem to reflect existing competitive environmental pressures and may vary across

locations. Strategic decision making may be tightly centralized around senior management in the home country; it may be decentralized to the national level, allowing for strategic planning to be tailored for each national market; or there may be more globally dispersed resources, although strategic decision making is still likely to be centralized with formal committees to facilitate communication and coordination among subunits. In 1989, Bartlett and Ghoshal categorized these different structures of multinational corporations as multinational, international, global, and transnational based on the degree of interdependence. The multinational model is decentralized and nationally self-sufficient, the international model has sources of core competencies centralized and all others decentralized, the global model is centralized and globally scaled, and the transnational model is dispersed, being the most interdependent and specialized model.

#### **3.4.1. Individual Consequences of Globalization**

As organizations become more global, individual employees are expected to travel more, including potential long-term assignments in countries other than their countries of origin. Adjustment to new cultures can be a powerful learning experience and a source of stress. This can affect individual employees as well as their families, and organizations frequently provide acculturation experiences for employees and their families before overseas assignments. It also has been shown that maintenance of social networks within organizations is important for expatriates' adjustment to the assignments and for the expatriates' successful return at the end of the assignments. Another aspect of globalization is communication across national boundaries on a regular basis. For example, coordination of telephone conferencing may involve individuals being available at hours outside of their regular working hours. This can create a sense of organizational injustice if it is not handled appropriately.

#### **3.4.2. Organizational Consequences of Globalization**

Multinational organizations vary in the degrees of complexity of their structures. The basis for interdependence in these organizations results from the flow of products, technologies, and people among subunits and between subunits and headquarters. The degree of interdependence can range from low in multinational models, to medium in international and global models, to high in

transnational models. Formalization may be able to handle the necessary coordination when there are relatively low levels of interdependence; however, as the level and complexity of interdependence increase, formalization generally is insufficient. There is a need for informal means of coordination and control.

### 3.5. Alternative Employment Relationships

Much of the literature on organizational behavior has focused on full-time employees. New employment relationships have emerged to include core part-time employees, peripheral part-time employees, independent contractors, guest workers, and temporary workers. These various work arrangements reflect alternative processes in organizing personnel.

#### 3.5.1. Individual Consequences of Alternative Employment Relationships

The advantage of the increasing number of employment relationships is that they give individuals greater choice in the number of hours they work, in the length of commitments to particular employers, and in opportunities for developing skills across work environments. In general, these effects are considered to have a positive effect on individuals' well being. A caveat to this positive aspect must be mentioned. If individuals are not able to match their employment arrangements with their desired employment arrangements, they may experience underuse, underemployment, and other negative consequences to their well-being. The existence of multiple work arrangements within an organization also can lead to status differences among employees, even though employees are essentially doing the same work. This can lead to perceptions of unfairness and discrimination.

#### 3.5.2. Organizational Consequences of Alternative Employment Relationships

Alternative employment relationships provide organizations with the flexibility of engaging the services of individuals as needed. Maintenance of core employees provides stability to organizations, whereas peripheral employees provide organizations with the ability to expand and contract the size of their labor forces based on production demands. Also, part-time, temporary, and short-term contracting allow organizations to expand

their labor pools to include individuals who might not be interested in full-time, long-term work arrangements. Thus, organizations can offer work arrangements that are consistent with their overall strategic decision making.

## 4. CONCLUSIONS

Organizational structures control social interactions within an organization. Centralization, formalization, and specialization have been shown to have effects on individuals' attitudes and behaviors as well as on organizations' stability, growth, and adaptability. Modern approaches to structure, such as the use of work teams, the adoption of a matrix structure, organizational learning, globalization, and the use of alternative employment relationships, reflect more of an organizing or process orientation to organizational structure, although they are not independent of centralization, formalization, and specialization. There are positive and negative aspects of all of these aspects of organizational structure, and context is important in understanding the choice of various aspects of organizational structure. It is recognized that structure should follow strategy so that the impact of the structure is consistent with the organization's strategy. However, as Burns and Wholey reported in 1993, adoption of a particular organizational structure may be explained somewhat by organizational characteristics, but it appears that what other organizations are doing in the local network influences the choice of a particular structure.

### See Also the Following Articles

Industrial/Organizational Psychology across Cultures  
 ■ Industrial/Organizational Psychology, Overview ■ Organizational Culture and Climate ■ Organizational Diagnosis ■ Organizational Participation ■ Organization Development

### Further Reading

- Bartlett, C., & Ghoshal, S. (1989). *Managing across borders: The transnational solution*. Boston: Harvard Business School Press.
- Burns, T., & Stalker, G. M. (1961). *The management of innovation*. London: Tavistock.
- Burns, L. R., & Wholey, D. R. (1993). Adoption and abandonment of matrix management programs: Effects of organizational characteristics and interorganizational networks. *Academy of Management Journal*, 36, 106–138.

- Coyle-Shapiro, J. A-M., Shore, L. M., Taylor, M. S., & Tetrick, L. E. (2004). *The employment relationship: Examining psychological and contextual perspectives*. Oxford, UK: Oxford University Press.
- Easterby-Smith, M., & Lyles, M. (Eds.). (2003). *The Blackwell handbook of organizational learning and knowledge management*. Malden, MA: Blackwell.
- Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison-Wesley.
- Kostova, T., & Roth, K. (2003). Social capital in multinational corporations and a micro-macro model of its formation. *Academy of Management Review*, 28, 297-317.
- Mintzberg, H. (1993). *Structure in fives: Designing effective organizations*. Englewood Cliffs, NJ: Prentice Hall.
- Moch, M. (1976). Structure and organizational resource allocation. *Administrative Science Quarterly*, 21, 661-674.
- Pearce, J. A., II, & David, F. R. (1983). A social network approach to organizational design-performance. *Academy of Management Review*, 8, 436-444.
- Rousseau, D. M. (1997). Organizational behavior in the new organizational era. *Annual Review of Psychology*, 48, 515-546.
- Schminke, M., Ambrose, M., & Cropanzano, R. S. (2000). The effect of organizational structure on perceptions of procedural fairness. *Journal of Applied Psychology*, 85, 294-304.
- Sitkin, S. B., & Bies, R. J. (1993). The legalistic organization: Definitions, dimensions, and dilemmas. *Organization Science*, 4, 351-354.
- Tetrick, L. E. (1997). Macro-organization: Organization structure. In J. M. Stellman (Ed.), *ILO encyclopaedia of occupational health and safety* (4th ed., pp. 34.35-34.36). Geneva, Switzerland: International Labor Office.



# Organization Development

W. Warner Burke

Teachers College, Columbia University, New York, New York, USA

1. An Organization Development Case
  2. Primary Characteristics of OD
  3. Roots of OD
  4. Related Theories That Underlie OD
  5. OD as a Process of Change
  6. Understanding Organizations: The Diagnostic Process
  7. Four Models Commonly Used by OD Practitioners
  8. Planning and Managing the Consequences of Change
  9. Does OD Work?
  10. OD Values and Considerations for the Future of the Field
- Further Reading

## GLOSSARY

**action research** Collecting relevant information (data) about a given issue or problem (the research aspect) for the express purpose of providing ways to help resolve the issue or solve the problem (the action aspect).

**appreciative inquiry** Acts of exploration and discovery (asking many questions) for the purpose of recognizing the best in people (e.g., affirming strengths, potential, and successes).

**cognitive map** A thought process, mental exercise of establishing and tracing paths throughout an organization to help clarify and take corrective actions for organizational improvement and change.

**development** Growth, advancement, strengthen, and/or to bring an idea into action.

**group unconscious** A theory emanating from Freudian psychoanalysis that argues that a group of people has a collective unconscious analogous to an individual's unconscious.

**information technology system** The use of computers for transporting information throughout an organization; system, in this case, means that all organizational members are (or can be) connected electronically and therefore form a communicative whole.

**interdependent multivariate systems** Statistical term that is used to convey the notion that all organizational variables and components are interconnected and form a whole; the statistical method is a way of numerically demonstrating this idea.

**intervention** The act of intervening, coming between two or more parties in an organization—boss and subordinates, two or more departments, divisions, or business function/units, an individual and the larger group of which he or she is a part, or the organization and another organization (e.g., acquisition or merger)—for the purpose of initiating or furthering organizational change and development.

**meta-analysis** A statistical procedure for comparing the results of many studies of like content.

**multirater feedback** Based on specific behavioral ratings, a process of providing information about a person's leadership, management of others, or simply a range of individual behaviors, whereby a comparison is made of how one rates oneself with how others rate him or her on the same set of behaviors.

**nonlinear complex systems theory** Originating from chaos theory in the physical sciences; the proposition is that organizational behavior is affected by many variables that interact powerfully and interdependently but not in a single direction; cause and effect therefore is rarely as simple as it sounds (i.e., being a "one-way street").

**norm** A standard or rule of conduct to which people conform (e.g., a dress code).

**open system framework** A scheme or model usually in a two-dimensional arrangement that depicts an organization as having input (external environment), throughput (the operation of the organization), and output (performance) with a feedback loop connecting performance with the external environment; any organization is dependent on its external environment for survival, as in a living organism, yet through performance has an impact on and influences its external environment.

**organization** A sociotechnical system that has some purpose in society; sometimes it is only a social system such as a club.

**refreezing** Once organization change has occurred, a process of stabilizing and institutionalizing the new behavior that supports the change.

**self-directed groups** Organizational units that conduct their work without a leader or manager who has formal authority; group leadership occurs but usually in an informal and rotating fashion.

**sensitivity training** Based on interacting with and learning from others in a small group, a process of helping individuals to develop greater self-awareness and become more sensitive to how they affect and are affected by others.

**sociotechnical system** The principle that a work organization is both a technical system—the content or nature of the work itself—and a social system—the people involved and how they interact to get the work done; to overlook either system is to misunderstand the organization.

**transactional** That aspect of an organization involving daily operations, including transactions between organizational members and between them and those the organization serves.

**transformational** That aspect of an organization involving the larger system in the context of long time frames; its purpose, mission, strategy, culture, and senior leadership and how these organizational variables interact with the external environment and with one another.

**unfreezing** The act of changing (“melting away”) current employee and managerial behavior that no longer leads to optimum organizational efficiency and effectiveness, with the act being education and training, the introduction of new technology requiring different work habits, providing evidence from the external environment that a new strategy with different products and services is required to ensure long-term survival, or all of these.

Organization development (OD) is about changing an organization—an institution, a company, a government agency, a community service agency, a university, a school, or even a church, synagogue, or mosque. Organization therefore is broadly defined—that is, a sociotechnical system that has some purpose in society. Sometimes, however, it is only a social system

such as a club. The term development means growth, advancement, strengthen, or to bring an idea into action. By implication at least, development is about change. The change may be gradual or sudden but change nevertheless. Organization change can happen in a number of ways—by an election or appointment of a new leader; coercion via some civil disturbance; acquisition by a larger, more powerful organization; introduction of a new technology that immediately replaces the old; etc. For OD, organization change occurs by way of the application of behavioral science—that is, using knowledge from psychology, sociology, and cultural anthropology to develop and hence change the sociotechnical system.

## 1. AN ORGANIZATION DEVELOPMENT CASE

The organization development (OD) example described here began with the merger of two banks in a large metropolitan area of the United States. The two organizations were very similar with respect to number of employees, length of existence (both founded in the mid-19th century), and annual revenues, but they were highly dissimilar regarding their respective cultures. For example, one focused much more on cost-effectiveness as the way to a profitable bottom line, whereas the other emphasized revenue growth via developing new markets and services. The latter's motto seemed to be “you have to spend money to make money.” In other words, much work would be needed to successfully merge the two banks.

The CEO of the newly merged bank sought consultation—one for strategic planning and another for OD. The two consulting firms that were brought in worked in tandem but focused on different initiatives to facilitate the merger and to foster change. The OD firm's consultation began with three primary initiatives. The first initiative was to conduct interviews with the top 19 executives of the merged bank. These interviews consisted of broad, open-ended questions, such as “What is going well regarding the merger?” and “What is not?” The interviews served as a data-gathering source for the second initiative, which was an organizationwide survey. The survey was conducted 6 months into the merger to assess the general state of morale and, specifically, to determine the degree of understanding and opinions about the bank's strategy, its leadership, the newly merged culture,

organizational structure and systems, and the adequacy of management in general. The third initiative was to discard the two previous banks' mission statements and craft a new one for the merged bank. To accomplish this initiative, a task force of 16 employees was formed that represented a cross section of the bank. Working in pairs, the task force sought input from the various parts of the bank they represented, eventually involving several hundred employees. They used the input from the expanded group of employees to craft the initial draft of a company mission statement. After a number of revisions the task force presented its statement to senior management and later to the board of directors. A final version was unanimously agreed on by both senior management and the board. In fact, the task force received a standing ovation for their work from the board of directors.

Results from the organizationwide survey using a 5-point Likert scale showed that most employees were positive about the merger (example of survey question: "To what extent have mergers and acquisitions made the bank more efficient?") and their leadership, at least at the very top; clear about the new strategy (e.g., "To what extent are the bank's products and services competitive with the top tier of banks in our geographical region?"); and strongly supported the new mission statement (e.g., "To what extent does the bank's mission include valuing cultural differences?"). On the other hand, employees were somewhat confused about culture (e.g., "To what extent is it safe to say what you feel?") and structure, especially with respect to accountability; critical of the reward system, especially compensation and incentives (e.g., "To what extent does the bank have rewards and benefits necessary to attract and retain the very best people?"), information technology (e.g., "To what extent are the computer systems that you use advanced and efficient?"), and the budgeting process; and quite negative about leadership below the very top five or so executives and highly critical of management (e.g., "To what extent does your supervisor provide people with the information they need to do their jobs effectively?") overall.

The survey results led to further initiatives and interventions. The information technology system was strengthened, the reward system improved, and a major initiative was launched to address the deficiencies in leadership and management beginning with the top 125 executives and managers. These programs for leadership and management emphasized self-awareness for individual development via multirater

feedback, personality and leadership assessment, and coaching.

Approximately 1 year after the first organizational survey, a second survey was conducted. Responses to 125 questions resulted in two questionnaire items remaining the same as those from the first survey and all remaining items showed higher ratings. Although it cannot be scientifically proven that the desired changes for the bank did indeed occur, the bank's executives, managers, and employees believed that positive change had occurred. The results of the survey were interpreted within the framework of an overall organizational model of performance and change. The model included dimensions such as external environment, mission and strategy, culture, leadership, structure, systems, climate, motivation, and performance.

## 2. PRIMARY CHARACTERISTICS OF OD

The previous case helps to illustrate what constitutes an OD approach to organization change. The following characteristics, when applied to the case, help exemplify the essence of OD.

### 2.1. Relevance

OD must be relevant to the organization's purpose and to what its employees actually do every working day. In the bank case, the OD practitioner worked initially on the organization's mission and ensured that this work linked directly to the bank's business strategy.

### 2.2. Sequential

OD practice involves a sequence of activities. After entry and contracting (reaching an agreement about the consulting services to be provided and when), the OD practitioner then gathers data about the organization and its leadership/management, determines a diagnosis about what the data mean, provides feedback to the client organization, and facilitates collaboratively with the client several interventions (initiatives) that help to address issues and needs for change embedded in the data and the diagnosis. After the interventions are launched, the next step in the sequence is to gather data again in order to evaluate progress. In the case,

1. Contracting occurred with the CEO and his immediate direct-report executives.

2. Data gathering took the form of executive interviews, which in part set the agenda for further data gathering via the initial organizationwide survey.

3. Based on the initial interviews of the top 19 executives, feedback was provided to them that led to off-site meetings facilitated by the OD consultant to address issues highlighted from a summary of the 19 interviews. Further larger system feedback was later provided for all managers in the bank and subsequently to all employees from the organizationwide survey.

4. Interventions from the initial feedback consisted of top executive team building, changes in the information technology and reward systems, and the launching of leadership/management development workshops.

5. Further data gathering and evaluation of the change initiatives (interventions) up to that point were conducted using a second organizationwide survey.

### 2.3. Systemic

OD is a system approach to organization change. In the bank case, this was manifested in the construction and feedback of the data from the survey. The 125 items comprising the survey were arranged in 12 categories—that is, according to the 12 boxes in the Burke–Litwin causal model of organization performance and change. This model is an open-system framework (i.e., input-throughput-output with a feedback loop connecting output to input) and consists of the following categories: external environment, mission/strategy, leadership, culture, structure, management practices, systems, work-unit climate, job-person match, individual needs and values, motivation, and performance. The survey feedback in the bank case was in the form of these 12 categories with priorities for action steps according to the arrangement of the boxes in the model in Fig. 1.

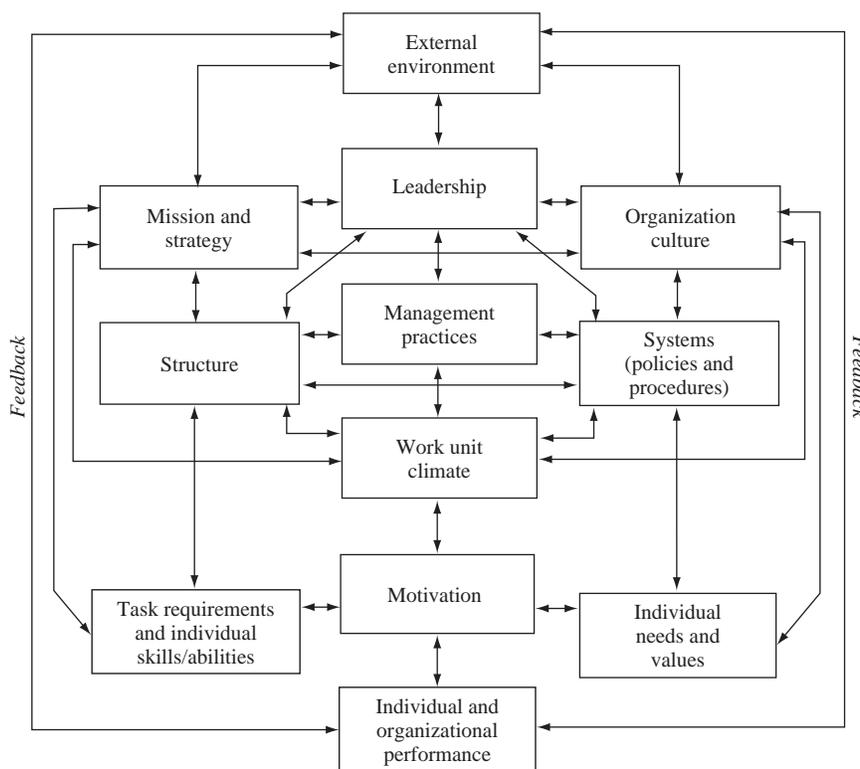


FIGURE 1 The Burke–Litwin model of organizational performance and change. From the Burke–Litwin model of organizational performance and change. © 1987, 1992 by W. Warner Burke, Associates, Inc.

## 2.4. Research Based

OD is steeped in the tradition of action research. This means that whatever action for change is taken, it is based on data, usually both qualitative and quantitative. OD interventions are “organic”—that is, based on what the data and the diagnosis reveal. Therefore, the action is based on the research (research loosely defined).

## 2.5. Humanistic

OD practice typically conforms to a set of values grounded in humanism and often takes the form of

- Involvement of organizational members in decisions that directly affect them
- Respect for differences among people
- Openness and honesty in communication
- A strong belief that the solution to organizational problems often lies within the knowledge and experience of people in the organization who have lived with the problem daily.

Usually this knowledge is either unspoken, perhaps due to an implicit norm not to disagree with the boss or with one another or the problem is a “sacred cow,” or the knowledge is tacit—that is, what may be known is not forbidden to discuss but, rather, is implicit and difficult to articulate. The OD practitioner’s role and function in these instances is to bring to the surface what organizational members are fearful of discussing or to help make explicit what is tacit, implicit. The previous four values are meant to be representative, not exhaustive.

## 3. ROOTS OF OD

Where did OD come from? Three primary streams of change methodologies and three institutions provide the basis for the emergence of OD.

### 3.1. Sensitivity Training

Evolving from the work of Kurt Lewin and colleagues during the 1940s, this form of training concerned the application of knowledge regarding group dynamics and interpersonal relations for individual learning. The learning occurred via feedback from others as to how one’s behavior affected these other group members. Learning also came from an analysis of interpersonal relations in small groups in which the primary source of information or data from which to learn was

the behavior and feelings of the group members. In the late 1950s and early 1960s, sensitivity training was employed in organizations to enhance interpersonal, managerial, and team effectiveness. These activities became the forerunner of what is a cornerstone of OD today—team building.

### 3.2. Sociotechnical Systems

Most any organization is at the same time a technical system and a social system, and to ignore either one during a change effort is to invite failure. Trist and Bamforth, both professional staff members of the Tavistock Institute, served as consultants to a coal mining company in the northern part of England. The company had made a technological change regarding the way coal was mined and instead of the change increasing productivity as management expected, morale of the workers plummeted and productivity declined. After studying the situation, Trist and Bamforth noted that the technological change had changed the social system. The miners could no longer work in small groups as before but were forced to work alone, resulting in a greater division of labor. The consultants explained to management what the real problem was and how to make another change that would not result in a return to the old method but modify the work process so that it would be more composite (i.e., less divided) and help to retain the social patterns that were so important to the miners. This way of thinking and changing organizations led to a number of interventions that are used by OD practitioners today, such as job redesign, self-directed work groups, and organizational redesign, especially in manufacturing companies.

### 3.3. Survey Feedback

Although not widespread, using a questionnaire to assess worker morale, attitudes, and perceptions in the 1950s was not new. What often occurred with the use of an organizational survey, however, was that management would retain the data and rarely report the results to those who responded to the questionnaire in the first place. Psychologists, particularly Rensis Likert and Floyd Mann at the University of Michigan’s Institute for Social Research, found that surveys did not work very well because of this lack of “reporting back.” Sometimes as a result of conducting a survey matters got worse. In other words, when asked to answer a questionnaire, employees thought that management

cared about what was on their minds. However, when nothing happened as a result of the survey, employees would become dissatisfied, even cynical, and productivity and morale would decline to levels lower than those before the survey was conducted. These Michigan psychologists believed that reporting the results to the people who answered the survey in the first place was a necessary step toward making the process a more positive experience. Necessary, they believed, but not sufficient for bringing about positive organization change. Mann, in particular, developed a way of stimulating action in an organization as a consequence of the survey. His procedure was to report the survey results to intact work and management teams—that is, to people who worked together. These work units would then analyze the results with respect to their own domain of work and responsibilities and then plan and subsequently take steps toward solving the problems identified in the survey. This procedure may not be groundbreaking today, but in the 1950s it was rare indeed. The procedure was called survey feedback simply because the results were “fed” back to managers, staff, and work units in a structured manner and the process set into motion action for resolving problems at the work unit level (i.e., where the knowledge resided about what to do). Survey feedback is common practice in OD today.

Although these three primary historical roots of OD differ, they all have in common the action research methodology and mode of thinking regarding organization change. It should also be recognized that theory, research, and methodology from industrial/organizational psychology have contributed significantly to the practice of OD.

## 4. RELATED THEORIES THAT UNDERLIE OD

There is no all-encompassing theory of OD. There are a number of “mini” theories that help to undergird the practice, however. The following sections provide a simple listing of these theories. They are grouped according to organizational level and emphasis of OD practice.

### 4.1. Individual Level

#### 4.1.1. *Need Theory of Motivation*

Maslow’s theory concerns a hierarchy of human needs that begin with physiological or creature comfort

needs (air, water, safety, etc.), include security, belongingness, and ego status, and conclude at the top of the hierarchy with self-actualization. Herzberg’s theory is more about job satisfaction and dissatisfaction than about needs per se. Factors that reduce dissatisfaction he called “hygiene” or preventive, such as fringe benefits, good working conditions, and salary. Factors enhancing job satisfaction he called motivators and include recognition, challenging work, and autonomy. OD practice that relies on these theories includes job enrichment, quality-of-work-life projects, empowerment interventions, and career development.

#### 4.1.2. *Expectancy Theory of Motivation*

This theory is cognitive in nature and consists of individuals’ expectations that (i) their efforts will lead to results, (ii) if enhanced these efforts will lead to rewards, and (iii) these rewards will be worth it to them. OD practice associated with this theory includes work redesign and modifying performance appraisal and/or systems.

#### 4.1.3. *Job Satisfaction*

Related to Herzberg’s theory, this theory posits that there are three primary psychological states that affect job satisfaction: experienced meaningfulness of the work, experienced responsibility for the work and its outcomes, and knowledge of results (i.e., performance feedback). OD practice that relies on this theory includes primarily job redesign, job enrichment, and performance feedback procedures.

#### 4.1.4. *Positive Reinforcement*

Although Skinner’s research and theory were more about individual animal and human behavior in the broad context of learning and change and not about organizational behavior per se, his principles can be applied to organizational reward systems. The primary emphasis is therefore on incentives and the belief that partial and variable reinforcement of behavior is more lasting and effective than continuous reinforcement. OD practice that relies on this theory includes modifying a reward system in general and putting into place some form of an incentive system.

## 4.2. Group Level

### 4.2.1. *The Group as the Focus of Change*

Applying concepts from the science of physics, Lewin analyzed human behavior in terms of “forces” impinging on an individual. These forces might be group norms to which the individual is expected to conform, supervisory pressure to exert more effort on the job, or deeply held beliefs and values emanating from the larger culture about what is the right thing to do. Changing norms and values is therefore tantamount to changing an organization’s culture. OD practice that relies on this theory includes team building, force field analysis, decision-making processes, and ultimately culture change (i.e., changing conformity patterns).

### 4.2.2. *Changing Values through the Group*

The work of Argyris is broad and thus difficult to reduce to one theory. Appropriate for OD are primarily ideas about change, particularly the motivation to change. The premise is that when individuals (think managers) are confronted with differences between what they espouse and what they actually do, they will be motivated to reduce the gap and as a result change their behavior. OD practice that relies on this theory includes “gap analysis”—that is, surfacing the incongruence between what managers say, the words they use (Argyris calls it espoused theory), and what their actions actually become (theory in action), especially decision making—and team building. Also, in collaboration with Schön, Argyris helped spawn another OD activity that is more at the larger system level—organizational learning.

### 4.2.3. *The Group Unconscious*

Steeped in psychoanalytic theory, Bion took Freud’s thinking one step further and argued that a group has a collective unconscious much like an individual’s. This collective unconscious revolves around the group leader, whether formal or informal, and the group’s motivation, albeit below conscious awareness, is to replace the leader irrespective of his or her performance. OD practice that relies on this theory includes group behavior diagnosis and team building, self-directed work groups, and sociotechnical systems interventions.

## 4.3. Larger System Level

### 4.3.1. *Participative Management—One Best Way*

A normative theory, Likert’s position was that any large organization could be considered as one of four possible systems. System 1 was authority centered, top-down, and unilateral; system 2 was a benevolent autocracy; system 3 was consultative (i.e., ask subordinates for their ideas and opinions but with the manager reserving the right to make the final decision); and system 4 was participative management with an emphasis on consensual decision making. Likert’s argument was that system 4 was the best way to manage an organization. OD practice that relies on this theory includes survey feedback and team consensus decision making (i.e., change to a participative form of management).

### 4.3.2. *It all Depends*

Opposite to a normative way of thinking, Lawrence and Lorsch stated that how an organization should be managed and designed depends on its relationship to its external environment (i.e., the organization should be as responsive as possible to external forces). Their key concept was interface—three of primary importance: the organization–external environment interface, intergroup interfaces within the organization, and the implicit relationship between the organization and the employee (the psychological contract). OD practice that relies on this theory includes organizational structure redesign, intergroup conflict resolution, and cross-functional groups for purposes of integration.

### 4.3.3. *The Organization as a Family*

Levinson considered top management as replicating the nuclear family, with the CEO being the father and the remaining executives acting as siblings, with rivalry among them being the dominant theme. Not surprisingly, Levinson’s work has been most applicable to family-owned businesses. OD practice that relies on this theory includes succession planning in a family-owned business, coaching the CEO, and team building for the top executive group.

Table I summarizes the minitheories discussed in the previous sections.

**TABLE I**  
**Summary of Primary OD Theorists According to Their Perspectives, Emphases, and Applications<sup>a</sup>**

<i>Perspective</i>	<i>Theorist</i>	<i>Emphasis</i>	<i>Application</i>
Individual	Maslow and Herzberg	Individual needs	Career development, job enrichment
	Vroom and Lawler	Individual expectancies and values	Reward system design, performance appraisal
	Hackman and Oldham	Job satisfaction	Job and work design, job enrichment
Group	Skinner	Individual performance	Incentive systems, reward system design
	Lewin	Norms and values	Changing conformity patterns
	Argyris	Interpersonal competence and values	Training and education
	Bion	Group unconscious, psychoanalytic basis	Group behavior diagnosis
System	Likert	Management style and approach	Change to participative management
	Lawrence and Lorsch	Organizational structure	Change contingent on organizational environment
	Levinson	Organization as a family, psychoanalytic basis	Diagnosis of organization according to familial patterns

<sup>a</sup>Source: W. W. Burke, *Organization development: A process of learning and changing* (2nd ed.). Copyright 1994 Addison-Wesley Publishing Co. Reprinted from p. 53 by permission.

## 5. OD AS A PROCESS OF CHANGE

As stated earlier, OD is based on action research as an approach (Fig. 2) and immersed in the Lewinian tradition exemplified by the following phrase: No action without research and no research without action.

### 5.1. Levin's Three-Step Change Procedure

Lewin provided a simple (although not easy to execute) three-step procedure for bringing about change. The following are the three steps:

1. Unfreeze the current level of behavior
2. Action that changes the system from its original state to a new level of behavior
3. Refreeze, meaning to secure the new behavior against further change or "slippage"

### 5.2. Schein's Elaboration on Levin's Three-Step Procedure

In 1987, Schein developed a useful elaboration of the three steps. He refers to them as stages rather than steps to emphasize the point that they overlap (i.e.,

they are not discrete). Schein's expansion of the three stages is presented in the following sections.

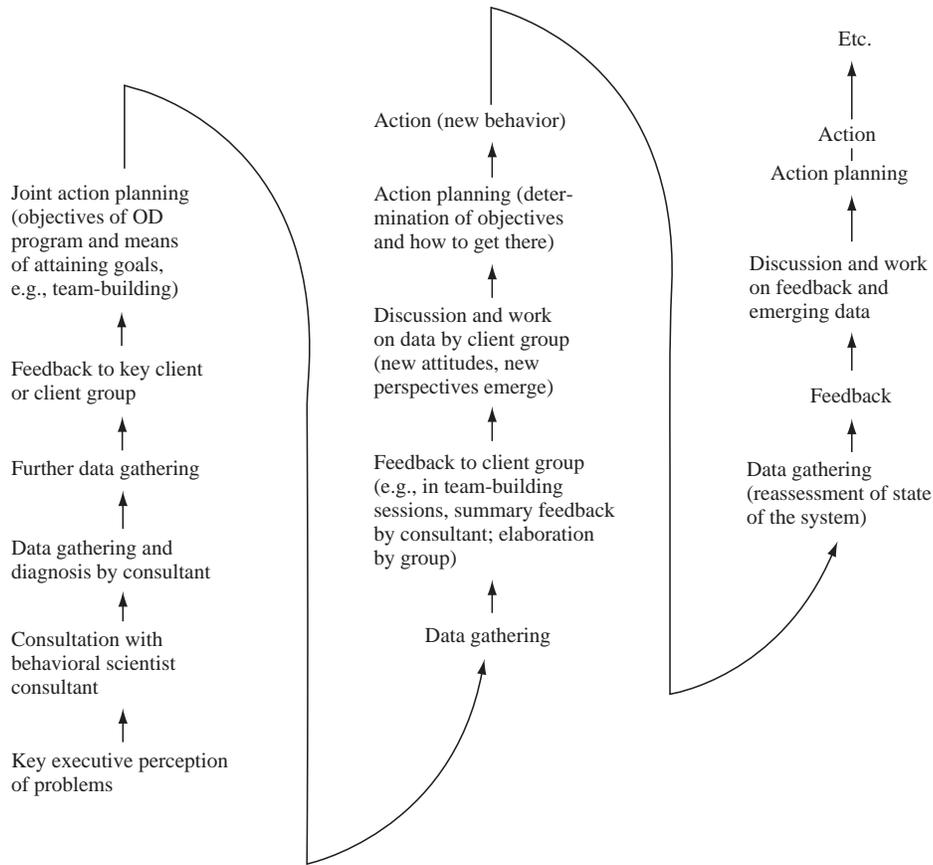
#### 5.2.1. Stage 1: Unfreezing

Schein defined this stage as creating motivation and readiness to change. The following are his three ways of unfreezing:

1. Disconfirmation of the way things are, which leads to dissatisfaction and a need for change (e.g., costs are too high, morale is too low, and productivity is declining).
2. Induction of guilt or anxiety by establishing a gap between the current state and what would be much better. In the face of this kind of gap, most people become motivated to reduce it.
3. Creation of psychological safety to help deal with disconfirmation, guilt, and anxiety so that people can take the required action. The action needs to be associated with feelings of accomplishment and worthiness instead of humiliation or a loss of self-esteem.

#### 5.2.2. Stage 2: Changing

Changing is the movement toward a new state that Schein believes can be accomplished via two processes. One process is the identification of organizational



**FIGURE 2** Action research model for organization development. From W. L. French, *Organization development: Objectives, assumptions, and strategies*, © 1969 by the Regents of the University of California. Reprinted from the *California Management Review*, 12(2), 26, by permission of the Regents.

members with a new role model, new leader, or perhaps the OD consultant to obtain a more attractive point of view and perspective. The second process is to scan the environment for new but relevant information that can facilitate the acquiring of a different perspective and point of view. This new information can come from a research project, outside expertise, or people in another organization who have gone through a similar organization change.

**5.2.3. Stage 3: Refreezing**

Schein conceptualizes this stage in two ways: the individual and relations with others. Refreezing with respect to the individual organizational member is a matter of feeling comfortable with the change—that it fits well with one’s self-concept, values, and new modes of behavior. Relations with others concerns also achieving a level of comfort. As one’s behavior

changes, the new behavior affects others differently than in the past. Will this be okay? Will others also change and work toward congruence of these new behaviors with one another? Reaching and maintaining these new levels is what Schein (and Lewin) means by refreezing.

**5.3. Lippitt *et al.*’s Five-Step Elaboration**

Another elaboration of Lewin’s three steps was developed in 1958 by Lippitt *et al.*, who expanded the three steps into five phases:

1. Development of a need for change (Lewin’s unfreezing)
2. Establishment of a change relationship between the client and the consultant(s)

3. Working toward change (Lewin's action movement step)
4. Generalization and stabilization of change (refreezing)
5. Achieving a terminal relationship between the client and consultant(s)

Lippitt and colleagues viewed an organizational change effort from the perspective of the consultant, which they referred to as the "change agent." To summarize, four models or approaches to organization development have been briefly described: action research, Lewin's three steps, Schein's elaboration of Lewin's steps, and Lippitt

*et al.*'s expansion. Figure 3 provides an overall summary and integration of these approaches.

Since these approaches overlap, they can be considered together in the form of an overall approach for OD that consists of the following elements:

- An outside consultant or change agent (the consultant may be an employee of the organization but would still be an "outsider" to the group in the organization being served)
- Data gathering from the client system (either the total system or some part that is relatively distinct—a

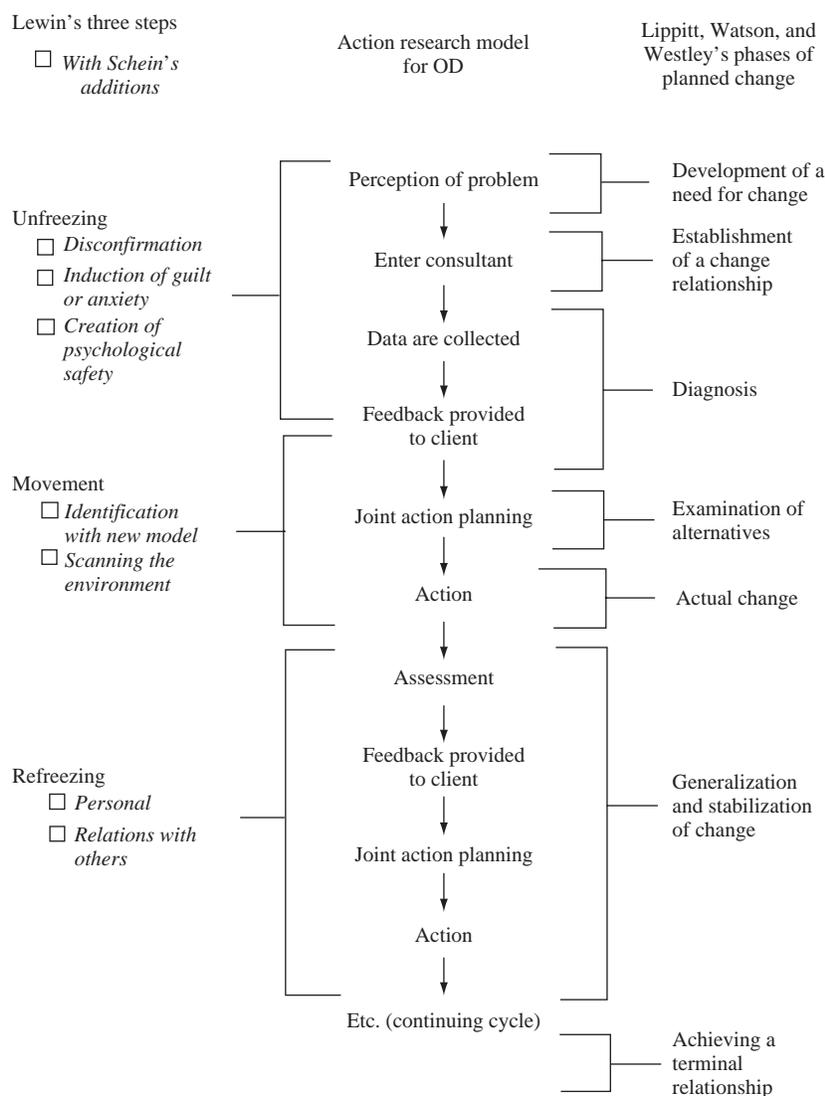


FIGURE 3 Comparison of the four models of change. From W. W. Burke, *Organization development: A process of learning and changing* (2nd ed.). © 1994 Addison-Wesley Publishing Co. Reprinted from p. 64 by permission.

division, a business unit, etc.) for the purpose of diagnosis (i.e., to understand in as much depth as possible the inherent nature of the system and its culture)

- Determine the major domains that need to be changed and problems to be solved
- Report this diagnosis to the client system so that appropriate action can be planned in a collaborative manner between the consultant (OD practitioner) and the client system
- Implementation of the planned change that is based on valid information in the form of interventions into the system that is usually conducted by the clients but with the continuing help of the consultant
- Institutionalization of the change

In short, the seven phases and processes of OD practice are as follows:

1. Entry: Establishing initial contact and a relationship with the client
2. Contracting: Agreement about the work to be done together (consultant and client); for example, who will do what and when
3. Diagnosis: Consists of data gathering and making sense of the information that has been collected (an organizational model helps with the latter process)
4. Feedback: Reporting to the client the summarized data and what the data mean
5. Planning change: Establishing action steps to accomplish the change
6. Intervention: Initiatives and actions that lead to change
7. Evaluation: Measures that help to determine if the change is taking effect

## **6. UNDERSTANDING ORGANIZATIONS: THE DIAGNOSTIC PROCESS**

Methods for data gathering in the OD process consist of (i) structured interviews, (ii) questionnaires either to embellish interviews (or vice versa) or to collect data in the form of an organizationwide survey, (iii) archival information (e.g., statements of mission and strategy, annual reports, and planning documents), and (iv) observations by the OD practitioner (e.g., paying attention to artifacts in the organization, sitting in on key management meetings, and listening carefully to the language of organizational members).

Often, these forms of data gathering produce more information than either the consultant or the client can deal with. Moreover, these data in their “raw form” are not obvious with respect to meaning. After all, the data are more about symptoms than causes of problems and issues. Thus, a framework or organizational model for summarizing and understanding a large amount of data can be useful. To be useful, the data need to be treated in organizational terms. OD represents a systems approach to change and a model can help to translate raw data into systems language. Thus, a model needs to be an effective representation of an organization.

An early organizational model was developed by Harold Leavitt in 1965. His diamond-shaped framework consisted of the following four concepts positioned at the corners of the diamond with interconnecting lines: structure, technology, people, and task. His model helped to set the stage for thinking about organizations as interdependent multivariate systems. In fact, the popular McKinsey 7-S model was derived from Leavitt’s work.

As stated previously, an organizational model can be useful in a number of ways:

- It can help to categorize and summarize.
- It can help to enhance understanding. For example, finding that problems exist in four categories but not in six others helps to establish priorities for action.
- It can help to interpret data about the organization. In other words, how categories in the model are arranged vis-à-vis one another can aid in determining what action to take regarding change. For example, in Weisbord’s six-box model, the leadership box (category) is central, with the remaining five boxes surrounding leadership. For Weisbord, leadership serves as the coordinating mechanism for the other five categories, attempting to keep them in balance. In Weisbord’s thinking, for any action to be taken toward change, leadership will inevitably play a key role.
- It can help to provide a common, shorthand language. Simply using the category culture, for example, covers a lot of ground.
- It can help to guide action for change. In other words, it can serve as a “cognitive map.” If some of the categories are more important than others with regard to change, how the categories are situated in the model vis-à-vis one another can provide direction for change that is likely to be clearer than otherwise might be the case.

As previously mentioned, an organizational model can indeed be useful, but there are at least two important caveats:

Content is critical: A model is only as good as the components selected and the arrangement of these components—that is, how each part relates to the others. Organizations consist of many components. The myriad parts need to be categorized into a workable number and it must be determined which components are the most important.

Metaphors are paradoxical: In 1997, Gareth Morgan stated this caveat eloquently: “Metaphor [a model is a metaphor after all] is inherently paradoxical. It can create powerful insights that also become distortions as the way of seeing created through a metaphor becomes a way of not seeing” (p. 5).

## 7. FOUR MODELS COMMONLY USED BY OD PRACTITIONERS

### 7.1. Weisbord’s Six-Box Model

The six boxes are leadership (which is central) surrounded by purposes, structure, rewards, helpful mechanisms, and relationships. The six are arranged in a circle surrounded by environment as a seventh variable. The primary diagnostic question that Weisbord asks for each box helps to define that particular category. The questions are as follows:

- Leadership: Does someone keep the boxes in balance?
- Purposes: What business are we in?
- Structure: How do we divide the work?
- Rewards: Do all needed tasks have incentives?
- Helpful mechanisms: Do we have adequate coordinating technologies?
- Relationships: How do we manage conflict among people? With technologies?

Weisbord’s model is useful when time is limited (i.e., when a simple framework is needed for quick service) and/or when the client is not particularly sophisticated regarding systems thinking.

### 7.2. Nadler–Tushman Congruence Model

This is a more sophisticated model but it is based on the assumption that Weisbord *et al.* posit that an

organization is an open system influenced by its external environment. The primary inputs from the environment are environmental forces, such as competition, resources (e.g., raw materials and/or capital), and history (the organization’s past). Strategy then connects the inputs to the four throughputs: informal organization (e.g., culture), formal organizational arrangements (structure and accountability systems), individual (the people), and task (nature of the work). These throughputs lead to outputs, which amount to performance at three levels: organizational, group, and individual. The primary premise of the model is the importance of congruence, the concept of fit. Therefore, this model goes beyond mere description and takes the position that an organization’s effectiveness is a function of how well the components of the model fit together. For example, is strategy congruent (i.e., responding well) to forces in the external environment? Do employees’ skills and abilities match well with the tasks of the organization? In summary, the components of the Nadler–Tushman model are fairly comprehensive and have good face validity.

### 7.3. Tichy’s TPC Framework

This model is much more explicit about organization change and states that the nine components of the model are “change levers”: (i) external interface, or the organization’s external environment; (ii) mission; (iii) strategy; (iv) managing organizational mission/strategy processes (i.e., realistically engaging the relevant interest groups); (v) task—change often requires new tasks; (vi) prescribed networks—more or less, the formal organizational structure; (vii) organizational processes—communicating, problem solving, and decision making; (viii) people; and (ix) emergent networks—more or less, the informal organization.

TPC stands for technical, political, and cultural. These three “systems” overlay or cut across the nine levers, forming a matrix consisting of subcategories of, for example, the technical system with mission and strategy or the political system with people, etc. The technical system represents rationality and is strongly associated with empiricism and the scientific method. With the political system, Tichy argues that organizations have dominant groups with bargaining being the primary mode of change. The cultural system is the organization’s norms and values as well as shared symbols. Although Tichy also believes that congruence among all these elements is important, he prefers the term alignment. Data are collected and then

categorized within the matrix of the three systems with the nine levers.

### 7.4. Porras's Stream Analysis

Porras developed a framework rooted in open system theory, as are the previously discussed models. His input, of course, is the external environment and the output is defined as both organizational performance and individual development, a unique aspect not included in other models. The throughput consists of two broad categories: organizational members' behavior (individual cognitions and on-the-job behavior) and work setting or context. The work setting is composed of four components that Porras labels as streams: (i) organizing arrangements (goals, strategies, structure, and systems), (ii) social factors (culture, networks, and individual attributes), (iii) technology (equipment, machinery, job design, etc.), and (iv) physical settings (space, interior design, ambiance, etc.). In a sense, Porras's model also concerns congruence. His diagnostic process emphasizes a pictorial chart of how the four main streams flow vis-à-vis one another. His stream diagrams remind one of flowcharts designed by an engineer.

### 7.5. Burke–Litwin Causal Model of Organization Performance and Change

This model is also about theory. Litwin and Burke make the argument that certain organizational variables have more “weight” than others and therefore tend to have effects on less powerful organizational variables. However, the arrows in the model go in both directions and therefore causes and effects are bilateral. Nevertheless, some variables are more powerful than others in causing change. The model is intentionally arranged in a hierarchical manner, with the top boxes referred to as transformational factors (Fig. 4). When practitioners are working to change these categories—mission/strategy, leadership, and culture—the organizational effect is more revolutionary than evolutionary, more discontinuous and episodic than continuous. On the other hand, when practitioners are working to change the categories in the lower remaining part of the model, referred to as the transactional factors, the organizational effect is more continuous than episodic, etc. (Fig. 5).

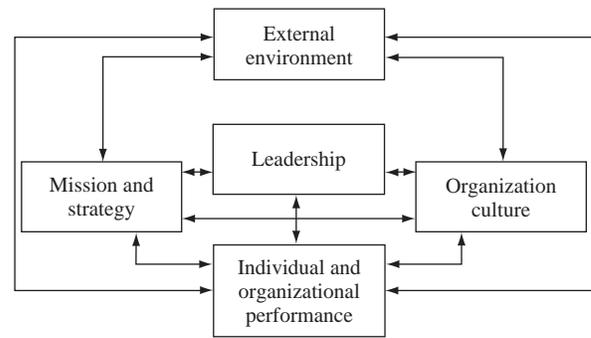


FIGURE 4 The transformational factors. From the Burke–Litwin model of organizational performance and change. © 1987, 1992 by W. Warner Burke, Associates, Inc.

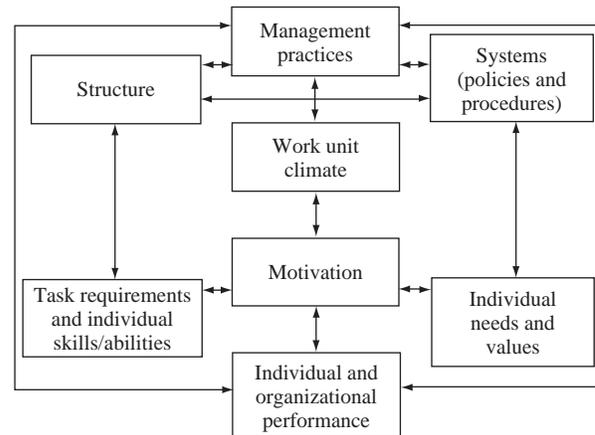


FIGURE 5 The transactional factors. From the Burke–Litwin model of organizational performance and change. © 1987, 1992 by W. Warner Burke, Associates, Inc.

When compared with the transformational categories, the transactional factors represent activities and organizational dimensions that are concerned more with daily operations, are more incremental with respect to change, and are more related to foreground (e.g., work unit climate) than to background (e.g., culture).

In summary, the five models discussed in the previous sections are the most prominent ones in the OD literature, but they may not be the ones used most. Many OD practitioners construct and develop their own models. Regardless of which model is used, more experienced OD practitioners rely on one that is grounded in open system theory and deployed in a manner consistent with action research.

## 8. PLANNING AND MANAGING THE CONSEQUENCES OF CHANGE

It is appropriate to plan change in terms of phases and steps, and it is appropriate to use models that help to frame data and to plan further. However, the actual process of organization change does not occur in the orderly way suggested in this article. This section addresses the realities of organization change.

The paradox of planned organization change is that the change is planned in a linear fashion—that is, phase 1 followed by phase 2, stage 1, 2, 3, etc., or steps 1–7; however, when the change is actually implemented, the process becomes nonlinear. As stated previously, the carrying out of change in an organization is messy: Things do not proceed exactly as planned; people do things their own way, not always according to the plan; some people resist or even sabotage the process; and some people who would be predicted to support or resist the plan actually behave in the opposite way. In short, unanticipated consequences occur. Leaders of change often say something like, “For every step forward we take, we seem to fall back two steps; something always needs fixing to get us back on track.”

Provided the change goals are clear and change leaders are willing to stay the course, over time the process may become somewhat linear, or at least a pattern may emerge. However, linearity is not what anyone experiences during the implementation process, which may seem chaotic, with people in the organization constantly asking “Who’s in charge here?”

With respect to managing the process, change leaders may be operating according to plan, especially with the change goal(s) clearly in mind, but the bulk of their time and energy is spent on dealing with issues and problems that were not planned for, were not anticipated, and must now be addressed and resolved. Otherwise, the change effort will become stymied, people will become frustrated if not disillusioned and eventually cynical, and ultimately true organization change will have failed. If planned and managed well, OD efforts can succeed.

## 9. DOES OD WORK?

This is the basic question that most managers in organizations ask when considering a proposal for getting involved in an OD effort. The answer to the question is “yes,” provided the principles and practices summarized in this article are applied.

With respect to research, some representative and key studies may be considered. In their study of 11 different types of OD initiatives employing a meta-analysis (a statistical procedure for comparing results of a number of studies that investigate related content) of 98 studies, Guzzo *et al.* concluded that “psychologically based organizational interventions,” as they called them, increased worker productivity. Some of the initiatives or interventions studied were job redesign, participative management, team building, and socio-technical systems methods. These interventions are common in OD practice. Also common in evaluating whether OD works are such measures of performance, broadly defined, as worker productivity, employee attitudes, job satisfaction, financial results (e.g., profits and/or cost reduction), customer satisfaction, turnover and absenteeism, quality of output, and degree of innovation.

In another study using a meta-analysis approach of 126 studies, Neuman *et al.* found that OD interventions had a significantly positive effect on worker attitudes and job satisfaction. In addition, their study showed that multiple OD interventions (e.g., intervening simultaneously with team building, leadership development, and an organizational redesign) had greater effect for positive change than relying solely on one OD intervention. In a less rigorous study of seven successful OD efforts, Burke found that the common denominator across all seven was the fact that multiple interventions were used in each case.

### 9.1. Evidence

There is evidence that OD works. However, there are problems concerning this body of work when matters of definition, purpose, and methodology are considered.

When discussing OD, what exactly are we talking about? An intervention or two or a total system change? Different OD interventions may result in different outcomes. With respect to purpose, is the study for knowledge generation or to evaluate the effectiveness of OD?

### 9.2. Methodology

It is not easy to conduct OD research using “normal” or traditional scientific methods. There are important considerations to be made when, for example, there is no control group and when it is not clear which variables of study are independent and which are dependent. Also, many studies have shown that the

researcher and methods used can have powerful consequences (biases) on outcomes.

Although there are complex issues to deal with regarding the determination of the effectiveness of OD, there seems to be cumulative evidence that, on the whole, OD works. Who makes it work? It is, of course, the managers and work unit members in the organization. The OD consultant plays a critical role as well.

### 9.3. The OD Consultant

Often, the term OD practitioner is used instead of OD consultant; in either case, the OD practitioner functions as a consultant. The OD consultant may be either internal (a full-time or part-time employee) or external (a member of a consulting firm, self-employed, or an academic who consults part-time). Although there are differences (e.g., it is difficult for an internal OD person to be a consultant to the CEO of his or her organization and to be as objective about issues in the organization as may be desirable, and the external consultant will never fully understand the culture of the client organization without living in it), the two still operate very similarly regarding the phases of OD practice (entry, contracting, diagnosis, etc.). Ideally, the internal consultant works in tandem with an external OD consultant.

With respect to the role of an OD consultant, Schein's distinctions are clarifying. He described three models or approaches: the purchase model, the doctor-patient model, and the OD model. The purchase model consists of expert services and information that a client organization may buy. For example, a client organization may want a market research study to be conducted and will purchase the services of a firm specializing in this form of expertise. The doctor-patient model means that the client organization calls in the consultant and describes a problem that needs attention, such as a recent loss of market share, an outdated information technology system, or the loss of too many talented people. The client then expects the consultant to provide a remedy. The OD consultant is more facilitative (i.e., helps the client organization to solve its own problems). Schein (1987) describes the process clearly:

It is a key assumption of change that the client must share in the process of diagnosing what may be wrong (or learn to see the problem for himself), and must be actively involved in the process of generating a remedy because only the client ultimately knows what is possible and what will work in his culture and situation. (p. 30)

The primary but not exclusive function of the OD consultant is therefore to help clients learn how to help themselves more effectively. The OD consultant may provide expertise, and even prescribe a remedy, but the role is most often one of being a facilitator (e.g., helping an individual manager to gain more self-insight about his or her impact on others, helping a work unit to operate more as a team, or helping the entire organization to understand the results of a systemwide survey).

### 9.4. Expert Facilitation

Expert facilitation requires certain individual skills and abilities. At least 10 are key:

1. Tolerance of ambiguity: Organizations are not neat and orderly and neither are the data that the OD consultant collects and has to analyze and report to the client.
2. Influence skills: The OD consultant's authority comes from knowledge, of course, but equally as important is the ability to impart that knowledge in a way that makes a difference but is not perceived as arbitrary.
3. Confronting difficult issues: Much of OD practice is bringing to the surface matters that clients are not eager to face.
4. Ability to support and nurture others: This skill is especially critical during times of conflict and stress.
5. Listening well and with empathy: This is particularly important for interviewing and when the client is under stress.
6. Being able to recognize one's feelings and intuitions quickly: This is especially important to discern between one's own perceptions and feelings and those of the client.
7. Conceptualization: This involves being able to express in clear language possible linkages in the organization, such as cause-effect and if-then combinations, and within a systems context.
8. Ability to discover and mobilize human energy: This includes within oneself and the client organization. One should remember that there is energy in resistance; channeling that energy is key.
9. Being a teacher: There are occasions when the OD consultant needs to give a "5-minute lecture" about knowledge that relates directly to the problem at hand.
10. Maintaining a sense of humor: The consultant and/or the client taking themselves too seriously deflects thoughts and energy that could otherwise be

applied more usefully. Also, humor can help to reduce tension when timed well.

This list is a compilation of highly important abilities and skills for the OD consultant. Other abilities could be itemized; this list is a starting point.

### 9.5. Becoming an OD Consultant

Becoming an OD consultant can be considered as a process of following three paths—apprentice, academic education, and professional development. First, there is no substitute for experience. Being an apprentice to someone with considerable OD experience is perhaps the most useful way to learn OD. After all, OD consultants are also known as practitioners, and the more practice the better. Practice alone is not enough, however. Thus, serving as an apprentice is valuable since one can receive feedback on his or her practice from a more experienced practitioner in the field.

Second, regarding academic education, many universities provide a curriculum in OD, some with master's or even doctoral degree programs in OD or in related fields. The following subjects are suggested as a guide for courses to take that would be the most useful:

- A basic, survey course in organizational psychology or organizational behavior
  - Group dynamics
  - Research methods that include both quantitative and qualitative skills—understanding and analyzing data are critical to effective OD practice
  - Adult learning, especially how to learn from experience
    - Career development
    - Counseling and coaching, including interviewing skills
  - Organization development and change—an overview course of the field
    - Training and development—designing programs and other learning activities
    - Action research
    - Human resource management—often the OD function in an organization is a part of the larger human resource or personnel function
      - Consulting skills with emphasis on process consultation
        - Organizational dynamics and theory—how organizations work (and do not work) and a comparison of different types of organizations
        - Functions of organization—learning the “business” of organizations (i.e., marketing, accounting

and finance, strategy, operations, information, technology, budgeting, etc.)

Third, regarding nonacademic training or professional development, the following are suggestions:

- Attend a laboratory, experiential training program to gain greater self-awareness.
- Attend workshops on how to consult; often these are activities that precede large conferences.
- Attend conferences that focus on OD or related practices (e.g., participating in the OD network's annual meeting and regional events).
- Attend specific professional development programs, such as the Principles and Practices of OD and the Advanced Program in OD and Human Resource Management sponsored by Teachers College, Columbia University, or the organization and systems development programs offered by the Gestalt Institute of Cleveland.

Finally, as noted previously, being an OD consultant means being a practitioner. Combining content learning from academic and professional development activities with practice and practice coupled with feedback on this practice are the necessary ingredients for learning and becoming an OD consultant.

## 10. OD VALUES AND CONSIDERATIONS FOR THE FUTURE OF THE FIELD

### 10.1. OD Values

Organization development emerged from beliefs that people in organizations had much of the knowledge required to make the system more effective. It was a matter of eliciting and applying this knowledge. This way of believing and acting was steeped in, if not dependent on, antecedent beliefs in and respect for human dignity, diversity, openness in communication, and the importance of empowering organizational members to take initiative and embrace change. Thus, OD grew from humanistic beliefs. However, there have always been the needs and demands of the organization. Thus, organizational requirements, particularly meeting budget demands and “bottom-line” pressures, are often seen as conflicting with humanistic values. Often, OD consultants spend more time and effort on the “D” and less on the “O,” but both are equally important. Weisbond (1977) addressed this potential conflict when he stated

OD's right goal—its central purpose—grows from its proper setting. If the proper setting is organizations, then there is only one right goal for OD: to confront an issue which most certainly predates the industrial revolution. That is the tension between freedom and constraint. OD's right purpose is to redress the balance between freedom and constraint. (p. 4)

There is always tension between the autonomy of the organizational member and the organization's requirements. The primary mission of the OD consultant, therefore, is to redress the balance whenever either one is out of kilter.

## 10.2. Considerations for the Future of OD

The field of organizational development originated during the late 1950s and early 1960s. OD today is well established. It now seems, however, that OD has reached a rather stagnant state. It is as if all that could be gleaned from the behavioral sciences to apply to organization change and development has been. Undoubtedly this is not true, but in any case, much of the creativity happened in the 1960s and early 1970s, with little innovation since then. Exceptions to this statement include the emergence of appreciative inquiry, self-directed groups, and large group interventions; otherwise, OD appears to be resting on precedence. New ways of thinking about organizations and changing them are needed. Within the field of OD, deeper understandings of the impact of the external environment on organizations and the difficulty of changing organizational culture, executive leadership, and organizational structure are needed. Perhaps of more importance is to infuse new ideas and theory from outside OD that are applicable and relevant nevertheless. An example is theory from the life sciences. The growing knowledge from chaos theory and non-linear complex systems theory may be nascent at present but brimming with ideas for how to think differently about organizational behavior and change. Finally, even from the more popular literature, a book such as that by Gladwell, titled *The Tipping Point*, is loaded with ideas directly applicable to OD.

For OD to survive and thrive in the future, an infusion of new and different thinking is required. The value of continuous or lifelong learning is absolutely necessary, and much of this learning needs to be done outside the behavioral sciences.

## See Also the Following Articles

Environmental Psychology, Overview ■ Industrial/Organizational Psychology, Overview ■ Organization Development ■ Work Motivation

## Further Reading

- Bunker, B. B., & Alban, B. T. (1997). *Large group interventions: Engaging the whole system for rapid change*. San Francisco: Jossey-Bass.
- Burke, W. W. (2001). The broad band of organization development and change: An introduction. In L. Carter, D. Giver, & M. Goldsmith (Eds.), *Best practices in organization development and change* (pp. 3–9). San Francisco: Jossey-Bass/Pfeiffer.
- Burke, W. W. (2002). *Organization change: Theory and practice*. Thousand Oaks, CA: Sage.
- Burke, W. W. (2004). Some key examples of knowledge management. In M. Goldsmith (Ed.), *Leading organizational learning* (pp. 269–280). San Francisco: Jossey-Bass.
- Burke, W. W., & Noumair, D. A. (2002). The role and function of personality assessment in organization development. In J. Waclawski, & A. H. Church (Eds.), *Organization development: Data-driven methods for change* (pp. 55–77). San Francisco: Jossey-Bass.
- Capra, F. (1996). *The web of life*. New York: Anchor.
- Church, A. H., & Waclawski, J. (1998). *Designing and using organizational surveys: A seven-step process*. San Francisco: Jossey-Bass.
- Cummings, T. G., & Worley, C. G. (2005). *Organization development and change* (8th ed.). Mason, OH: South-Western.
- Foster, R. N., & Kaplan, S. (2001). *Creative destruction: Why companies that are built to last underperform the market—and how to successfully transform them*. New York: Currency.
- French, W. L., & Bell, C. H., Jr. (1995). *Organization development: Behavioral science interventions for organization improvement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Gladwell, M. (2000). *The tipping point: How little things can make a big difference*. Boston: Little, Brown.
- Hultman, K. (2002). *Balancing individual and organizational values*. San Francisco: Jossey-Bass/Pfeiffer.
- Kraut, A. I. (Ed.). (1996). *Organizational surveys: Tools for assessment and change*. San Francisco: Jossey-Bass.
- Morgan, G. (1997). *Images of organizations* (2nd ed.). Thousand Oaks, CA: Sage.
- Pascale, R. T., Milleman, M., & Gioga, L. (2000). *Surfing the edge of chaos: The laws of nature and the new laws of business*. New York: Crown Business.
- Schein, E. H. (1987). *Process consultation: Vol. 2. Its role in organization development* (2nd ed.). Reading, MA: Addison-Wesley.

- Schein, E. H., DeLisi, P. S., Kampas, P. J., & Sonduck, M. M. (2003). *DEC is dead, long live DEC: The lasting legacy of Digital Equipment Corporation*. San Francisco: Berrett Koehler.
- Trist, E. (1993). Introduction to Volume II. In E. Trist, & H. Murray (Eds.), *The social engagement of social science. A Tavistock anthology: Vol. 2. The socio-technical perspective* (pp. 36–60). Philadelphia: University of Pennsylvania Press.
- Weisbord, M. (1987). *Productive workplaces: Organizing and managing for dignity, meaning, and community*. San Francisco: Jossey-Bass.
- Zaccaro, S. J. (2001). *The nature of executive leadership: A conceptual and empirical analysis of success*. Washington, DC: American Psychological Association.



# Organizations, Careers in

Ayse Karaevli and Douglas T. Hall

*Boston University, Boston, Massachusetts, USA*

1. Organizational Careers
  2. The New Career Contract: Protean and Boundaryless Careers
  3. Predicting Career Effectiveness
  4. Conclusion
- Further Reading

## GLOSSARY

**adaptability** The capacity and motivation to adapt or change in response to changing task or environmental circumstances.

**boundaryless career** A career not bounded, not tied to a single organization, not represented by an orderly sequence, and marked by less vertical coordination and stability than an organizational career.

**career** In both the popular and behavioral science literature, there are four distinct meanings of the term career: career as advancement, career as profession, career as a lifelong sequence of jobs, and career as a lifelong sequence of role-related experiences. Based on all these different interpretations, career can be defined as the individually perceived sequence of attitudes and behaviors associated with work-related experiences and activities over the span of a person's life.

**identity** The person's awareness of his or her values, interests, abilities, and plans; the degree of integration between past, present, and future concepts of self; and the extent to which the person feels integrated over time.

**organizational career** A career conceived to unfold in a single employment setting.

**protean career** A career based on self-direction in the pursuit of psychological success in one's work.

**psychological success** The feeling of pride and personal accomplishment that comes from knowing that one has done one's "personal best."

This article discusses the general developmental stages that people pass through during the course of their working lives in organizations. The ways in which careers have changed since the mid-1970s are also discussed, with a focus on the changing career contract and the protean, or self-directed, career. It also examines what happens after a career choice has been made, particularly how the effectiveness of a career can be predicted. Performance, career attitudes, adaptability, and identity are discussed as the main dimensions of career effectiveness.

## 1. ORGANIZATIONAL CAREERS

### 1.1. Career and Life Stages

People go through a regular series of life experiences and tasks that constitute general phases of human development. Although in an era of rapid change and complexity the concept of regular and predictable stages or phases has diminished in significance, we have not lost completely the experience of regular, predictable changes in our roles, expectations, concerns, and self-images that are packaged as life and career stages. There are several life and career models that explain these regular stages or phases.

In Erik Erikson's 1963 theory of life stages, there are eight stages of development, each of which is characterized by a particular developmental task that the person must work through before advancing fully to the following stage. In the realm of vocational behavior, Donald Super and associates employed a model of five developmental stages: childhood, adolescence, young adulthood, maturity, and old age. According to their model, in childhood the main task is growth. In the adolescence stage, a person explores his or her interests and different specific career opportunities. In adulthood, a person may initially flounder a bit but eventually establish himself or herself in a particular field. In maturity, when the generativity concerns would probably be the most important, the person continues to hold his or her own in a sort of career plateau. Old age is a period of disengagement when ego integrity is achieved if the person has resolved all the earlier stages.

Psychologist Daniel Levinson and colleagues developed a theory of life stages and transitions based on in-depth clinical case studies of a sample of men and women. They utilized the concept of life structure, which is formed by the individual's sociocultural world, aspects of the self that are played out and those that are inhibited or neglected, and the person's participation in the world. In 1968, Hall and Nougaim identified organizational career stages by using a three-stage model of career development: establishment, advancement, and maintenance. At the establishment stage, people have strong concerns about safety, gaining recognition, and establishing themselves in the organization. The advancement stage is characterized by concerns for promotion and achievement. The maintenance stage represents the onset of a plateau, where the incumbent has cues that he or she is nearing the limit of his or her advancement.

### **1.2. Schein's Model of the Organizational Career**

A model of career development in organizations needs to describe the career from two separate but related perspectives: the career as described by the characteristics and experiences of the person, who moves through the organization, and the career as defined by the organization, which involves policies and expectations. A useful approach for understanding this relationship was proposed by Edgar Schein in 1971, who viewed the organization as a three-

dimensional space like a cone, in which the three dimensions represent

- Vertical: Moving up or down, representing one's changing rank or level in the organization
- Radial: Moving more (or less) "inside" in the system, becoming more (or less) central, part of the "inner circle" acquiring increased (or decreased) influence in the system
- Circumferential: Transferring laterally to a different function, program, or product in the organization

### **1.3. Careers as Learning Cycles**

One of the keys to understanding the new employment contract is the fact that the employees' needs and career concerns change during the course of the career in a much more dynamic way than in the past. Therefore, in Douglas Hall's 1996 concept of the flexible career or protean career, the career is not measured by age-linked life stages but, rather, by a series of short (3–5 year) learning cycles. Each learning cycle has the following phases: exploration, trial of new activity, mastery, and then exploration of some other new activity. Because the life cycle of technologies and products is short, so too are personal mastery cycles. As a result, people's careers will increasingly become a succession of "ministages" (or short-cycle learning stages) of exploration–trial–mastery–exit as they move in and out of various product areas, technologies, functions, organizations, and other work environments.

## **2. THE NEW CAREER CONTRACT: PROTEAN AND BOUNDARYLESS CAREERS**

### **2.1. The Protean Career Contract**

Another way to frame the changes in career patterns is in terms of the psychological contract between the employee and the organization. Ian MacNeil discussed two forms of what he called the social contract. The first was relational, which was based on assumptions of a long-term relationship and trust that the relationship would be a mutual one. In contrast, the other form was transactional, based on a shorter term exchange of benefits and contributions. MacNeil was discussing the role of an individual in a larger society, but his concepts seem applicable to organizations as well.

From the perspective of the individual, there is a shift from the organizational career to what Hall described as the protean career—a career based on self-direction in the pursuit of psychological success in one's work. The driving forces in the protean career are the person's core values and need for control of his or her life and career choices. In addition, the protean person has a strong learning or exploratory orientation that would predispose him or her to be quite comfortable in a career characterized by continuous learning cycles, as described previously.

The measure of success for the protean careerist is psychological success, or success as measured by the person's own goals and values. Thus, in Everett Hughes' terms, to understand the protean career, the observer must examine the subjective career, which is the way the individual experiences his or her career. In contrast, the more traditional organizational career was often assessed from the perspective of the objective career, which is the way career success would be judged by an external observer, usually in terms of outcomes such as money, power, or position.

## 2.2. The Boundaryless Career

In 1996, Michael Arthur and Denise Rousseau identified a related concept, the boundaryless career. They defined the boundaryless career as not bounded, not tied to a single organization, not represented by an orderly sequence, and marked by less vertical coordination and stability. In their words, "boundaryless careers are the opposite of 'organizational careers'—careers conceived to unfold in a single employment setting" (p. 5). In 1993, Defillipi and Arthur offered a congruent explanation of the boundaryless career that evolved around interfirm mobility. Peiperl and Baruch, in their discussion of the "post-corporate career," agreed that the emerging career clearly centered outside the organization—increasingly identifiable with professional and familial as opposed to organizational communities.

Thus, if the old contract was with the organization, in the protean career the contract is with the self and one's work. In a study comparing data collected in 1978 and 1989, Linda Stroh and colleagues found evidence of this shift from an organizational focus to an investment in one's own work; satisfaction with the company decreased from 1978 to 1989, but job involvement and job satisfaction increased. (They also reported that managers in 1989 were changing jobs and relocating more often than those in 1978.)

## 2.3. Mobility via Contingent Employment

One of the ways that career mobility and protean careers are currently manifest is in the rise of contingent employment. In the United States, the largest employer at the time of this writing is a temporary agency, Manpower, Inc. More than 90% of U.S. employers use temporary workers as part of their staffing mix, as identified by von Hippel *et al.* in 1997, and nearly 3 million people worked as temporary workers in a single year. It is clear that one of the major ways that careers have become boundaryless has been through the practice of temporary employment.

## 2.4. Life and Work Roles

Another way that change and flexibility manifest themselves in contemporary careers is in the interplay between the person's work role and the roles in his or her private life. There are three common forms of stress between work and home, and it is important to consider each one separately. (There is a huge literature on work-life issues, so this discussion is simplified by considering married couples.) The first and most common source is role overload. This is caused by the sheer number and intensity of demands on a couples' time. Overload pressure is especially high for young couples (at a career stage in which financial resources are tight) with children. The sheer number of demands on the partners exceeds the time and energy to do them.

The second type of work-life stress derives from role conflict, the incompatibility of different role expectations. One type of conflict is interrole conflict, in which one partner's different roles make mutually conflicting demands, such as when a person, as an employee, is expected to be out of town for a client meeting and is also expected, as a parent, to be at his or her child's school for the fifth-grade play. It is not possible to be in both places at the same time (although people may be tempted to do so electronically). There is also intersender conflict, within a role. An example is when a father is home taking care of his daughter and she wants him to let her stay up past her bedtime to watch a special TV show, even though the mother had left strict instructions that the daughter should be in bed on time.

A third type of conflict is intrapersonal conflict (within the person), in which a person feels torn between two valued activities (e.g., going to the son's band concert or to the daughter's all-star softball game). It could also be a conflict of unmet expectations, in which the person

feels that he or she is simply not measuring up to the standards set by other people he or she respects (“I could never be the father that he is” or “I’ll never be as successful as she is”).

The third source of stress is change. This is especially problematic for dual-career couples, who may be more likely to experience relocations more often than one-career families. A relocation can trigger all sorts of other changes—new jobs, new home, new friends, new schools, new community, new culture, often a more expensive house and the concomitant financial stress, etc. Again, synchronicity is an issue: The more roles that are changing simultaneously, the more the stress is compounded.

### 3. PREDICTING CAREER EFFECTIVENESS

In the research literature, career effectiveness has been defined in terms of performance and the popular symbols of success—money and position. However, there are other important measures of career effectiveness: career attitudes, adaptability, and identity. These four dimensions of career effectiveness can be considered as representing a  $2 \times 2$  grid, as shown in Table I, where one axis involves the time span (short term vs long term) and the other involves the focus (task vs personal) aspect of career development.

#### 3.1. Performance

The major category of performance effectiveness is related to typical financial indicators, such as current salary, average yearly salary increases, or salary in relation to that of other people with equivalent length of service or of similar age. The other major category involves the position, which has been employed as rank or level in an organizational hierarchy, number of promotions received over a given time period (i.e., rate of advancement), or, conversely, length of time in current position.

TABLE I  
Four Types of Career Effectiveness<sup>a</sup>

Focus	Time span	
	Short term	Long term
Task	Performance	Adaptability
Self	Attitudes	Identity

<sup>a</sup>Adapted from Hall (2002).

#### 3.2. Career Attitudes

The way the career is perceived and evaluated by the individual defines an individual’s attitudes toward his or her career. The major attitudes related to career are organizational identification and commitment, which tap the psychological involvement that the person feels with the organization and the extent to which his or her organizational membership is a significant aspect of his or her identity; the involvement and commitment to career and family and negative experiences related to involvement and commitment, such as burn-out; the psychological success, which is the feeling of pride and personal accomplishment that comes from knowing that one has done one’s personal best, as Hall and Mirvis discussed; and self-efficacy, which is the individual’s beliefs about his or her ability to perform successfully a given task, as identified by Bandura.

#### 3.3. Adaptability

Morrison and Hall define adaptability as the capacity and motivation to adapt or change in response to changing task or environmental circumstances. Hall proposed adaptability as a higher order quality called metacompetency—the capacity to master many more specific skills when one masters this metacompetency.

#### 3.4. Identity

An individual’s sense of identity can be defined as (i) the person’s awareness of his or her values, interests, abilities, and plans; (ii) the degree of integration between past, present, and future concepts of self; and (iii) the extent to which the person feels integrated over time. Role transitions, socialization, life stage, as well as race, culture, and history are critical factors that facilitate the growth of the person’s identity. There are two facets of the person’s overall sense of self. Personal identity refers to those qualities in the person’s self-perceptions that are unique to him or her. Social identity describes the portion of the overall identity that derives from the person’s membership in a particular social group, such as a gender group or a racial or ethnic group.

### 4. CONCLUSION

In the postindustrial era, individual career effectiveness entails more than traditional “objective” indicators of

success, such as money and position. In a knowledge economy, the rapid rate of change puts positive attitudes, identity growth, and adaptability at a premium. These are all qualities that are more “subjective,” or internal and unique to the person. The ultimate evaluation of these career outcomes (i.e., career success) can be assessed from the individual’s perspective. This individual or subjective definition of success is called psychological success.

All these personal career outcomes, including psychological success, are developable—that is, they can be influenced by the individual and the organization. Also, they tend to go together and to reinforce one another. For example, people with positive attitudes tend to make jobs into satisfying and developmental experiences; that is, they adapt effectively to frustrations in the work environment. Knowing oneself enables the person to make better choices that are in line with the person’s interests and values and to know whether it is better to make adaptations or to move on and select a new place to work. In turn, effective adaptation creates new skills and relationships, which in turn lead to a more positive and proactive self-identity.

### Acknowledgment

Portions of this article are based on and adapted from Hall (2002).

### See Also the Following Articles

Career Counseling ■ Holland’s Theory (Vocational Personality Types) ■ Indecision, Vocational ■ Vocational Assessment in Schools ■ Vocational Interests ■ Vocational Psychology, Overview ■ Work Role, Values Sought in the

### Further Reading

- Arthur, M. M., & Rousseau, D. M. (Eds.). (1966). *The boundaryless career: A new employment principle for a new organizational era*. New York: Oxford University Press.
- Hall, D. T. (1996). Protean careers of the 21st century. *Academy of Management Executive*, 10(4), 8–16.
- Hall, D. T. (2002). *Careers in and out of organizations*. Thousand Oaks, CA: Sage.
- Hall, D. T., & Mirvis, P. H. (1996). The new protean career: Psychological success and the path with a heart. In D. T. Hall et al. (Eds.), *The career is dead—Long live the career: A relational approach to careers* (pp. 15–45). San Francisco: Jossey-Bass.
- Levinson, D. J. (1986). A conception of adult development. *American Psychologist*, 41, 3–13.
- Levinson, D. J. (with Levinson, J.). (1996). *The seasons of a woman’s life*. New York: Knopf.
- Levinson, D. J., Darrow, C. N., Klein, E. B., Levinson, M. H., & McKee, B. (1978). *The seasons of a man’s life*. New York: Knopf.
- MacNeil, I. R. (1980). *The new social contract*. New Haven, CT: Yale University Press.
- Peiperl, M., & Baruch, Y. (1997, Spring). Back to square zero: The post-corporate career. *Organizational Dynamics*, 23–34.
- Schein, E. H. (1971). The individual, the organization, and the career: A conceptual scheme. *Journal of Applied Psychology*, 7, 401–426.
- Stroh, L. K., Brett, J. M., & Reilly, A. H. (1994). A decade of change: Managers’ attachment to their organizations and their jobs. *Human Resource Management*, 33, 531–548.
- Super, D. E. (1992). Toward a comprehensive theory of career development. In D. H. Montross, & C. J. Shinkman (Eds.), *Career development: The theory and practice* (pp. 35–64). Springfield, IL: Thomas.
- Super, D. E., & Bohn, M. J., Jr. (1970). *Occupational psychology*. Belmont, CA: Wadsworth.
- Wiese, B. E., Freund, A. M., & Baltes, P. B. (2002). Subjective career success and emotional well-being: Longitudinal predictive power of selection, optimization, and compensation. *Journal of Vocational Behavior*, 60, 321–335.



# Overtraining and Burnout in Sports

Michael Kellmann

Ruhr-University of Bochum, Bochum, Germany

1. Introduction
  2. Recovery
  3. Overtraining
  4. Burnout
  5. Interrelation of Stress States and Recovery Demands
  6. Treatment of Overtraining and Burnout
- Further Reading

## GLOSSARY

**burnout** A psychological, physical, and emotional withdrawal from a formerly enjoyable and motivating activity resulting from prolonged or chronic stress.

**long-term overtraining** During long-term overtraining, lasting 3 weeks or more, athletes are on a chronic performance plateau that cannot be influenced positively by short amounts of rest and recovery periods.

**overtraining** This is due to a long-term imbalance between stress and recovery (i.e., too much stress combined with too little regeneration).

**recovery** An inter- and intraindividual multilevel (e.g., psychological, physiological, and social) process for the reestablishment of performance abilities. Recovery includes an action-oriented component, and self-initiated activities (proactive recovery) can be systematically used to optimize situational conditions to build up and to restore personal resources and buffers.

**short-term overtraining** Short-term overtraining, lasting less than 3 weeks, is a regular part of athletic training that leads to a state of overreaching in affected athletes. This state of overreaching is characterized by a transient

underperformance that is reversible within a short-term recovery period of 1 or 2 weeks and can be rewarded by an increase in performance ability.

This article addresses underrecovery as a key component in the development of overtraining and burnout. A constant lack of recovery or disturbed recovery results in overtraining or burnout, which frequently occur in sports. Even being slightly underrecovered over a longer period of time results in underperformance.

## 1. INTRODUCTION

At the 2002 Soccer World Championships in Japan and Korea, teams that were favored to win the World Cup, such as France and Portugal, did not survive the first round in the tournament. During and immediately after the World Cup, some journalists and the World Soccer Association (FIFA) started a discussion about why these teams were not able to perform at their expected level. A connection was quickly made to burnout, overtraining, and underrecovery due to the game schedule of the players during the season. Especially athletes from successful clubs playing in international competitions, such as the Champion League or the Uefa-Cup in Europe, have a very tight competition calendar. For example, Hertha BSC played more games than any of the other German professional soccer teams in the first half of the 1999/2000 season—29 games in 18 weeks, which is an

average of 1 game every 4.3 days, including travel time. By midseason, the players described themselves as physically and mentally drained. In addition to this schedule, some players are drafted to their national teams and have to play for them in addition to playing regular club games. Therefore, a common tendency seems to be that players purposely miss games of their national teams by claiming an injury to reduce the number of games they must play and to get more time off. Interestingly, often it is not the training and competition that are perceived to be stressful by the players but, rather, traveling and the continuous adaptation to new setting and surroundings. The schedules of professional sports teams in North America, such as basketball, baseball, and hockey teams, are similar to or even more intense than those of European soccer teams. In high-performance sports, especially professional sports, the drills performed in practice are often driven by the competition schedule so that there is only time for regenerative practice. For example, National Hockey League (NHL) teams played 82 games during the regular season in 2000/2001. According to the official game schedule, the Colorado Avalanche, like other NHL teams, played an average of 1 game every 2.3 days (including travel time), and reaching the playoffs added more games to the schedule. How much recovery time is left between games? When 2 or 3 games are played per week, physical and mental recovery is often incomplete. However, lack of recovery or underrecovery is not only due to the frequency of competitions. It can also be enforced by training mistakes, such as (i) having a monotonous training program, (ii) more than 3 hours of training per day, (iii) more than a 30% increase in training load each week, (iv) ignoring the training principle of alternating difficult and easy training days or by alternating 2 difficult days followed by an easy training day, (v) no training periodization and respective regeneration microcycles after 2 or 3 weeks of training, or (vi) no rest days.

## 2. RECOVERY

Often, recovery is defined as the compensation of deficit states of an organism (e.g., fatigue or decrease in performance) and, according to the homeostatic principle, a reestablishment of the initial state. In 2001, Kellmann and Kallus stated that recovery encompasses active processes of reestablishing psychological and physiological resources and states that allow one to tax these resources again. This more precisely describes the complex issue of recovery. However, it

is still not sufficient to define recovery as an elimination of fatigue or system restart. Kellmann und Kallus developed a more precise definition:

Recovery is an inter- and intraindividual multilevel (e.g., psychological, physiological, social) process in time for the reestablishment of performance abilities. Recovery includes an action-oriented component, and those self-initiated activities (proactive recovery) can be systematically used to optimize situational conditions to build up and to refill personal resources and buffers. (p. 22)

Recovery processes in competitive sports support the restoration of individual activity conditions and well-being following training and competition demands. Several studies have shown that recovery is faster when subjects exercise moderately instead of resting passively. Active recovery consists of cool-down activities, such as muscle relaxation and stretching, after practice and competition. The purpose of these exercises is to eliminate the effects of fatigue through physical activity. Passive recovery, on the other hand, includes treatments such as massages, hot and cold baths, steam baths, and sauna baths, which initiate physiological reactions through physiological stimuli (heat, cold, and pressure), affecting blood flow, respiration rate, and muscle tone. Optimal use of available recovery time is imperative for success in sports. Top performances can only be achieved by athletes who can recover quickly during competition and optimally carry out the changes between stress, recovery, and future stress.

## 3. OVERTRAINING

Kellmann summarized the literature and stated that 60% of females and 64% of males have been overtrained at least once during their running careers. Ninety-one percent of swimmers who were overtrained during their freshmen year in college also experienced overtraining during the following years, whereas only 30% of athletes who were not overtrained in their freshman year had problems later. In addition, overtraining was seen in 33% of Indian national-level basketball players during a 6-week training camp, in 21% of swimmers on the Australian National Team during a 6-month season, and in more than 50% of semiprofessional soccer players after a 5-month competitive season.

Gould and Dieffenbach pointed out the relevance of overtraining in high-performance sports and its importance as a performance-influencing factor identified in

several studies of high-performance athletes. They summarized a study on positive and negative factors influencing U.S. athletes and coaches at the Olympics in Atlanta in which 84 (28%) of 298 athletes reported that they had overtrained for the games and this overtraining had a negative impact on performance. Similarly, 35 of the athletes identified overtraining/not getting enough rest as the primary coaching action that negatively affected their performance. Ten percent of U.S. athletes who participated in the 1998 Nagano Winter Olympics reported that they were overtrained and that this had negative effects on their performance. The need to taper, rest, not to overtrain, travel less, and stay healthy were changes that they would employ if they could prepare again for the Olympics. Hence, these findings demonstrate that both summer and winter Olympic Games athletes identified overtraining as a concern. Gould and Dieffenbach note that the findings should not be assumed to result only from inappropriate physical training. Other factors, such as psychological stress, travel, personality, inadequate rest, type of recovery activity, and sociological issues must be examined in multifaceted models as well.

It is well accepted that overtraining is due to a long-term imbalance between stress and recovery (i.e., too much stress combined with too little regeneration). The term stress includes all training, competition, and additional nontraining stress factors. Social, educational, occupational, economical, nutritional, and travel aspects, time stress, and the monotony of training increase the risk of an overtraining syndrome.

The same training can have different effects on athletes depending on their personal situation, including emotional stressors (e.g., fights with friends or partners and also parents getting divorced), obligations at school, or difficulties in time management (e.g., practice/school/friends). Often, individuals can easily handle these situations, but when a heavy training load is added to an already high "personal package load," the total impact can be too high. Therefore, coaches should be aware of athletes' personal situations and design training accordingly. The most frequent causes of overtraining cited by athletes are (i) too much stress and pressure, (ii) too much practice and physical training, (iii) physical exhaustion and all-over soreness, (iv) boredom due to too much repetition, and (v) poor rest or lack of proper sleep. Therefore, training holism encompasses two ideas: Training must be balanced and varied, and non-training time has a major influence on training. It is therefore important that all factors outside the realm of the training session be evaluated with regard to their possible negative influence on total fatigue. Insufficient

and/or lack of recovery time between practices is the main cause of overtraining syndrome. Factors such as nutrition, sleep deficit, sickness, travel, and competitions increase the negative effect of insufficient recovery.

In 2002, Kellmann discussed the different uses of the term overtraining and related terms in the literature and concluded that the definition by Lehmann *et al.* provides the best integration and description of overtraining. Lehmann and colleagues distinguished between short-term overtraining, which lasts less than 3 weeks, and long-term overtraining that lasts 3 weeks or more. They noted that short-term overtraining (also called overreaching or supercompensation training) is a frequent part of athletic training, which leads to a state of overreaching in affected athletes. This state of overreaching is characterized by a transient underperformance that is reversible within a short-term recovery period of 1 or 2 weeks and can be rewarded by a state of supercompensation (increase in performance ability after 1 or 2 regeneration weeks after a short-term phase of overtraining). Therefore, short-term overtraining or overreaching is a regular part of athletic training and desired by coaches. When overreaching is done inappropriately or done for too long, short-term overtraining turns into long-term overtraining.

During long-term overtraining, athletes are on a chronic performance plateau that cannot be influenced positively by short periods of rest and recovery. Long-term overtraining can be characterized by an inability to train at the customary level. The symptoms include feelings of depression, general apathy, decreased self-esteem, emotional instability, irritability, impaired performance, lack of supercompensation, restlessness, disturbed sleep, weight loss, loss of appetite, increased resting heart rate, increased vulnerability to injuries, and hormonal changes. An important clinical feature of overtraining syndrome is the increased susceptibility to infections with corresponding symptoms, suggesting some kind of impaired immune response. Unfortunately, when athletes realize that performance has stagnated or declined, it is a common tendency to increase the training effort, which leads to an even deeper stage of long-term overtraining.

#### 4. BURNOUT

Although there is no common definition for burnout, there is a consensus that burnout is a reaction to chronic stress. Smith defines burnout as a psychological, physical, and emotional withdrawal from a formerly

enjoyable and motivating activity resulting from prolonged or chronic stress. It is an exhaustive psychophysiological response exhibited as a result of frequent, sometimes extreme, and generally ineffective efforts to meet excessive physical training and psychological demands. It is also important to note that when athletes burnout of sport they do not always discontinue involvement altogether. Some athletes who experience burnout remain active. They believe that they cannot discontinue for various reasons (e.g., family pressure), and this entrapped participation is characterized by lower motivation, emotional detachment, less satisfaction, and poor performance. Burnout results in a decline in sport participation by an athlete as a result of chronic excess stress.

Maslach defines burnout as an individual syndrome including emotions, attitudes, motives, and expectancies. Burnout is a negative individual experience encompassing problems, distress, negative mood states, dysfunction, and negative consequences. Maslach and Jackson characterize the burnout syndrome through emotional exhaustion (i.e., loss of social interest) and reduced personal accomplishment and involvement. In addition to a lack of performance, burnout eventually results in the quitting of prior enjoyable activities. Research on burnout was originally performed in the area of social professions, and it was expanded to other professional groups, such as athletes and coaches. With an increasing level of burnout, the performance of athletes becomes inconsistent and eventually declines. The effect is supported by overtraining, routine, monotony, boredom, and a lack of diversity. This indicates that burnout is caused by a lack of recovery. In the case of an intense schedule during the season and intense training during the off season, a complete physical and psychological recovery seems to be impossible. Therefore, a long recovery phase or a temporary reduction in athletic activities are the only effective approaches to prevent athletes from dropping out of sport. Consequently, coaches and staff need to be sensitive to the recovery demands of athletes.

On the behavioral level, burned out athletes have less input with regard to training, fewer practice days, differences in motivation, poorer coping behavior, and a longer history within the sport. Gould *et al.* showed for tennis players that Smith's model well described the antecedents and consequences of burnout. Thus, one can view a model of chronic stress as a good basis for explaining burnout in athletes. Cohn identified too much practice or competition, lack of enjoyment, and too much pressure from self or others as the main sources of burnout. Furthermore, stress associated

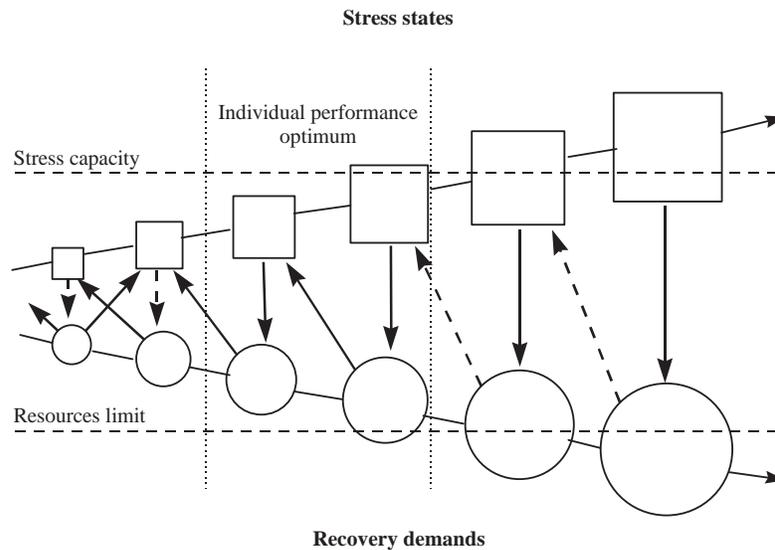
with injuries and resumption of training can drastically increase the risk of dropout and burnout.

Burnout in coaches has also been examined. The work environment of coaches involves long hours, demands much mental and emotional energy, and exerts considerable subjective pressure due to standards set by the coaches or by fans and athletes. A survey by Vealey *et al.* demonstrated the importance of the coach's personality as a factor in the development of burnout. The results also supported the contribution of reward factors as buffers against burnout, which can be derived from Smith's model. Perceived rewards and perceived value of the social role, together with excitement, are the most important antiburnout buffers, as evidenced by research in organizations.

In 1996, Kallus *et al.* reported results from a study of 195 semiprofessional and professional coaches. They found (i) differences in the coping strategies of younger coaches as well as between coaches of team and individual sports; (ii) coaches of individual sports who used more cognitive reevaluation strategies, actively and positively coped with stress, and searched for self-affirmation pitied themselves more than coaches from team sports; and (iii) full-time coaches were more burned out and highly stressed, were experiencing less recovery, and were less fit than semiprofessional coaches. In addition, professional coaches showed deficits in coping strategies and therefore may possibly react less effectively in critical situations. It can be concluded that stress in addition to a recovery deficit and a coping deficit leads to a high risk of burnout. More research is needed to fully support the results of Kallus *et al.*; however, these results strongly indicate that recovery is a key variable in the development of burnout.

## 5. INTERRELATION OF STRESS STATES AND RECOVERY DEMANDS

The previous discussion has shown that recovery is the key factor in preventing overtraining and burnout. In this context, Kellmann proposed a general model that describes the interrelations of stress states and recovery demands (Fig. 1). The basic assumption of this model is that with increasing stress, increased recovery is necessary to remain at the same stress state. However, limited resources (e.g., time) initiate a vicious cycle: Under increased stress and due to the inability to meet increased recovery demands, a person experiences more stress. Recovery demands are defined as



**FIGURE 1** The “scissors model” of the interrelation of stress states and recovery demands. Reprinted, by permission, from K. W. Kallus & M. Kellmann, 2000, “Burnout in athletes and coaches” in *Emotions in sport*, edited by Y. L. Hanin (Champaign, IL: Human Kinetics), p. 212.

the quality and/or quantity of recovery activities required to level out the current recovery stress state. People may be stressed to the point that they fail to find or make time to recover adequately or to consider better ways of coping with the situation.

With intermediate levels of stress, one can find an area of optimal performance and thus an area of adequate recovery. Beyond this point, one cannot meet recovery demands without additional recovery activities. Stress will accumulate, and without intervention burnout symptoms are likely to develop. The state of balanced stress and recovery is related to optimal performance. In a state of adequate recovery, the individual can react appropriately and cope successfully with stress without additional recovery activities. A lack of recovery or underrecovery can trigger a process that leads to a state of elevated stress. Because increasing stress limits the possibility of recovery, the athlete must be given opportunities to recover in order to reestablish an optimal level of performance.

With regard to sports, the model may also explain how overtraining develops. The axis of the stress states can be seen as a continuum of an increasing training load, which can be labeled at the end points “no training” and “overtraining.” With extended training load, the organismic recovery demands increase proportionally. A short-term planned period of recovery enhances long-term performance effects (e.g., supercompensation). If the

training load and intensity increase over a longer time period without adequate or with merely inappropriate recovery, the individual experiences long-term underrecovery, which may result in the overtraining syndrome. To reach the optimal recovery stress state, athletes have to increase their self-initiated activities to fulfill their recovery demands. At each state of the model, recovery can work as a regulation mechanism, which is caused by an increasing distance between the two axes into a longer recovery period (days to weeks). The higher a person is on the stress states or the more extensive the overtraining syndrome, the more recovery efforts are needed to reach the individual optimal recovery stress state. The model interrelation of stress states and recovery demands implies that it is not necessarily bad to be highly stressed as long as a person knows how to recover.

## 6. TREATMENT OF OVERTRAINING AND BURNOUT

When performance plateaus or instability on performance occur, athletes and coaches have the tendency to increase their effort and the training load, which mostly leads to a vicious cycle. The best treatment for overtraining is rest and recovery phases from 2 to 6 weeks long (sometimes longer), depending on the degree of overtraining. In some cases of overtraining syndrome,

athletes may require months without any training and without any physical activity in order to recover completely. Changes in training can also positively affect the practice routine. A reduction of training load, a change in intensity, or a change in training technique can positively affect athletes. When a coach believes that an athlete is showing signs of overtraining, the athlete should consult a doctor to rule out injuries and damage due to overstrain. If overtraining syndrome is diagnosed, treatment with electrolytes can be considered. Active and proactive recovery is the best treatment for overtraining syndrome. Physical exercises, such as gymnastics, games, regenerative runs, or swimming, as well as a balanced and healthy diet play an important role in the recovery process. Environment and climate changes help athletes to recover both physically and mentally. Activities that were denied because of the rigid training regime can be important too, and psychoregulative techniques, such as progressive muscle relaxation or autogenic training, can accelerate the recovery process.

In the case of burnout, psychoregulative techniques also play an important role, enabling people to deal with chronic stress. Psychoregulative techniques enhance the ability of athletes to have appropriate reaction patterns when needed. Also, when someone is burned out, the setting of short-term and long-term goals is important to gain perspective on a new or formerly enjoyable activity. Setting short-term goals and providing rewards or incentives every time a goal is accomplished is motivating. In addition, having fun goals, especially during the end of a long season, is very beneficial. These goals should be set for competition as well as for practice, especially since practice often comprises more than 90% of the participation time in sports. Gould and colleagues' recommendations for dealing with and preventing burnout include playing for personal reasons, balancing sport with other activities such as school clubs or socializing with friends, stopping participation if the sport is no longer fun to play, focusing on making the sport fun, doing things to relax, and taking time off to recover.

### See Also the Following Articles

- Performance Slumps in Sport: Prevention and Coping
- Psychological Skills Training in Sport

### Further Reading

- Cohn, P. J. (1990). An exploratory study of sources of stress and burnout in youth golf. *The Sport Psychologist*, 4, 95–106.
- Gould, D., & Dieffenbach, K. (2002). Overtraining, under-recovery, and burnout in sport. In M. Kellmann (Ed.), *Enhancing recovery: Preventing underperformance in athletes* (pp. 25–35). Champaign, IL: Human Kinetics.
- Gould, D., Tuffey, S., Udry, E., & Loehr, J. (1997). Burnout in competitive junior tennis players: III. Individual differences in the burnout experience. *The Sport Psychologist*, 11, 257–276.
- Kallus, K. W., Kellmann, M., Eberspächer, H., & Hermann, H.-D. (1996). Beanspruchung, Erholung und Streßbewältigung von Trainern im Leistungssport [Strain, recovery, and coping with stress of coaches in elite sports]. *Psychologie und Sport*, 3, 114–126.
- Kellmann, M. (Ed.). (2002). *Enhancing recovery: Preventing underperformance in athletes*. Champaign, IL: Human Kinetics.
- Kellmann, M., & Kallus, K. W. (2001). *Recovery–Stress Questionnaire for athletes; User manual*. Champaign, IL: Human Kinetics.
- Kreider, R. B., Fry, A. C., & O'Toole, M. L. (Eds.). (1998). *Overtraining in sport*. Champaign, IL: Human Kinetics.
- Lehmann, M., Foster, C., Gastmann, U., Keizer, H. A., & Steinacker, J. M. (Eds.). (1999). *Overload, fatigue, performance incompetence, and regeneration in sport*. New York: Plenum.
- Maslach, C., & Jackson, S. E. (1986). *Maslach burnout inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Smith, R. E. (1986). Toward a cognitive–affective model of athletic burnout. *Journal of Sports Psychology*, 8, 36–50.
- Vealey, R. S., Udry, E. M., Zimmerman, V., & Soliday, J. (1992). Intrapersonal and situational predictors of coaching burnout. *Journal of Sports and Exercise Psychology*, 14, 40–58.



# Pain Management

Dennis C. Turk

University of Washington, Seattle, Washington, USA

1. Pain: Magnitude of the Problem
  2. Pharmacological and Surgical Treatments for Pain
  3. Factors Influencing the Perception of Pain
  4. Psychological Treatment of Chronic Pain Patients
  5. Interdisciplinary Pain Rehabilitation Programs
  6. Subgroups of Chronic Pain Sufferers
  7. Concluding Comments
- Further Reading

## GLOSSARY

**biofeedback** A procedure that combines voluntary control of autonomic nervous system activity with continuous recording and presentation of information (“feedback”) regarding activity of (e.g., muscles) believed to be involved in various pain syndromes; the physiological information provided to the patient by means of electrical activity converted into signals, such as tones and lights, is used by the patient to help him or her learn how to control the autonomic activity being monitored.

**cognitive-behavior therapy (applied to chronic pain)** Involves multiple interventions that target behavioral contributors to pain and the maladaptive thoughts that can accompany adjustment to any chronic physical condition; the focus of therapy centers around four central components: (a) education to increase the patient’s knowledge about his or her condition, reassurance, and reduction of fear; (b) behavioral exercises and activities designed to increase physical and functional activities such as pacing programs to enhance physical exercise and activity; (c) cognitive exercises and activities designed to help the patient alter responses to stressors and pain; and (d) skills training to help the patient

deal with problems created by a chronic disease and to modify the perception of pain using techniques such as relaxation, biofeedback, and other self-control strategies.

**implantable drug delivery systems** Consist of the implantation of a pump and reservoir containing analgesic medication (usually opioids); the pump provides a steady administration of the medication following a preprogrammed schedule directly into the spinal canal that may be more potent than oral administration.

**interdisciplinary pain rehabilitation programs** Comprehensive treatment programs that integrate physical and occupational therapy, medication management, vocational counseling, and psychological (e.g., operant conditioning, cognitive-behavior therapy) components, with an emphasis on self-management and functioning improvements despite residual pain rather than cure.

**operant conditioning (applied to chronic pain)** Has the main focus of modifying the frequency of a given behavior; if the consequence of the given behavior is rewarding, the likelihood of its occurrence increases, whereas if the consequence is aversive, the likelihood of its occurrence decreases.

**spinal cord stimulation** Involves surgical placement of electrodes along the spinal cord connected to an external device capable of generating a current along the electrodes, with the current designed to interrupt or mask the transmission of noxious sensations from the periphery to the brain.

Pain is an extremely common perceptual event that is essential for survival. However, when pain persists for extended periods of time, it no longer serves as a warning function, it affects all aspects of a person’s functioning, and it can become a problem in its own

right. Despite advances in understanding of the anatomy and physiology of pain and the development of treatments based on sophisticated technology, chronic pain continues to be a prevalent and costly problem. Because chronic pain is a perceptual event, psychosocial and behavioral factors as well as physical ones contribute to the experience. Over time, cognitive, affective, and environmental factors play an increasingly greater role in the maintenance of pain, emotional distress, and functional disability. An understanding of the role of psychological factors has resulted in innovative treatment approaches that have been shown to have significant effects on pain intensity, associated physical symptoms, and physical functioning. These approaches are not cures; rather, they focus on functional restoration, alleviation of emotional distress, and reduction of disability. The effectiveness of psychological treatments for diverse chronic pain syndromes has been demonstrated in numerous studies. Moreover, psychological principles and techniques have been incorporated into successful interdisciplinary pain rehabilitation programs.

### **1. PAIN: MAGNITUDE OF THE PROBLEM**

In 1998, Abbott and Fraser noted that pain is the single most frequent reason for physician consultation, with nearly one-half of those who seek treatment reporting pain as their primary symptom. Precise estimates of the prevalence of chronic pain syndromes are difficult to ascertain. In 1994, Joranson and Lietman indicated that there may be more than 30 million people with chronic or recurrent painful conditions in the United States. Based on a survey conducted by Marketdata Enterprises in 1995, an estimated 2.9 million Americans (1.1% of the population) are treated annually by health care professionals specializing in chronic pain.

The U.S. Bureau of the Census in 1996, as well as the National Research Council in 2001, estimated that the total annual costs of chronic pain (e.g., treatment, lost work days, disability payments, legal fees) in the United States range from \$150 billion to \$215 billion. In 1995, Cousins suggested that the costs of health care for patients with chronic pain might exceed the combined costs of treating patients with coronary artery disease, cancer, and acquired immune deficiency syndrome (AIDS).

### **2. PHARMACOLOGICAL AND SURGICAL TREATMENTS FOR PAIN**

The first-line treatment for pain consists of a host of pharmacological agents. In 1998, Schappert noted that pain medications were the second most prescribed drugs (after cardiac-renal drugs) during visits to physicians' offices and emergency rooms, and the results of the National Ambulatory Care Medical Survey conducted in 1998 suggested that pain accounted for 12% of all medications prescribed during ambulatory office visits in the United States. Pharmaceutical industry data provided in a personal communication indicate that more than 312 million prescriptions for analgesic medications were written in the year 2000 in the United States, that is, more than one prescription for analgesics for every man, woman, and child.

The most frequently used medications to control pain in patients with chronic pain are nonsteroidal anti-inflammatory drugs, acetaminophen, opioid and nonopioid analgesics, low-dose tricyclics, and anti-convulsants. Despite their frequent use, currently available medications do not eliminate pain. For example, in 2002, Turk reviewed the available literature and concluded that the average pain reduction for patients placed on the most potent agents, opioids, is approximately 35%.

Persistence of some chronic pain syndromes (e.g., back pain) frequently leads to surgery. However, a number of studies, including those by North and colleagues in 1991 and Franklin and colleagues in 1994, reveal that significant pain may persist following surgery and that a significant percentage of chronic pain patients treated with surgery report that their pain is actually worse following surgery. Several studies of implantable devices (e.g., spinal cord stimulators, implantable drug administration systems) have reported impressive results in pain reduction, but only for carefully selected patients with pain of long duration.

### **3. FACTORS INFLUENCING THE PERCEPTION OF PAIN**

There has been a growing recognition that pain is a complex perceptual experience influenced by a wide range of psychosocial factors, including emotions, social and environmental contexts, sociocultural background, the meaning of pain to the person, and beliefs, attitudes, and expectations, as well as by

biological factors. Biological factors may initiate, maintain, and modulate physical perturbations, but psychological factors influence the appraisal and perception of internal physiological signs and social factors shape the behavioral responses of patients to the perceptions of their physical perturbations and may have a direct effect on physiological processes associated with exacerbation and maintenance of symptoms.

Pain that persists for months or years, called chronic pain, will have an impact on all aspects of a person's functioning—emotional, interpersonal, avocational, and physical. Consequently, treating chronic pain patients successfully requires attention not only to the organic basis of the symptoms but also to the range of factors that modulate nociception and moderate the pain experience and related disability.

#### 4. PSYCHOLOGICAL TREATMENT OF CHRONIC PAIN PATIENTS

A large volume of research exists demonstrating the efficacy of the psychological treatments for low back pain, fibromyalgia syndrome, noncardiac chest pain, arthritis, headaches, temporomandibular disorder (TMD), and whiplash-associated disorders, to name a few. Furthermore, cognitive-behavioral approaches appear to prevent the development of disability due to pain. Psychological modalities have been used in the treatment of chronic pain patients either on their own or, more typically, as an essential component of comprehensive interdisciplinary rehabilitation.

Table I lists the most common psychological approaches applied to chronic pain. The effectiveness of relaxation, biofeedback, and cognitive-behavior therapy (CBT) has been the most thoroughly evaluated. Electromyographic (EMG) biofeedback is a procedure that combines muscle relaxation with continuous recording and presentation of information (i.e., “feedback”) regarding activity of muscles believed to be involved in different pain syndromes. In 1979, Turk and colleagues noted that biofeedback is based on the idea that stress-induced hyperactivity is often an important component in muscular pain syndromes. Biofeedback makes use of surface records of muscle activity to provide patients with information about levels of muscle tension. Levels of muscular arousal are recorded and converted into signals (usually auditory or visual) that permit patients to observe their current levels of muscle tension. The signals provided indicate changes

TABLE I  
Treatments for Chronic Pain

Pharmacological
Nonsteroidal anti-inflammatory agents
Acetaminophen
Opioids
Antidepressants
Anticonvulsants (e.g., gabapentin)
Topical agents (e.g., lidocaine)
Physical modalities (e.g., hot packs, ultrasound, exercise, transcutaneous nerve stimulation)
Surgery
Regional anesthesia (e.g., neural blockade, epidural steroid injections)
Spinal cord stimulators
Psychological
Relaxation
Electromyographic biofeedback
Hypnosis
Counseling (e.g., family, marital)
Behavior modification (e.g., exposure, desensitization, contingency contracting, operant conditioning)
Cognitive-behavior therapy
Complementary and alternative medicine (e.g., acupuncture, chiropractic, diet supplements, Tai chi)

in muscular activity resulting from patients' efforts to relax the muscles. Thus, the information about muscular activity is “fed back” to the patients and assists them in learning to alter high levels of muscular tension that may be contributing to the onset, maintenance, or exacerbation of pain.

The main focus of operant conditioning is modification of the frequency of a given behavior. If the consequence of the given behavior is rewarding, the likelihood of its occurrence increases; if the consequence is aversive, the likelihood of its occurrence decreases. Behaviors associated with pain, such as limping and moaning, are called “pain behaviors”—overt expressions of pain, distress, and suffering. When an individual is exposed to a stimulus that causes tissue damage, the immediate behavior is withdrawal in an attempt to escape from noxious sensations. Such pain behaviors are adaptive and appropriate. Over time, however, environmental contingencies of reinforcement may serve to maintain the pain behaviors long after the initial cause of the pain has been resolved (Table II). The behaviors are

**TABLE II**  
Operant Conditioning: Modification of Behavior

Schedule	Consequences	Probability of the behavior recurring
Positive reinforcement	Reward the behavior	More likely
Negative reinforcement	Prevent or withdraw aversive results	More likely
Punishment	Punish the behavior	Less likely
Neglect	Prevent or withdraw positive results	Less likely

conditioned based on their consequences. Often, the consequences are provided by significant others, including health care providers, who may inadvertently reinforce the pain behaviors by providing attention, treatment, and/or avoidance of activities that are undesirable to the patient. The operant learning paradigm does not uncover the etiology of pain but rather focuses primarily on the maintenance of pain behaviors and on deficiency in well behaviors such as activity. Treatment focuses on extinction of pain behaviors and on increased well behaviors by changing the reinforcement contingencies—withdrawal of rewards for pain behaviors and increased rewards for well behaviors.

#### 4.1. Cognitive–Behavior Therapy: Components

CBT programs usually involve multiple interventions that target behavioral contributors to pain and the maladaptive thoughts that can accompany adjustment to any chronic physical condition. In 1999, Turk and Okifuji found that these programs focus on four primary components: (a) education to increase patients' knowledge about their conditions, reassurance, and reduction of fear; (b) behavioral exercises and activities designed to increase physical and functional activities such as pacing programs to enhance physical exercise and activity; (c) cognitive exercises and activities designed to help patients alter their responses to stressors and pain; and (d) skills training to help patients deal with problems created by chronic diseases and to modify the perception of pain using techniques such as relaxation, biofeedback, and other self-control strategies (Table III).

**TABLE III**  
Components of Cognitive–Behavior Therapy

- Provide education about pain and particular syndrome.
- Encourage acceptance.
- Foster a self-management perspective.
- Focus on function rather than cure.
- Assist in goal setting.
- Teach relaxation.
- Provide information and guidance about pacing and increasing activities.
- Provide guidance on ways in which to improve sleep.
- Emphasize identifying and eliminating maladaptive thoughts.
- Provide strategies to help cope with relapse.
- Provide skills and skills training for ways in which to cope with interpersonal problems.
- Discuss appropriate uses of medications, exercise, and physical modalities (e.g., heat, cold).

#### 4.2. Cognitive–Behavior Therapy: Efficacy

In 1999, Morley and colleagues conducted a systematic review and meta-analysis of randomized controlled trials, comparing the effectiveness of CBT to waiting list controls and alternative treatment control conditions. They found that CBT produced significantly greater changes in pain experience, cognitive coping, and reduced behavioral expressions of pain. However, significant differences were not found for dysphoric mood, negative appraisals, and social role functioning.

### 5. INTERDISCIPLINARY PAIN REHABILITATION PROGRAMS

The large body of evidence demonstrating the importance of psychosocial and behavioral factors has led to the proliferation of interdisciplinary pain rehabilitation programs (IPRPs). In 2001, Loeser and Turk noted that these IPRPs typically include psychological (e.g., operant conditioning, CBT) components as integrated within comprehensive treatment plans that include physical and occupational therapy, medication management, and vocational counseling.

In 1992, Flor and colleagues concluded that IPRPs are more effective than no treatment, waiting list controls, and monodisciplinary treatments in terms of reductions of health care use and medication,

increased activity and return to work, closure of disability claims, and reductions of affective distress. The failure to find significant improvement in affective distress, noted by Morley and colleagues in 1999, is at variance with the results of this meta-analysis. The explanation for this inconsistency is unclear and may involve differences in the effects of psychological treatments on their own compared with when they are integrated within IPRPs. IPRPs are often referred to as being based on operant, behavioral, or cognitive-behavioral principles. The actual treatments are broadly defined and often include diverse ingredients. It is difficult to determine what the active components of the treatment packages actually are. Turk in 1990, as well as Kole-Snijders and colleagues in 1999, indicated that this is important because it is not cost-effective to provide comprehensive treatments with multiple components, some of which may be unnecessary. Patients with different characteristics may benefit from different treatments, and not all of the components included may be necessary for all patients.

The reduction of pain following treatment at IPRPs has been reported to be statistically significant in meta-analyses conducted by Flor and colleagues in 1992. Flor and colleagues reported the mean pain reduction for patients treated at IPRPs at 37%. However, the majority of patients continue to experience considerable pain. Interestingly, Flor and colleagues found that the pain reduction achieved at IPRPs, which is comparable to the pain reduction reported with opioids, is accompanied with a significant decrease (63%) in prescription pain medication. However, even when successful, people continue to experience significant levels of pain with which they must cope.

## 6. SUBGROUPS OF CHRONIC PAIN SUFFERERS

Although biobehavioral treatments appear to be important components of rehabilitation treatments for chronic pain patients, not all patients benefit equally. Focusing solely on group effects may mask important issues related to the characteristics of patients who respond successfully to a treatment. Chronic pain syndromes are made up of heterogeneous groups of people even if the individuals have the identical medical diagnosis. A common pitfall in clinical research and practice is the assumption of patient homogeneity. Inclusion of a diverse group of patients into the same

category simply because they present a common set of symptoms may result in inconsistent research results, as is commonly observed in the pain literature.

Several studies, including one by Kight and colleagues in 1999 and two by Turk and colleagues in 1998, have identified subgroups of patients based on psychosocial and behavioral characteristics. In 1997, Dahlstrom and colleagues found that when patients were classified into different subgroups based on their psychosocial and behavioral responses, they responded differentially to treatments. Similarly, Turk and colleagues noted differential responses to a common treatment for patients with distinctive psychological characteristics but identical physical diagnoses. The results of these studies could be used for developing treatments that are matched to patients' characteristics. Various studies comparing the subgroups have yielded evidence supporting differential responses to the same intervention.

## 7. CONCLUDING COMMENTS

Despite the advances in the knowledge of sensory physiology, anatomy, and biochemistry and of the prevalence of chronic pain, and despite the many treatments available to provide relief (see Table I), total elimination of pain is elusive for many pain sufferers. Pain and disability have not been eliminated as problems for a significant portion of the population and for society as a whole. Comprehensive biobehavioral treatments appear to have the best outcomes for patients with many chronic pain syndromes. Combining CBT with more traditional medical, surgical, and pharmacological modalities appears to be the most reasonable approach. Matching treatments with patient characteristics has the potential to improve outcomes and reduce costs.

## Acknowledgments

Preparation of this article was supported in part from grants provided by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (AR/AI44724 and AR47298), the National Institute of Child Health and Human Development/National Center for Medical Rehabilitation Research (HD33989), and the Fetzer Institute.

## See Also the Following Articles

Coping ■ Psychological Rehabilitation Therapies

### Further Reading

- Abbott, F. V., & Fraser, M. I. (1998). Use and abuse of over-the-counter analgesic agents. *Journal of Psychiatry and Neuroscience*, 23, 13–34.
- Cousins, M. J. (1995). Foreword. In W. E. Fordyce (Ed.), *Back pain in the workplace: Management of disability in nonspecific conditions* (p. ix). Seattle, WA: IASP Press.
- Dahlstrom, L., Widmark, G., & Carlsson, S. G. (1997). Cognitive-behavioral profiles among different categories of orofacial pain patients: Diagnostic and treatment implications. *European Journal of Oral Science*, 105, 377–383.
- Dworkin, S. F. (1995). Behavioral characteristics of chronic temporomandibular disorders: Diagnosis and assessment. In B. J. Sessle, P. S. Bryant, & R. A. Dionne (Eds.), *Progress in pain research and management* (pp. 175–192). Seattle, WA: IASP Press.
- Dworkin, S. F., Turner, J. A., Wilson, L., Massoth, D., Whitney, C., Huggins, K. H., Burgess, J., Sommers, E., & Truelove, E. (1994). Brief group cognitive-behavioral intervention for temporomandibular disorders. *Pain*, 59, 175–187.
- Epker, J., & Gatchel, R. J. (2000). Coping profile differences in the biopsychosocial functioning of patients with temporomandibular disorders. *Psychosomatic Medicine*, 62, 69–75.
- Flor, H., Fydrich, T., & Turk, D. C. (1992). Efficacy of multidisciplinary pain treatment centers: A meta-analytic review. *Pain*, 49, 221–230.
- Franklin, G. M., Haug, J., & Heyer, N. J. (1994). Outcome of lumbar fusion in Washington State Workers' Compensation. *Spine*, 19, 1897–1904.
- Gatchel, R. J., & Turk, D. C. (1999). *Psychosocial factors in pain: Critical perspectives*. New York: Guilford.
- Goossens, M., Rutten-Van Molken, M., Kole-Snijders, A., Vlaeyen, J., van Breukelen, G., & Leidl, R. (1998). Health economic assessment of behavioral rehabilitation in chronic low back pain: A randomized clinical trial. *Health Economics*, 7, 39–51.
- Hasenbring, M., Ulrich, H. W., Hartmann, M., & Soyka, D. (1999). The efficacy of a risk factor-based cognitive behavioral intervention and electromyographic biofeedback in patients with acute sciatic pain: An attempt to prevent chronicity. *Spine*, 24, 2525–2535.
- Hildebrandt, J., Pflingsten, M., Saur, P., & Jansen, J. (1997). Prediction of success from a multidisciplinary program for chronic low back pain. *Spine*, 22, 990–1001.
- Holroyd, K., & Lipchik, G. (1999). Psychological management of recurrent headache disorders: Progress and prospects. In R. J. Gatchel, & D. C. Turk (Eds.), *Psychosocial factors in pain: Critical perspectives* (pp. 193–212). New York: Guilford.
- Joranson, D. E., & Lietman, R. (1994). *The McNeil National Pain Survey*. New York: Louis Harris & Associates.
- Kight, M., Gatchel, R. J., & Wesley, L. (1999). Temporomandibular disorders: Evidence for significant overlap with psychopathology. *Health Psychology*, 18, 177–182.
- Kole-Snijders, A. M. J., Vlaeyen, J. W. S., Goossens, M. E. J. B., Rutten-van Molken, M. P. M. H., Heuts, P. H. T. G., van Breukelen, G., & van Eck, H. (1999). Chronic low-back pain: What does cognitive coping skills training add to operant behavioral treatment? Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 67, 931–944.
- Loeser, J. D., & Turk, D. C. (2001). Multidisciplinary pain management. In J. D. Loeser, S. D. Butler, C. R. Chapman, & D. C. Turk (Eds.), *Bonica's management of pain* (3rd ed., pp. 2069–2087). Philadelphia: Lippincott, Williams, & Wilkins.
- Lorig, K., Mazonson, P., & Holman, H. (1993). Evidence suggesting that health education for self-management in patients with chronic arthritis has sustained health benefits while reducing health care costs. *Arthritis & Rheumatism*, 36, 439–446.
- Main, C. J., & Spanswick, C. C. (2000). *Pain management: An interdisciplinary approach*. London: Churchill-Livingstone.
- Marketdata Enterprises. (1995). *Chronic pain management programs: A market analysis*. Valley Stream, NY: Author.
- Mayou, R. A., Bryant, M. M., Sanders, D., Bass, C., Klimes, I., & Farfar, C. (1997). A controlled trial of cognitive-behavioural therapy for non-cardiac chest pain. *Psychological Medicine*, 27, 1021–1032.
- Morley, S., Eccleston, C., & Williams, A. (1999). Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. *Pain*, 80, 1–13.
- National Center for Health Statistics. (1998). *National Ambulatory Medical Care Survey, 1998*. Washington, DC: Department of Health and Human Services.
- National Research Council. (2001). *Musculoskeletal disorders and the workplace*. Washington, DC: National Academy Press.
- North, R. G., Ewend, M. G., & Lawton, M. T. (1991). Failed back surgery syndrome: Five-year follow-up after spinal cord stimulator implantation. *Neurosurgery*, 28, 692–699.
- Schappert, S. M. (1998). *Ambulatory care visits to physicians' offices, hospital outpatient departments, and emergency departments: United States, 1996* (Vital and Health Statistics, Series 13, No. 134). Atlanta, GA: Centers for Disease Control and Prevention.
- Turk, D. C. (1990). Customizing treatment for chronic pain patients: Who, what, and why. *Clinical Journal of Pain*, 6, 255–270.
- Turk, D. C. (2002). Clinical effectiveness and cost effectiveness of treatments for chronic pain patients. *Clinical Journal of Pain*, 18, 355–365.
- Turk, D. C., & Gatchel, R. J. (2002). *Psychological approaches to pain management: A practitioner's handbook* (2nd ed.). New York: Guilford.

- Turk, D. C., Meichenbaum, D., & Berman, W. H. (1979). The application of biofeedback for the regulation of pain: A critical review. *Psychological Bulletin*, 86, 1322–1338.
- Turk, D. C., & Melzack, R. (2001). *Handbook of pain assessment* (2nd ed.). New York: Guilford.
- Turk, D. C., & Okifuji, A. (1999). A cognitive-behavioral approach to pain management. In P. D. Wall, & R. Melzack (Eds.), *Textbook of pain* (4th ed., pp. 1431–1444). London: Churchill-Livingstone.
- Turk, D. C., & Okifuji, A. (2002). Psychological factors in chronic pain: Evolution and revolution. *Journal of Consulting and Clinical Psychology*, 70, 278–290.
- Turk, D. C., Okifuji, A., Sinclair, J. D., & Starz, T. W. (1998). Interdisciplinary treatment for fibromyalgia syndrome: Clinical and statistical significance. *Arthritis Care and Research*, 11, 186–195.
- Turk, D. C., Okifuji, A., Starz, T. W., & Sinclair, J. D. (1998). Differential responses by psychosocial subgroups of fibromyalgia syndrome patients to an interdisciplinary treatment. *Arthritis Care and Research*, 11, 397–404.
- U.S. Bureau of the Census. (1996). *Statistical abstract of the United States: 1996* (116th ed.). Washington, DC: Author.
- Vendrig, A. A., Van Akkerveeken, P. F., & McWhorter, K. R. (2000). Results of multimodal treatment program for patients with chronic symptoms after whiplash injury of the neck. *Spine*, 25, 238–244.



# Panic

Martin M. Antony

St. Joseph's Healthcare and McMaster University, Hamilton, Ontario, Canada

1. Introduction
  2. Definitions and Description
  3. Epidemiology of Panic Disorder
  4. Causes of Panic Disorder
  5. Treatment of Panic Disorder
- Further Reading

Panic is an emotional state associated with extreme fear and discomfort, a sense of dread or apprehension, and a number of intense physical sensations, such as racing heart, sweating, shaking, shortness of breath, and various other symptoms of arousal. It tends to begin suddenly, is often triggered by a feared situation or object, and may occur out of the blue, without any obvious trigger or cause.

## GLOSSARY

**agoraphobia** A fear of situations in which escape might be difficult, or in which help might be unavailable, in the event of experiencing a panic attack or panic-like sensations.

**anxiety sensitivity** A fear of anxiety-related symptoms, such as dizziness, shortness of breath, and racing heart.

**cognitive behavior therapy** A form of psychotherapy used to change negative patterns of thinking and maladaptive behaviors using cognitive restructuring, exposure-based interventions, relaxation training, and other methods.

**interoceptive exposure** Exposing oneself to feared physical symptoms using exercises designed to trigger the sensations.

**panic attack** A rush of fear or discomfort that peaks quickly and that is associated with intense symptoms of physical arousal, including racing heart, shaking, and feeling lightheaded.

**panic disorder** An anxiety disorder in which an individual experiences recurrent unexpected panic attacks, accompanied by anxiety over having more attacks, worry about the consequences of the attacks, and changes in behavior aimed at preventing future attacks.

## 1. INTRODUCTION

The word panic has its origins in Greek mythology. It refers to Pan, the ancient Greek god of woods, fields, and flocks, who has a human torso and head, and the legs, horns, and ears of a goat. Pan was believed to be very mischievous, often inciting terror among travelers making their way through the forest by rustling trees and bushes. Pan's victims probably experienced many of the same sensations that people report during modern day panic attacks.

The experience of panic is often referred to as a "fight or flight" response, in which the body is mobilized to facilitate immediate escape from a perceived threat, or to meet the threat head-on with an aggressive response. During panic, the heart races to get blood to the large muscles that may be needed in the event of an escape. Breathing becomes more rapid in order to increase the flow of oxygen to the bodily tissues. Perspiration increases to cool off the body so it can perform more efficiently. In fact, almost all of the

symptoms that occur during a typical episode of panic can be explained in terms of their adaptive functions. In other words, the symptoms of panic serve to improve an individual's ability to escape from perceived danger.

However, panic is often not adaptive or helpful. In many cases, panic occurs in the absence of true danger or threat. Further, panic attacks may lead individuals to avoid feared situations, and may therefore cause considerable impairment in functioning. As a result, frequent panic attacks often lead an individual to seek treatment in one form or another.

## 2. DEFINITIONS AND DESCRIPTION

In the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), a panic attack is described as a discrete period of intense fear or discomfort that peaks in 10 minutes or less, and is accompanied by at least four of the following 13 symptoms: (1) racing or pounding heart, (2) sweating, (3) trembling or shaking, (4) shortness of breath, (5) feeling of choking, (6) chest pain or discomfort, (7) nausea or abdominal distress, (8) feeling dizzy, unsteady, or faint, (9) feeling unreal or detached, (10) paresthesias (i.e., numbness or tingling sensations), (11), chills or hot flushes, (12) fear of dying, and (13) fear of going crazy or losing control.

Panic attacks are usually triggered by a feared object, situation, or thought. For example, a person who fears dogs may experience a panic attack each time he or she encounters a dog, or panic attacks may occur in anticipation of giving a speech for someone who is frightened of public speaking. Similarly, a person with obsessive compulsive disorder may experience panic attacks in response to obsessional thoughts (e.g., irrational thoughts about harming another individual).

Although panic attacks are often triggered, they can also appear unexpectedly or out of the blue. The presence of recurrent, unexpected panic attacks is the hallmark feature of panic disorder. In addition, people with panic disorder tend to be fearful of experiencing panic attacks and the symptoms that are associated with their attacks. They often anticipate the attacks, wondering when the next one will occur. They also are concerned about the possible consequences of the attacks, worrying about a physical catastrophe such as a heart attack, a stroke, dying, or fainting, or about an emotional or psychological catastrophe such as losing control, embarrassing oneself, or "going crazy."

Finally, people with panic disorder often change their behavior in response to the attacks. They may actively avoid situations where the attacks occur, and they may do things to protect themselves from experiencing further attacks (e.g., distraction, avoiding being alone, checking their body for panic symptoms, carrying medication). To receive a diagnosis of panic disorder, it must also be established that the attacks are not due to a medical illness (e.g., cardiac disease or hyperthyroidism) or a substance (e.g., cocaine use, caffeine overuse, alcohol withdrawal).

Most people with panic disorder also develop some degree of agoraphobia. Agoraphobia is a fear of situations in which escape might be difficult, or in which help might be unavailable, in the event of experiencing a panic attack or panic-like sensations. Typical situations avoided by people with agoraphobia include driving, flying, crowds, enclosed places, public transportation, malls, supermarkets, and traveling. In extreme cases, people with agoraphobia may avoid leaving the house altogether. Often, agoraphobic fear is milder when the individual is accompanied by a close friend, relative, or spouse.

## 3. EPIDEMIOLOGY OF PANIC DISORDER

Studies on the prevalence of panic disorder estimate that between 2 and 5% of individuals in the general population suffer from this problem, although up to one-third of individuals may experience a panic attack from time to time. Panic disorder occurs more frequently in women than in men. In addition, women with panic disorder tend to experience more severe symptoms (e.g., more intense and frequent attacks, more severe agoraphobic avoidance) than men with panic disorder. This condition most often begins in early adulthood (in a person's 20s or early 30s), though it can begin at any age, ranging from childhood to late in life. The onset often occurs following a period of stressful life events. Panic disorder occurs across ethnic and religious groups and does not appear to be significantly associated with intelligence, marital status, education, or socioeconomic status.

Panic disorder is often accompanied by significant functional impairment and leads to considerable costs to society in the form of increased healthcare utilization, inability to work, and decreased productivity at work. Without treatment, the course is usually

chronic, with up to 90% of individuals still experiencing panic attacks a year after their initial assessment. Panic disorder often occurs concurrently with other psychological problems, including other anxiety disorders and depression. The rate of substance use disorders is also higher in people with panic disorder than it is in the general population.

#### 4. CAUSES OF PANIC DISORDER

There is evidence that both psychological and biological factors contribute to the development and maintenance of panic disorder. From a psychological perspective, cognitive factors seem to play an important role. A risk factor for the development of panic disorder is heightened anxiety sensitivity (i.e., a fear of experiencing symptoms of anxiety and arousal). In addition, people with panic disorder tend to misinterpret the meaning of their physical symptoms, and often assume that their symptoms are dangerous when in fact they are not. Panic disorder is also associated with a tendency to attend to threat-related information (e.g., to scan the body for uncomfortable symptoms) and to remember information that is consistent with anxiety-related beliefs and assumptions. It is believed that the tendency to pay attention to threat-related cues and to interpret benign physical symptoms as dangerous contributes to the development of panic disorder and to the chronic course of this condition.

From a biological perspective, a number of different factors have been studied and found to be relevant. Neurotransmitters in the brain (e.g., norepinephrine, serotonin, cholecystokinin) may play a role. For example, substances that increase levels of norepinephrine or cholecystokinin in the brain (including injections of cholecystokinin and inhalation of carbon dioxide-enriched air) have been found to trigger panic attacks, and medications that block the re-uptake of norepinephrine and serotonin are effective for preventing panic attacks. Genetics also influences the development of this condition. Panic disorder tends to run in families, and a portion of the transmission of panic disorder from generation to generation is directly attributable to the effects of genetics. Finally, brain imaging studies suggest that certain areas in the brain (e.g., the hippocampus) may be more or less active depending on whether someone has panic disorder. The nature of these differences is still not fully understood, and research findings are often inconsistent.

#### 5. TREATMENT OF PANIC DISORDER

Evidence-based approaches to treating panic disorder include cognitive-behavior therapy (CBT), medications, or a combination of these approaches. On average, these options are about equally effective in the short term. However, any one individual may benefit more from one approach over another. In the long term, there is evidence favoring CBT. Rates of relapse and recurrence tend to be lower following the discontinuation of CBT than they are following the discontinuation of medication. Although most people benefit considerably from these treatments, fewer than half of individuals experience a nearly complete remission of their symptoms. Occasional panic attacks, anxiety, and avoidance are common following the end of treatment.

CBT is designed to help an individual to change the beliefs and behaviors that are thought to maintain the problem. The client is taught to replace anxious and unrealistic thoughts with more realistic beliefs, by critically examining the evidence for both anxious and non-anxious interpretations. Clients are also taught to confront feared situations by engaging in prolonged exposure on a regular basis (e.g., practicing driving for an hour each day) until the situations no longer trigger fear. They also learn to confront the fear of physical sensations through interoceptive exposure, that is, exposure to arousal symptoms (e.g., spinning to induce dizziness, aerobic exercise to increase one's heart rate). Finally, relaxation-based strategies (especially slow diaphragmatic breathing) are sometimes used to reduce the intensity of physical symptoms, although it is not clear that these strategies provide any benefit over and above the effects of the other CBT techniques. Usually, education about the nature and treatment of panic and anxiety is included throughout the treatment.

The most commonly used medications for panic disorder include certain antidepressants as well as a number of anxiolytics. Effective antidepressants include the selective serotonin re-uptake inhibitors (e.g., citalopram, fluoxetine, paroxetine, sertraline), certain tricyclic antidepressants (e.g., imipramine, clomipramine), as well as other medications (e.g., venlafaxine). Anxiolytics, such as the benzodiazepines (e.g., alprazolam, clonazepam) have also been shown to reduce the symptoms of panic disorder. Decisions regarding which medication to use should be based on the individual's past treatment history, medical history, possible interactions between the medication and other drugs

that person may be taking, potential side effects, and other factors.

### See Also the Following Articles

Anxiety Disorders in Late Life ■ Test Anxiety

### Further Reading

- Antony, M. M. (2001). Measures for panic disorder and agoraphobia. In M. M. Antony, S. M. Orsillo, & L. Roemer (Eds.), *Practitioner's guide to empirically-based measures of anxiety* (pp. 95–126). New York, NY: Kluwer Academic/Plenum.
- Antony, M. M., & Swinson, R. P. (2000). *Phobic disorders and panic in adults: A guide to assessment and treatment*. Washington, DC: American Psychological Association.
- Craske, M. G., & Barlow, D. H. (2000). *Mastery of your anxiety and panic*, 3rd ed. Boulder, CO: Graywind Publications, Inc.
- Craske, M. G., & Barlow, D. H. (2001). Panic disorder and agoraphobia. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders*, 3rd ed. New York: Guilford Press.
- McNally, R. J. (1994). *Panic disorder: A critical analysis*. New York: Guilford Press.
- Pollard, C. A., & Zuercher-White, E. (2003). *The agoraphobia workbook: A comprehensive program to end your fear of symptom attacks*. Oakland, CA: New Harbinger.
- Rachman, S., & de Silva, P. (1996). *Panic disorder: The facts*. New York, NY: Oxford University Press.
- Rosenbaum, J. F., & Pollack, M. H. (1998). *Panic disorder and its treatment*. New York, NY: Marcel Dekker.
- Taylor, S. (2000). *Understanding and treating panic disorder: Cognitive and behavioral approaches*. Chichester, UK: Wiley.
- Zuercher-White, E. (1997). *An end to panic: Breakthrough techniques for overcoming panic disorder*, 2nd ed. Oakland, CA: New Harbinger Publications.
- Zuercher-White, E. (1997). *Treating panic disorder and agoraphobia: A step by step clinical Guide*. Oakland, CA: New Harbinger Publications.



# Paralinguistic Behaviors and Culture

Hede Helfrich

Universität Hildesheim, Hildesheim, Germany

1. Introduction
2. The Nature of Paralinguistic Behaviors
3. The Relevance of Culture
4. Evolutionary Continuity in Communication
5. Paradigms for the Study of Paralinguistic Behaviors
6. Externalization of Paralinguistic Behaviors
7. Accuracy of Listeners' Inferences
8. Use of Paralinguistic Cues by Listeners
9. Conclusion

Further Reading

**silence** The absence of vocal behavior for a period of time; often, it is misunderstood to refer to the absence of any and all communication.

**tone of voice** Pitch, loudness, and timbre of a spoken utterance.

**turn taking** The allocation of speaker and listener roles in a conversation; a person takes his or her turn when adopting the role of the speaker.

## GLOSSARY

**baby talk** A voice and speech style used by adults, and even by children, when speaking to infants.

**contact cultures** Cultures from which people stand closer, face more directly, touch, and look more than do people from noncontact cultures.

**distance between bodies** Physical space between people in their conversations; there are wide cultural differences in what is considered appropriate distance.

**facial expression** Patterns of movements of brow, eye, and mouth regions.

**paralinguistic behaviors** Vocal and nonvocal phenomena that are either blended or associated with speech utterances but have no direct function according to linguistic rules; they include tone of voice, speech and conversation style, facial expressions, gestures, head and body movements, and distance between bodies.

Successful communication includes knowing when to speak and when to remain silent; which tone of voice, which facial expression, and which body position are appropriate in various contexts; what the routines for turn taking are in conversation; and how to exhibit the appropriate listener signals. These behaviors are referred to as "paralinguistic." Some of them may be common to all people, and some of them may vary from one culture to another. When communicating with persons who are culturally different, violations of the rules for paralinguistic behaviors often lead to misunderstandings, even if the words and grammar of the foreign language are appropriately mastered. The aim of the cross-cultural study of paralinguistic behaviors is to learn how the principles of paralinguistic behaviors vary from one culture to another and how universal principles must be qualified and adjusted on the basis of cultural differences. This understanding will help people to use communication effectively in intercultural encounters.

## 1. INTRODUCTION

One of the most striking differences among members of various cultures is that they speak different languages. Thus, it will obviously be useful to persons involved in international transactions to learn the languages of their interlocutors. However, successful communication involves knowing not only what to say to whom but also how to say it appropriately in any given situation. It includes knowing when to speak and when to remain silent; which tone of voice, which facial expression, and which body position are appropriate in various contexts; what the routines for turn taking are in conversation; how to request, how to offer, and how to decline assistance or cooperation; and how to exhibit the appropriate listener signals. These behaviors are not regarded by linguists as being an integral part of the verbal utterances with which they are associated; therefore, they are referred to as “paralinguistic.” Some of these behaviors may be universal (common to all people), and some of them may vary from one culture to another. As a person learns his or her “mother tongue” during childhood, that person usually acquires and reproduces the appropriate paralinguistic behaviors typical of speakers of that language because a first language is learned in the context of behavioral settings where the norms of the culture are acted and enunciated. However, these norms are less explicit and less under conscious control than are the rules of languages. Thus, learning a second language has traditionally stressed the acquisition of words and grammar while ignoring paralinguistic behaviors. When communicating with persons who are culturally different, violations of the rules for paralinguistic behaviors often lead to misunderstandings, even if the words and grammar of the foreign language are appropriately mastered. The aim of the cross-cultural study of paralinguistic behaviors is to learn how the principles of paralinguistic behaviors vary from one culture to another and how universal principles must be qualified and adjusted on the basis of cultural differences. This understanding will help people to use communication effectively in intercultural encounters.

## 2. THE NATURE OF PARALINGUISTIC BEHAVIORS

### 2.1. Description of Phenomena

Speech utterances not only convey information via spoken words and grammatical structure but also

include phenomena that, according to linguistic rules, have no direct function. These phenomena refer to the voice in which words are uttered and to their patterning in time. In face-to-face communication, speech utterances are accompanied by various body movements such as face and head movements, gestures, body postures, and body orientation. These vocal and nonvocal phenomena, which are either blended or associated with speech utterances, are called “paralinguistic behaviors.” Not all vocal modifications of spoken words are nonverbal; linguists distinguish between prosodic and paralinguistic features. Prosodic features, such as stress, tone, and juncture, follow the rules of language and affect the linguistic meaning of utterances. Paralinguistic features convey other kinds of information. They not only refer to the speech behavior of a single speaker but also comprise nonverbal aspects of a conversation between two or more speakers such as regulation of speaker roles and distance between bodies. Sometimes, the term “paralinguistic” is restricted to vocal features, excluding the nonvocal characteristics that are referred to as nonverbal behaviors or body language. This article uses the broader sense, comprising both vocal and nonvocal aspects of speech, although the focus is primarily on vocal aspects.

### 2.2. Measurement of Paralinguistic Behaviors

#### 2.2.1. Levels of Analysis

Paralinguistic behaviors have been analyzed at four levels: at the motor–physiological level of the body organs that produce the signals, at the physical level using acoustic or spatial parameters, at the perceptual level of the various sensory (mainly auditory and visual) modalities, and at the functional level referring to “meaningful” voice or movement categories. Analysis techniques at the first three levels may be summarized as “microanalytic techniques,” whereas functional techniques may be referred to as “macroanalytic techniques.” Studies using direct recordings of muscle activities such as electromyography (EMG) are rare; most of the analyses carried out are based on audio or video recordings.

In many studies, human observers are used to assess paralinguistic behaviors. For some types of studies, the human assessment is indispensable because such studies focus on the perceptual properties of the utterances or on the inferences drawn by listeners. In other studies, the judgment of human observers is used as a

more readily available or more holistic measure as opposed to the rather analytically defined physical measures. The disadvantage might be that the human system is rarely able to isolate the various physical parameters from each other. The advantage might be that sometimes information is conveyed based on complex physical configurations that are more easily judged as holistic impressions than formalized in terms of physical parameters.

### 2.2.2. Vocal Behavior

Human vocal behavior is the joint product of the action of three muscle systems: the respiratory, phonatory, and articulatory systems. The action of these muscle systems leads to air pressure changes, which can be perceived by listeners and described in terms of acoustic parameters based on audio recordings.

Three major acoustic dimensions are distinguished: time, frequency, and amplitude. Time-based parameters are related to the number and length of sound and silence segments within the speech flow, to onset and offset parameters of speech passages, and to speaker switching. Amplitude-based parameters refer to the energy level of voice. Of the parameters based on frequency, the one used most often is fundamental frequency ( $F_0$ ). It is defined as the number of repetitions of a complex wave per second, measured in hertz (Hz), and corresponds to the frequency at which the vocal cords open and close. Averaged over a speech utterance, it can be expressed as mean  $F_0$ . Changes in  $F_0$  may be measured in terms of  $F_0$  variability or in terms of  $F_0$  contour, referring to the course of consecutive  $F_0$  values. Some important measures involve more than one dimension. For example, the voice spectrum refers to the frequency–amplitude pattern of a speech signal, showing the component frequencies of a complex waveform and their relative amplitudes.

At the perceptual level, the tone of voice, speech style, and conversation style are distinguished. Although there is no one-to-one relationship, the voice characteristics of pitch, loudness, and timbre are the perceptual counterparts of  $F_0$ , energy level, and voice spectrum, respectively. Speech style can be assessed in terms of temporal properties, such as pauses and speech tempo, and in terms of speech disfluencies, such as “filled pauses” (i.e., interruptions of continuity filled by nonlinguistic vocalizations such as “ah,” “mhm,” and “er”) and slips of the tongue. The conversation style refers to features such as opening and ending a dialogue, speaker switching and

interruptions of the speaker by another person, turn-taking, and feedback signals.

At the functional level, voices may be characterized by holistic labels such as harsh, powerful, and pleasant, and speech styles may lead to impressions such as halting and hectic speaking. Feedback signals and interruptions in conversation may signal interest or attentiveness as well as criticism or disagreement.

Table I illustrates the common measures at the acoustic and perceptual levels and gives some examples of functional characterizations of voice, speech, and conversation style.

### 2.2.3. Nonvocal Behavior

Analysis of nonvocal paralinguistic behaviors sometimes relies on impressionistic descriptions of real-time behavior, but it is based mostly on video recordings. The most common microanalytic technique for the analysis of facial movements is the Facial Action Coding System (FACS) developed by Ekman and Friesen in 1978. It is an anatomically based system for identifying the muscular contractions (so-called action units) in the brow, eye, and mouth regions responsible for changes in facial movement that roughly correspond to functional descriptions such as various kinds of smiling and various emotions. The original manual coding system has been automated in terms of digital feature point tracking. A microanalytic technique for the study of gestures and body movement is provided by the Bernese System, which allows a quantitative description of body and gestural positions and movements. Functional characterizations of gestures distinguish emblems, illustrators, and adaptors. Emblems directly translate words or phrases, for example, the signs for “yes” and “no.” Illustrators accompany verbal utterances and clarify their verbal meanings, for example, upward movements that accompany the verbal expression “It’s up there.” Adaptors refer to movements that have no communicative function but serve some personal need, for example, scratching one’s head until the itch is eliminated.

## 2.3. Possible Functions of Paralinguistic Behaviors

On the one hand, paralinguistic behavior is a product of the speaker’s biological endowment as well as of the speaker’s cognitive and emotional state. Compared with language, information conveyed by paralinguistic cues is less controllable by the speaker. For example,

TABLE I  
Acoustic, Perceptual, and Functional Characteristics of Vocal Behaviors

	<i>Acoustic level</i>	<i>Perceptual level</i>	<i>Functional level</i>
Voice	Frequency- and amplitude-based parameters Mean fundamental frequency ( $F_0$ ) $F_0$ variability, $F_0$ range $F_0$ contour Energy level Measures of frequency spectrum	Tone of voice Pitch level Pitch variation, pitch range Intonation, speech melody Loudness Timbre, voice quality	Voice characteristics High voice Exaggerated intonation Soft voice, powerful voice Harsh voice, pleasant voice
Speech	Time-based parameters Duration of speech utterances Number and duration of silences Sound–silence ratio	Speech style Amount of speaking Number and length of pauses Speech rate, tempo Filled pauses Intrusions Disfluencies	Speech style Hectic speaking, halting speaking Hesitant speaking
Conversation	Time-based parameters Onset and offset of speech passages Speaker switching	Conversation style Opening and closing of conversation Speaker switching Turn-taking signals Interruptions Feedback signals	Conversation style Cooperative, intrusive Attentive, inattentive, criticizing

Note.  $F_0$ , mean fundamental frequency.

the experience of stress often constricts the vocal cords, thereby creating an increase in vocal pitch. This unintentionally given information has been called “nonverbal leakage.” On the other hand, many aspects of voice, speech style, and face and body movement are managed and applied deliberately. For example, a speaker may increase loudness and raise pitch to prevent an interruption by the listener. In that sense, paralinguistic behavior may serve communicative purposes similar to those of verbal behavior. Yet there are some important differences. First, the information is less explicit because there is no unambiguous relationship between the cues emitted and their meanings. Second, compared with language, the relationship between cues and meanings is more iconic than arbitrary (i.e., the cues bear some intrinsic resemblance to the meanings that are expressed). For example, a loud voice by itself expresses a kind of potency. Third, compared with language, paralinguistic signals can be

encoded continuously instead of discretely; that is, modifications can be applied in a gradually varying manner without being subjected to fixed steps. For example, the voice can be raised in varying degrees, whereas a verbal appeal requires decisions about the use of specific terms. Implicitness and ambiguity sometimes may be an advantage for the speaker as well as for social interaction. For example, it might be more polite and less disruptive for social harmony to express anger toward the addressee through one’s voice or facial expressions rather than through words.

### 3. THE RELEVANCE OF CULTURE

Culture is not opposed to nature; rather, as a consequence of phylogenetic evolution, it is part of the natural endowment of humans. However, what can vary is

that there are specific forms of culture. In this sense, the term “culture” is used as a set of guidelines that individuals share as members of a particular society and that tell them how to view the world and how to behave in relation to other people. Cultural guidelines are transmitted from one generation to the next mainly by symbols. The most important system of symbols by which culture is transmitted is language. Although the general ability to use language must be viewed as species-specific, the specific language spoken by an individual is not. It is, as Hofstede stated, a “vehicle of culture.” In some sense, it is even a part of culture given that it has evolved simultaneously to the cultural tradition. Similar to the norms and rules of culture that are transmitted by language, language itself must be acquired by an individual during ontogenetic development. Therefore, because paralinguistic behavior occurs simultaneously with speech behavior, the question arises as to what extent it can be considered independent of the specific language that an individual uses. Whether such an abstraction is possible must be demonstrated by analyses showing equivalences in paralinguistic behaviors among users of various languages.

If cultural differences are to be investigated as antecedents of individual behavior, meaningful dimensions that differentiate cultures must be specified. The most important approach for doing so was used by Hofstede. Based on factor-analytic evaluation of questionnaires administered to members of more than 50 countries, he extracted four bipolar dimensions that he called “individualism–collectivism,” “power distance,” “masculinity–femininity,” and “avoidance of uncertainty.” The most investigated of these four dimensions is individualism–collectivism. Individualist cultures foster the needs, wishes, and desires of individuals; as such, they encourage autonomy, separateness, and uniqueness. Collectivist cultures foster the needs, wishes, and desires of in-groups over individuals; these cultures encourage values such as social harmony, cohesion, cooperation, and conformity. In individualist cultures, the self–other relationship is perceived as more independent, whereas in collectivist cultures, it is construed as more interdependent. Sometimes, collectivist cultures are also characterized as “high-context” cultures, as opposed to the “low-context” cultures typical of the individualist orientation. A high-context culture is one in which much of the information is implicitly inferred from the situational context and from shared experiences. A low-context culture is one in which most of the information is explicitly stated in the verbal message.

#### 4. EVOLUTIONARY CONTINUITY IN COMMUNICATION

If human language communication is compared with primate vocal communication, some qualitative differences can be found. First, the relationship between sound and meaning is arbitrary and not iconic. Second, the minimal sound units (phonemes) have no meaning by themselves but distinguish different meanings (e.g., the phonemes [r] and [l] have no meaning in English but discriminate the meanings of the words “read” and “lead”). Third, the most dominant principle of sound production is articulation and not vocalization. Fourth, single units are not just repeated or intensified but are combined by rules at different hierarchical levels (i.e., phonological, morphological, lexical, syntactic, and semantic).

Although the symbol system of language must be considered a species-specific feature of humans, vocal communication has its precursors in nonhuman animal behavior. The qualitative differences between human language and nonhuman animal vocal communication do not rule out that there also exist evolutionary continuities of communication from animals to humans in forms and functions. As noted previously, paralinguistic behaviors show at least two properties similar to those of animal communication: iconic relationship to meaning and continuous encoding. Therefore, it makes good sense to study the evolutionary roots of language. The communication of sender identity, emotional states, and interpersonal relationships seems to be particularly well suited for study. For example, birds and mammals use harsh, relatively low-frequency sounds when they are hostile and use higher frequency, more tone-like sounds when they are frightened, appeasing, or approaching in a friendly manner. In 2002, Soltis and colleagues reported that certain responses of squirrel monkeys to audiotaped vocal calls reflect the affiliate relationship between sender and receiver. Although Darwin assumed that it serves the adaptation of an organism to communicate its state as precisely as possible to other organisms of the same species to allow cooperation and minimize the danger of lethal conflict, this assumption has been challenged by modern sociobiologists. According to their view, the basis for evolutionary adaptation is not cooperation but rather struggle between individuals, and natural selection favors individuals who successfully manipulate the behavior of other individuals. Applied to paralinguistic behavior, this means that expressive displays serve strategic purposes rather than simply mirroring emotional states and goals.

Based on those assumptions and results, it can be fruitful to look at similarities and continuities in animal and human communication. It is not impossible to find instances of communication that do not obey the rules of the system of language but reflect phylogenetic older layers of communication in humans. For example, some researchers have claimed that the vocalizations of infants are strikingly similar for humans, lions, polar bears, cats, and gorillas.

## **5. PARADIGMS FOR THE STUDY OF PARALINGUISTIC BEHAVIORS**

### **5.1. A Functional Framework**

Paralinguistic behaviors have been investigated in various disciplines. Among them are linguistics, psychotherapy, social psychology, evolution biology, and expression psychology. To evaluate results and hypotheses with respect to cross-cultural issues, a systematic framework is required. Research may be organized around a classification of material aspects (e.g., acoustic or perceptual properties) as well as around functional issues (e.g., questions closely related to social interaction processes). It seems more parsimonious to focus on the functional perspective because in that case the significance of results, in terms of human universals and cultural differences, can be evaluated easier. A basic assumption of the functional approach is that different patterns of behavior (in terms of material properties) may serve the same function in different cultures, and in turn the same pattern of behavior may serve different functions in different cultures. Starting from a functional classification, researchers must determine to what degree a given function is present in different cultures and by what material properties it is realized.

Two different functional paradigms are presented here: the communication model proposed by Schulz von Thun in 1999 and the Brunswikian lens model proposed by Scherer in 1982. The former is orientated toward verbal communication, whereas the latter was originally developed in the context of general perception and later adapted to social perception research.

### **5.2. The Communication Model of Schulz von Thun**

Schulz von Thun's communication model represents an elaboration of Bühler's organon model, the speech act theory, and Watzlawick and colleagues' communication

theory. According to that model, the communication process consists of three distinct components: the sender, the message, and the receiver. The sender and receiver do not act completely independent of one another; they may change their roles. The receiver may give feedback signals that affect the sender. The message, which includes verbal as well as nonverbal properties, is emitted by the sender and decoded by the receiver. There is no one-to-one correspondence between what the sender emits and what the receiver decodes; the sender may emit more than the receiver can decode, and the receiver may perceive or infer other characteristics than those emitted by the sender. Concerning the message or, more specifically, the speech utterance, a distinction is drawn between what is said and how it is said. The former refers to the content of the message, whereas the latter refers to the relationship between the sender and the receiver. The relationship is subdivided into three functional aspects: expression of situational and habitual states of the sender, appeal to the receiver, and the kind of interpersonal relationship between the sender and the receiver.

The four functional aspects may be encoded by all material properties, although to different degrees. For example, an appeal to the receiver can be expressed through words underlining the urgency or legitimacy of a request or through characteristics of the voice signaling that the sender does not expect any resistance. Not all aspects are equally present in all messages; the relative weight of each aspect may vary in different situations. One can assume that cultural norms regulate which aspect is displayed to what degree as well as what verbal and nonverbal properties are to be used in which situation.

### **5.3. The Brunswikian Lens Model**

Schulz von Thun's communication model focuses on products or outcomes of communication and does not give much attention to the processes leading to those products. To throw light on these processes, some investigators have adopted Brunswik's lens model as their research paradigm. Two kinds of processes are distinguished: (a) the externalization of the speaker's traits and states leading to objectively measurable or subjectively perceivable cues of a trait or state and (b) the listener's use of different cues to make judgments on the speaker's states or traits. Implementing the lens model involves three steps. One step is to investigate the process of externalization by measuring various physical qualities, perceptual qualities, or both and

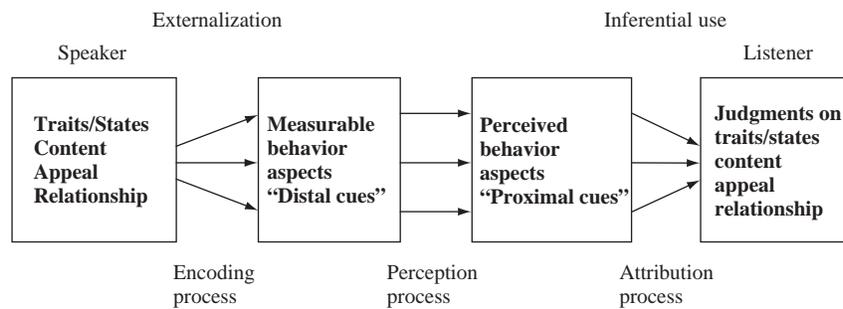


FIGURE 1 A modified version of the Brunswikian lens model.

determining which of these correlate with various speaker features that have been either manipulated or assessed by some objective index (e.g., test scores). This step serves to identify the ecological validity of various distal (i.e., objectively measurable) or proximal (i.e., perceivable) cues. A second step is to determine the accuracy with which listeners can identify the speaker's traits and states from paralinguistic cues. This is done by correlating listeners' judgments with the objective indexes of the speaker's traits and states. A third step is to investigate the specific kind of paralinguistic cues that listeners use to make their attributions. This is performed by correlating various physical or perceptual cues with listeners' judgments. This step identifies listeners' inferential use of cues.

A shortcoming of the lens model is that the communicative function of speech utterances is nearly completely ignored. In terms of Schulz von Thun's communication model, the expression function is stressed without taking into account the aspects of appeal, the kind of interpersonal relationship, and the content of the message. To overcome this weakness, a modified version of the lens model is presented (Fig. 1).

It is assumed that both externalization and use of cues are affected by cultural norms and rules. The model provides a guideline to examine these cultural variations.

## 6. EXTERNALIZATION OF PARALINGUISTIC BEHAVIORS

### 6.1. Gender and Age of the Speaker

There is little doubt that there are voice and speech characteristics that mark the personal identity, gender, age, and ethnic background of a speaker. Such

long-term speaker-characterizing features are of two different sorts. One type arises from physical characteristics of the vocal and respiratory apparatus and is, to a high degree, beyond volitional control. For example, men and boys usually have louder and lower pitched voices than do women and girls as a result of having a larger respiratory volume as well as longer and thicker vocal cords. Elderly individuals often experience changes in voice quality and loudness due to factors such as ossification of the vocal cords, diminished respiratory capacity, and reduced coordination within the central nervous system. The second type of characterizing features is the product of the way in which the speaker habitually sets his or her vocal apparatus for speaking. These features are under potential control, yet they are not usually deliberately managed by the speaker.

In all societies, age and gender are important social categories, and individuals of different genders and ages may be discernible on the basis of specific voice and speech characteristics. However, both the specific features and the magnitude of differences may be partly affected by cultural factors.

Usually, women and girls show higher pitch and greater pitch variability than do men and boys; however, the magnitude of female-male differences varies across cultures and even across social classes. It may be hypothesized that the magnitude of these differences relates to differences in the masculinity-femininity dimension. Because masculine countries stress gender role differences, voice differences should be more pronounced than in more feminine cultures. Consistent with this hypothesis is that in Japan, a country with a high masculinity index, the difference in pitch (as well as in verbal style) has been found to be higher than that in England and The Netherlands, two countries with a moderate masculinity index. However, the three

countries differ even more with respect to power distance, so that one cannot unambiguously attribute the voice and speech differences to differences in masculinity. It would be of interest to compare various countries of equal power distance differing exclusively in masculinity (e.g., with Japan as a masculine country and Thailand as a feminine country).

In a number of studies,  $F_0$  has been investigated as a function of age. The average  $F_0$  in babies has been found to be extremely high when they are crying (approximately 400–600 Hz) and somewhat lower when they are babbling (approximately 300–500 Hz). It decreases for both males and females with increasing age, at least until adulthood is attained. The changes during adulthood show a marked difference between the sexes, at least in Western societies. For women, no evidence of any systematic changes has been found. On the other hand, average  $F_0$  in men appears to decrease slightly from young adulthood to middle adulthood and then to shift upward from approximately 65 years of age onward. One explanation relates the upward shift in old age to a weakening of the gonads in secreting hormones, a process comparable to puberty, a period when one generally finds a marked decrease in  $F_0$ . However, it is also possible to explain this shift without recourse to biology.  $F_0$  has been shown to be affected by both habitual and transitory stress; most people tend to respond to stress with an increase in mean  $F_0$ . Thus, it seems reasonable to hypothesize that the upward shift reflects not only age-related physiological changes but also emotional tension caused by factors such as forced retirement, fewer social contacts, decreasing self-sufficiency, and low status associated with old age in Western societies. Cross-cultural comparisons might well shed some light on this issue, but there has been a lack of such studies so far. One may assume, for example, that factors such as the evaluation of old age, in terms of status and institutional support in a particular culture, do play an important role in mediating voice changes.

## **6.2. Personality Traits of the Speaker**

Many attempts have been made to show that paralinguistic behaviors express long-term personality traits. In most of the studies, the central issue is the accuracy with which listeners can judge the speaker's personality on the basis of his or her voice. According to the

lens model, however, the first step must be to investigate whether or not personality traits are externalized in objectively measurable distal cues. This can be done by assessing the actual correlations between objectively measured characteristics and valid criteria of personality traits.

There is some evidence from research done in Western countries that personality traits such as extraversion and dominance are indicated by acoustical cues such as energy level and  $F_0$ . However, the direction and magnitude of correlations, especially with respect to  $F_0$ , turned out to depend on particular situations. In situations where competitive achievement is encouraged, dominant and extraverted people tend to speak in a higher pitched voice (i.e., with a higher mean  $F_0$ ) than do submissive and introverted people. However, in neutral situations or in situations where communicative abilities are stressed, extraversion and dominance are associated with a lower pitch. Presumably, different facets of extraversion are addressed. Achievement-orientated situations favor ergotropic facets of energy, ascendance, and ambition, whereas neutral or interpersonal communication situations evoke a preponderance of trophotropic (rest) facets of warmth and gregariousness. Moreover, there are considerable cultural differences even within different Western countries. For example, the correlations are more pronounced in Americans than in Germans.

One may assume that in collectivist countries, the correlations between voice and personality are even smaller. When group conformity is stressed instead of individuality, personality traits as markers of the self should be externalized to a much lower degree than in individualist cultures, where the independent self is more likely exhibit cues that clearly mark individuality and serve self-presentation. Thus, it might not be surprising that the correlations between personality and voice obtained in Germany are smaller than those obtained in the United States given that Germany showed a lower index of individualism than did the United States.

## **6.3. Situational States of the Speaker**

Often, the voice is considered to be a direct means of expression of perceived stress. Hence, there have been numerous attempts to find paralinguistic correlates of situational stress. The significance of acoustic cues has

been tested by comparing the differences against some baseline conditions such as neutral or positive situations. Stress has been experimentally induced by threatening movies, increased cognitive workload, and negative feedback.

Stress is often accompanied by increased muscle tension. Thus, it has been hypothesized that increased tension in the respiratory muscles leads to louder speech and that increased tension in laryngeal muscles leads to higher  $F_0$ . Increased tension in the respiratory muscles increases subglottal pressure, thereby producing a slightly higher  $F_0$ . Because of the increased muscle tension, the  $F_0$  range of the voice should be narrowed and the number of harmonics should also decrease. Moreover, it has been hypothesized that the speech rate and the number of disfluencies increase with higher stress. In general, these hypotheses have been confirmed; most people respond to increased stress with an increase in  $F_0$  relative to baseline, a decrease in the  $F_0$  range, an increase in the speech rate, and an increase in the number of disfluencies. However, there also are considerable interindividual differences. Some people even show a reversed pattern, that is, a decrease in the mean  $F_0$  and an increase in the  $F_0$  range. There is some evidence that these interindividual differences mirror different coping styles such as repression-sensitization. Although they refer to personality differences in arousal control, there is no doubt that cultural norms strongly affect the expression of arousal by so-called display rules. With respect to facial expression of fear, it could be documented that the displays differ across cultures depending on the social situation. When alone, Japanese and American participants displayed exactly the same facial expressions of fear; however, when in the presence of another person, their expressions differed dramatically, with the Japanese invariably smiling rather than displaying their fear. Because articulatory and facial muscles are partially identical, one may expect some parallels to facial expression in the voice. Developmental studies suggest that the regulation of socially disruptive emotions is acquired at an early age. For example, Japanese preschool children respond to frustrating situations with less aggression and anger (as rated on the basis of voice, face, and body) than do U.S. children. To account for the differences, it is hard to determine whether they reflect differences in open expression or differences in experience of emotions due to different interpretations of frustrating events.

#### 6.4. Emotional Content

A person would not be accepted as speaking a language properly if he or she did not deliver the utterances in the pitch pattern, stress pattern, and temporal pattern of grouping and pausing appropriate for that language. Although these prosodic features affect the linguistic meaning of a sentence, the emotional content of an utterance may be expressed independently. There is a long history of researchers trying to identify specific paralinguistic patterns associated with specific emotional contents, including the contributions of the German "expression psychology" and research by evolution biologists. It is a basic assumption that paralinguistic features exist and that they have an isomorphic relationship to their meaning. The validity of distal and proximal cues indicating a particular meaning can be tested by asking speakers to deliberately express a specific emotion, typically one of the six or seven basic emotions: anger, disgust, fear, happiness, sadness/grief, surprise, and (sometimes) contempt. If physically measurable or proximally perceivable correlates of the expressed meanings can be established, the validity may be ensured. Consistent correlations have been found only for anger and sadness/grief, two emotions that may be characterized as the poles of the dimension of activity. Anger seems to be indicated by high  $F_0$ , high  $F_0$  variability, high energy level, and high speech rate. The opposite characteristics seem to be associated with sadness/grief: low  $F_0$ , narrow  $F_0$  range, low energy level, and low speech rate. So far, there is little evidence for universal patterns discriminating between the poles of other emotional dimensions such as valence (positive vs negative). One reason might be that a particular emotion or emotional dimension is indicated by specific combinations of different cues (i.e., amplitude, frequency, and temporal patterns) that are difficult to quantify. Another reason might be that culture-specific display rules determine in which situations certain patterns may undergo a so-called ritualization, indicated by exaggeration and repetition of single-pattern elements.

#### 6.5. Appeal to the Listener

Although speech behavior may occur without reference to listeners, in most natural speech situations the speaker addresses listeners with the purpose of exerting some influence on the listeners' thoughts, feelings, and behaviors. This function of speech behavior is often referred to as "appeal." It implies an adaptation

toward the specific addressees. The listener variables to which the speaker may adapt can be of different kinds. For example, the speaker may talk to persons of higher, lower, or equal status; to persons of the opposite sex; to familiar persons, or to strangers. Within empirical studies conducted in this field, the most salient feature of listeners that speakers adapt to is age. Thus, a speaker may talk in a loud voice and very slowly to older people, assuming that they are hard of hearing and have difficulty in comprehending. The best-known phenomenon in this area is so-called baby talk (*Ammensprache*, *Motherese*, or infant-directed speech), a voice and speech style used by adults, and even by children, when speaking to infants. Although the phenomenon has been found in a variety of cultures and so may assume the status of a universal, the extent of use and the upper limit of the addressee's age vary from culture to culture. Evidence from diverse linguistic and cultural backgrounds shows that infants prefer the baby talk to the adult speech style by showing more attention, more affective behavior, and more positive vocalizations. Infant-directed speech differs from adult-directed speech by phonetic, morphological, syntactic, and paralinguistic features. Most salient are the paralinguistic features. The most notable acoustic features of infant-directed speech, as compared with adult-directed speech, are its higher  $F_0$ , wider  $F_0$  range, slower tempo, and longer pauses. The differences in  $F_0$  between infant-directed speech and adult-directed speech seem to be generally more exaggerated in Western cultures (e.g., United States, Australia, England, Italy, Germany) than in Eastern cultures (e.g., China, Japan, Thailand). This may be partly due to the differences between tonal and nontonal languages (e.g., Mandarin Chinese and Thai vs English, German, and Italian). For example, in tonal languages, overly exaggerated  $F_0$  contours superimposed on a sentence may disrupt the lexical information conveyed by word tone. However, one cannot rule out the possibility that the differences also map onto cultural differences in the acceptability of open emotional expression. For example, in Japanese (a nontonal language), the exaggeration of  $F_0$  contours in infant-directed speech, compared with that in adult-directed speech, is also less pronounced. There appear to be three main functions of infant-directed speech in that (a) it serves to initiate and maintain the child's attention, (b) it communicates empathy, and (c) it facilitates language acquisition. Because language-specific features, such as prosodic, lexical, and syntactic features, increase with increasing age of the child, it can be

concluded that facilitating infant socialization by engaging attention and empathy dominates at younger ages, whereas facilitating language acquisition prevails during later infancy.

A specific kind of appeal is used in the context of street traffic. For example, a pedestrian crossing a road must ensure that an approaching car will permit him or her to cross. In this case, eye contact and hand and body movements exchanged between the driver and the pedestrian have the function of emblems. Although these emblems may have evolved from universally valid iconic gestures, they have been culturally transformed and their situation-specific meanings must be learned. Eye contact between the driver and the pedestrian may universally signal "I see you," but in some cultures the action-relevant message is "Okay, cross the street," whereas in other cultures the appeal that is expressed is "Pay attention or it will be dangerous for you." Thus, merely decoding the literal meaning of the gesture will be misleading and may have severe consequences.

## 6.6. Interpersonal Relationship between Speaker and Listener

When two or more people are engaged in conversation, they take turns speaking. Usually, they manage to achieve a fairly smooth synchronization of utterances without too many interruptions and silences. The emerging "rhythms of dialogue" or "accommodation" patterns are highly dependent on the particular speech situation (e.g., formal vs informal, face-to-face vs telephone conversation) and especially on the kind of relationship between the speakers (e.g., mother-child, teacher-student, inferior-superior, like-dislike). Although considerable variations do exist within a culture due to interindividual differences (e.g., extraversion), cross-cultural differences appear to be even larger. For example, the physical duration of silence following an utterance by one of the communication partners that is experienced as appropriate, polite, or uncomfortable is highly regulated by cultural norms and conventions. Roughly speaking, utterances tend to be longer and interspeaker silences tend to be shorter in Western cultures than in Eastern cultures. This may reflect the difference between low-context cultures and high-context cultures. To a member of a high-context culture, what is omitted is an essential part of the communication. Therefore, silence is highly valued. To a member of a low-context culture, what is omitted

creates ambiguity that will be eliminated by explicit verbalizing.

The coordination between the speaker and the listener is perfect when the speaker sends the correct signals for a change of turn and the listener responds to these signals. Also, the listener may interrupt the speaker to take a turn. Interruptions imposed by the listener may signal positive feedback or act as a power device to get control of the conversation. In this vein, two main types of interruptions are distinguished: cooperative and intrusive. An example of the former is a vocalization signaling agreement, whereas an example of the latter is a loud voice abruptly interrupting the speaker. It has been hypothesized that in a collectivist culture that stresses social harmony, individuals send more cooperative interlocutor signals, whereas in an individualist culture that encourages self-assertion and competition, intrusive interlocutor signals dominate. In 2001, Li tested this hypothesis using simulated doctor–patient conversations between either Canadian or Chinese participants. The comparison of the two intracultural conversation conditions showed that Chinese participants engaged in significantly more cooperative interruptions than intrusive ones, regardless of their roles as doctors or patients. In contrast, the Canadian doctors performed significantly more intrusive interruptions than cooperative ones. Although the coding procedures did not differentiate between linguistic and paralinguistic cues, it can be assumed that the latter determined categorizing to a considerable extent. Researchers have observed similar differences in interruption patterns between members of other collectivist cultures (Japanese and Thai dyads) and members of another individualist culture (British dyads). These findings may be taken as providing strong support for Markus and Kitayama's theory that collectivists construe the self–other relationship interdependently, whereas individualists clearly mark their personal boundaries.

In addition to sending audible feedback signals, interlocutors coordinate their conversations through spatial behavior, eye contact, and touch. Cultures develop more or less explicit rules about the appropriate use of these behaviors. Those who stand or sit too far away are seen as cold, aloof, and withdrawn, whereas those who come too close are seen as embarrassingly intrusive and overly intimate. Anthropologists distinguish between contact cultures and noncontact cultures. People from contact cultures stand closer, face more directly, touch, and look more than do people from noncontact cultures. Contact cultures include

Arabs, Latin Americans, Southern Europeans (e.g., Greeks, Turks), and a number of African cultures, whereas noncontact cultures include Asians and Northern Europeans. Although it has generally been observed that U.S. Americans stand less close and touch less than do Arabs and South Americans, there are considerable differences within the United States: Hispanics stand closer than do Anglos, and most studies find that African Americans interact at greater distances, look less, and adopt a less direct orientation than do Whites. Moreover, these differences vary with age; African American children stand closer than do Whites, whereas adolescents and adults stand farther away from one another. Age differences have also been observed in Japan, which is generally regarded as a noncontact culture: in public situations, whereas adults show very little body contact, young Japanese women touch each other often.

## 7. ACCURACY OF LISTENERS' INFERENCES

### 7.1. Methodological Considerations

To study the accuracy with which listeners can infer traits or states of the speaker and emotional contents of the message on the basis of the speaking voice, listeners are presented with a voice sample and must make a judgment. The accuracy is assessed by correlating the judgments with some objective criteria of states, traits, or content. The voice samples used as stimuli may be, to a great extent, similar to those elicited in externalization studies. However, to prevent listeners from being biased by the linguistic content of the message, either the stimuli must consist of a standard text (spoken or read) or the linguistic content must be masked. Most masking techniques applied are low-pass filtering and random splicing. Low-pass filtering preserves pitch and loudness contours as well as temporal information, such as pauses and speech rate, but it removes most features of voice timbre due to the elimination of upper frequency parts of the voice spectrum. Random splicing is carried out by splitting the speech signal into short segments and then concatenating the segments in random order. It preserves the static voice characteristics of pitch, loudness, and timbre but eliminates the dynamic characteristics such as pitch contour and temporal properties. Judgments are obtained by having listeners either rate a particular trait, state,

or emotion on a rating scale or choose an item from a list of response alternatives.

## **7.2. Personality Judgments**

In many of the earlier studies investigating the accuracy of personality judgments on the basis of voice, researchers have found that there is consensual agreement among the judges but that the validity of judgment is low. Based on these findings, it has been concluded that the judgments reflect “vocal stereotypes” rather than accurate inferences. This conclusion was challenged by arguing that invalid personality measures have been used and that the functional perspective of perceived personality traits has been neglected. In this vein, it has been suggested that people should be most attuned to those traits about which correct or incorrect identification has implications for actions relevant for survival and reproduction in the evolutionary past (e.g., social dominance, extraversion). More recent studies have shown that these traits can be inferred from voice with at least moderate accuracy. This does not necessarily contradict the situation specificity of distal cues indicating these traits given that judges may take into account the particular speech situation when making inferences. However, similar accuracy scores might not be obtained universally. Cultural differences may emerge due to several factors. First, real differences may exist in mean levels of personality traits such as extraversion and dominance. Second, the measurement of personality traits through questionnaires might not be equivalent across cultures due to various self-accessibility of personality traits and various response styles. Third, personality traits as markers of personality may be externalized to various extents due to cultural rules.

## **7.3. Judgment of Emotional Content**

Considering that the emotional content of an utterance can be expressed independently of lexical and syntactic structure, a question arises as to whether listeners are able to accurately recognize the particular emotion the speaker has portrayed. The criterion for accuracy typically used is whether the judges correctly identify the emotion that the speaker has been asked to portray. Psychologists have often debated whether the recognition of emotions is universal or whether it varies by culture. Three competing hypotheses have been proposed. The strong universalist position follows Darwin and states that the expression and recognition of

emotions is biologically programmed and, hence, universal. The weak universalist hypothesis asserts that specific emotional categories are largely culturally specific but that broad dimensions, such as activity/arousal and valence, are universal. The interactionist position acknowledges that emotions are programmed biologically but claims that the process of learning to control both expression and perception is highly dependent on cultural factors.

Empirical studies have yielded the following results:

1. Emotions are recognized from voice and face at a better-than-chance level within and across cultures. In addition, accuracy differences among various emotions are highly similar within and across cultures.

2. Cross-culturally, some emotions are better recognized than others. Among the seven basic emotions, fear and disgust are the most poorly recognized, whereas happiness is the most accurately recognized.

3. Accuracy is higher when emotions were both portrayed and identified by members of the same national, ethnic, or regional group, suggesting a so-called in-group advantage.

4. Cross-cultural accuracy is different across channels. In general, the face seems to be the best source of information, although there are emotion-specific differences. Happiness is most accurately recognized from the face but least accurately recognized from the voice. In comparison, anger and sadness are most accurately recognized from the voice but relatively less accurately recognized from the face. Likewise, the in-group advantage seems to be smallest for happiness in the face and greatest for happiness in the voice. Fear shows a large in-group advantage in both the face and the voice, whereas anger shows a relatively small in-group advantage in both the face and the voice.

5. When emotions are expressed by speakers of a foreign language, accuracy is higher for judges whose mother tongue is close in linguistic structure to the language used by the speaker than for judges whose mother tongue is completely different.

6. The evidence for differences in culture per se is not clear-cut. Some results suggest that collectivists have difficulties in both expressing and recognizing emotions (e.g., anger, contempt) that might be potentially disruptive to social interaction. However, it is not clear to what extent the results can be attributed to a mismatch between the cultural background of the speaker and the listener.

Evaluating the three hypotheses formulated previously, it seems that the strong universalist position

cannot be held. There is some support for the second hypothesis in that the dimension of activity/arousal, but not the dimension of valence, seems to be present universally. The interactionist hypothesis fits the data best. Certain core components of emotions seem to be universal and are likely to be biological. However, the style of expression and recognition is governed by cultural conventions. Linguistic rules also appear to have a strong impact. Although linguistic content has been controlled experimentally, language-specific prosodic rules are likely to affect both expression and recognition. This parallels the results found in infant-directed speech, where language-specific and language-independent principles also seem to be intricately linked.

There has been some criticism concerning the studies of emotional recognition centered around the fact that both stimuli and elicitation of responses may lack ecological validity. Also, because responses are typically elicited as forced choices, the high accuracy scores obtained merely show that a particular emotion can be reliably discriminated from other emotions. They do not show that a particular emotion can be correctly identified in reality when no response alternatives are provided. In addition, some researchers have questioned the ecological validity of stimuli. Although deliberately portrayed emotions may exhibit stylized forms of paralinguistic behaviors intended to communicate, these might not be typical in the spontaneous expression of experienced emotions. Instead of mirroring actual emotional states, the latter may be influenced by subjective control mechanisms such as coping and interpersonal manipulation strategies. Moreover, the same emotion may produce different types of expression.

Finally, it has been argued that in face-to-face communication, the listener is presented with information from multiple channels. Thus, optical information may override acoustic information. Compared with judgments of the voice alone, the addition of video information—especially of the face—has been found to be associated with a large additional increase in cross-cultural accuracy. Moreover, in natural communication settings, judgments on voice and behavioral responses to voice usually are affected by the linguistic content of the message. In some cases, linguistic and paralinguistic sources of information may lead to similar inferences, but in others there may be discrepancies. The source of information on which the listener relies most may vary according to the particular speech situation, but it may also be affected by cultural factors.

## 8. USE OF PARALINGUISTIC CUES BY LISTENERS

### 8.1. Methodological Considerations

To investigate the kind of acoustic cues that listeners use, correlations between inferences and acoustic measures of the speech signal may be applied. However, results might not always be conclusive because one cannot be sure which of the many cues available have been used and to what extent important cues went unmeasured. Thus, it is useful (and eventually necessary) to manipulate the speech signal. Although speakers can be asked to deliberately vary the paralinguistic features, they may unintentionally vary other features as well. Therefore, experimenters must be able to manipulate features synthetically. This can be done either by systematically changing the digitized speech sample with a computer or by using a speech synthesizer. Typically,  $F_0$ , energy level, and pauses have been manipulated.

The aim of some researchers is to identify proximal representations that lead to particular inferences. In that case, listeners are asked to judge the speech signal with respect to perceptual attributes such as soft voice, high voice, powerful voice, unpleasant voice, and halting or hectic speaking.

### 8.2. Cues to Personality Traits and Situational States

Results obtained in studies on acoustic cues leading to inferences of personality traits and situational states are not quite conclusive. Typically, mean energy level and energy variation are positively associated with dominance and extraversion judgments. With respect to  $F_0$  and speech rate, the results are less consistent and seem to be both situation- and gender-specific. Some researchers have found that speech rate is positively correlated with extraversion and dominance, whereas other researchers have found a negative correlation. High mean  $F_0$  seems to lead to the attribution of high dominance in male voices but not in female voices. Situational stress may be inferred from an increased  $F_0$  relative to a baseline.

Results obtained in studies using perceptual cues indicate that extraversion and dominance are inferred from loud and powerful voices. This is consistent with the findings on acoustical cues. Results on inferences drawn from pitch are as inconsistent as those drawn from  $F_0$ . High-pitched male voices are sometimes

judged as extraverted and dominant and, at the same time, as emotionally labile and disagreeable, although they are sometimes judged as less dominant and more agreeable. High-pitched female voices are usually judged as kind, humorous, immature, and agreeable.

Two reasons may account for the inconsistent results. The first is that judges rely more on configurations of multiple cues than on single cues. In addition,  $F_0$  and energy level may be acoustically insufficient to describe relevant configurations. The spectral structure of speech, which determines how the voice sounds, may outweigh the importance of  $F_0$  and energy level. For example, depending on the spectral structure, a high-pitched voice may sound childlike and be perceived as friendly, or it may sound shrill and be perceived as angry and aggressive.

The second reason is that judges are guided not only by vocal stereotypes but also by implicit personality theories. These are characterized by overgeneralizations and illusory correlations such as a negative correlation between dominance and emotional stability.

Although systematic cross-cultural studies on cues to attributions of traits and states are rare, there seems to be evidence of cultural differences. Because average loudness and tempo vary from culture to culture, misattributions from persons not familiar with the cultural standard may occur frequently. For example, Anglos who commonly speak in a softer voice and are not familiar with the Latin standard may erroneously attribute dominance and aggression to the louder voices of Latin people. Even familiarity with the general cultural standards does not prevent misattributions given that sometimes it is essential to analyze the situational demands in a culture-specific way. For example, in cultures stressing power distance, a higher ranked person will be judged as competent when speaking in a loud voice and a low tempo, whereas the same cues may lead to the attribution of boldness when exhibited by a subordinate person.

Cultural differences also seem to exist with respect to implicit personality theories. For example, the Chinese *shi gu*, which refers to “someone who is worldly, devoted to his or her family, socially skilful, and somewhat reserved,” is not an easily found stereotype in Western countries. On the other hand, the Western concept of an “artistic type” seems absent in Chinese. Thus, although it is easy for speakers of English or Chinese to refer to single personality traits (e.g., socially skilled, creative), each culture generates its own personality syndromes.

### 8.3. Cues to Emotional Content

Attempts to identify a set of vocal cues that reliably differentiate among the basic emotions have not yet been successful. In a classic experiment in 1972, Uldall presented listeners with a number of sentences on which 16 different  $F_0$  contours had been imposed synthetically. The contours differed with respect to  $F_0$  range,  $F_0$  reached at end of contour, and shape. The listeners were asked to rate each sentence on several bipolar adjective scales representing the three dimensions of emotional meaning—valence, activity, and potency—defined by Osgood and colleagues in 1957. Both high activity and high potency were inferred from contours characterized by a wide  $F_0$  range and a change of direction. A change of direction combined with a final rise led to the impression of a positive valence. The more smooth or monotonous a contour was, the more the judged meaning was dependent on the linguistic content of the utterance. The more lively a contour was, the more the judged meaning was stable and independent of linguistic content.

### 8.4. Selective Attention to Various Cues

In collectivist cultures, given that direct confrontation of another person is considered rude and undesirable and that paralinguistic information is less direct and less explicit than verbal information, one may assume that collectivist persons are more sensitive to paralinguistic cues than are individualists, who should be more attuned to linguistic information. In 2002, Kitayama and Ishii tested this hypothesis by using an interference paradigm similar to a Stroop task (i.e., a recognition task in which the print color interferes with the lexical meaning of a written word) to assess spontaneous attention to either linguistic or paralinguistic cues. In this modified Stroop task, lexical emotional content of a word interfered with vocal emotional content. Lexical and vocal meanings had previously been checked as equivalent in extremity in the two languages. Japanese participants (members of a collectivist culture) and English participants (members of an individualist culture) had to make a judgment of either vocal emotion or lexical meaning of an emotionally spoken evaluative word. As hypothesized, an interference effect by competing word evaluation in the vocal emotion judgment was significantly stronger in English than in Japanese. In contrast, an interference effect by competing vocal emotion in the word

evaluation judgment was stronger in Japanese than in English. This finding suggests that individualists are more attuned to direct communication of feelings involving words, whereas collectivists are more attuned to less explicit communication involving paralinguistic features. This does not contradict the notion that open expression of feelings is generally less acceptable in collectivist cultures than in individualist cultures; instead, it suggests that the communication of feelings in collectivist cultures involves paralinguistic behaviors rather than linguistic behaviors.

## 9. CONCLUSION

### 9.1. Universals and Cultural Differences

The aim of cross-cultural research on paralinguistic behaviors is to investigate to what extent such behaviors are used in a ubiquitously uniform manner and to what extent they are shaped by cultural factors. Addressing this issue, a modified version of the Brunswikian lens model may serve as a functional framework. Two kinds of processes are distinguished: externalization of paralinguistic behaviors by the speaker and use of paralinguistic cues by the listener. Both kinds of processes seem to be characterized by both universal and culture-specific principles.

The most uniform use of paralinguistic behaviors across cultures can be observed in adult speech directed toward infants. When speaking to an infant, adults accommodate their vocal behavior to the infant's cognitive ability and signal empathy, and the infant responds with attentive and affective signals. The paralinguistic features indicating infant-directed speech seem to be universal, although they are less pronounced in Eastern cultures than in Western cultures. The cultural differences cannot be attributed entirely to language differences (nontonal vs tonal languages). They also appear to map onto cultural differences in the social acceptability of open expression of emotion.

There is additional evidence for universality in decoding emotional meaning from paralinguistic behaviors. When emotions are intentionally portrayed, they are recognized quite accurately from voice and face both within and across cultures. This suggests that certain core components of emotional expression are not conventional and are probably biological. However, this does not imply that emotional experience is uniform across cultures or that specific emotions are

universally expressed in the same manner and with similar frequency. The wide variations observed in various cultures seem to reflect different display rules, that is, cultural rules that govern which paralinguistic behaviors are permissible and which are inappropriate in a social setting.

The most striking differences among cultures have been observed in paralinguistic behaviors that regulate the interpersonal relationship between speakers and listeners. It seems that members of individualist cultures send more signals that indicate self-focusing tendencies, whereas members of collectivist cultures exhibit more cues that stress harmony and cooperation. When giving feedback, listeners from individualist cultures tend to show their own feelings toward the utterance of the speaker, whereas in collectivistic cultures it is more important for listeners to signal a positive attitude even when they do not feel it.

Cultural differences have also been observed in the extent to which listeners use paralinguistic cues. Being presented with conflicting information from various channels, it seems that collectivists tend to be more sensitive to the less explicit information conveyed by the voice, whereas individualists rely more on the explicit information transmitted by the spoken words. This difference may reflect the difference between high-context cultures, in which the social meaning of situations is processed in a holistic mode, and low-context cultures, in which an analytic processing mode stressing the abstract meaning of words is preferred.

### 9.2. Applications to Intercultural Encounters

Cultural differences in paralinguistic behaviors have consequences with respect to intercultural encounters. The appropriate use of paralinguistic behaviors is needed to communicate effectively in a wide variety of intercultural situations. Success in business, teaching, politics, and private interpersonal relationships will depend in great part on the understanding and ability to display the culturally appropriate paralinguistic behavior when communicating with persons who are culturally different. Misunderstandings due to violations of cultural rules can result even if the words and grammar of the foreign language are appropriately mastered. In some cultures, the expectation is to mask, attenuate, or neutralize vocal and facial expressions of one's feelings. The listener must

recognize finer nuances of paralinguistic cues and pay more attention to the situational context in these cultures, as compared with cultures where feelings are openly expressed, to draw inferences about the feelings of persons from these cultures. Most notably, self-disclosure on the part of the speaker is likely to be interpreted as crude or even exhibitionistic in these cultures and may cause embarrassment.

On the other hand, in some cultures feelings are supposed to be dramatized, and so their withdrawal or suppression may lead to the impression of a cool and reserved person. However, as a practical maxim, one can assume that the lost benefit resulting from acting too reservedly in a culture stressing self-disclosure will be less than that resulting from acting too overtly in a culture where self-disclosure is taboo. The reason is that people generally pay more attention to behavior that is shown than to behavior that is not shown.

Lack of effectiveness in intercultural encounters may also result from inappropriate listening. The key to effective listening is participation. However, participation is inferred from different paralinguistic feedback cues in different cultures. In individualist cultures, feedback is expected to reflect the listener's comment on and "honest" attitude toward the content of the message conveyed by the speaker (e.g., feedback to the proposal of a business partner). In collectivist cultures, the listener will do better to just signal that he or she is actively attending to the speaker without giving his or her own comments.

### Acknowledgments

The author thanks John Berry, Richard Block, and Harry Triandis for valuable comments.

### See Also the Following Articles

Emotion ■ Emotion and Culture ■ Interpersonal Behavior and Culture

### Further Reading

- Argyle, M. (1996). *Bodily communication* (2nd ed.). London: Routledge.
- Aronson, E., Wilson, T. D., & Akert, R. M. (1994). *Social psychology: The heart and the mind*. New York: HarperCollins.
- Bühler, K. (1987). Das Organon-Modell der Sprache. In K. R. Scherer, A. Stahnke, & P. Winkler (Eds.), *Psychobiologie* (pp. 321–327). Munich, Germany: Deutscher Taschenbuch Verlag.
- Cohn, J. F., Zlochower, A. J., Lien, J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology*, *36*, 35–43.
- Dabbs, J. M., Jr., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences*, *27*, 801–804.
- Darwin, C. (1859). *On the origin of species by means of natural selection*. London: John Murray.
- Darwin, C. (1872). *The expression of emotions in man and animals*. London: John Murray.
- Dawkins, R. (1994). *The selfish gene* (2nd ed.). Oxford, UK: Oxford University Press.
- DeVito, J. A. (1997). *Human communication* (7th ed.). New York: Longman.
- Ekman, P., & Friesen, W. V. (1978). *Manual for the facial action code*. Palo Alto, CA: Consulting Psychologists Press.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–235.
- Grammer, K., & Eibl-Eibesfeldt, I. (1993). Emotionspsychologische Aspekte im Kulturenvergleich. In A. Thomas (Ed.), *Kulturvergleichende Psychologie* (pp. 298–322). Göttingen, Germany: Hogrefe.
- Helfrich, H., & Wallbott, H. G. (1986). Contributions of the German "expression psychology" to nonverbal behavior research: IV. The voice. *Journal of Nonverbal Behavior*, *10*, 87–204.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across cultures* (2nd ed.). Thousand Oaks, CA: Sage.
- Itakura, H. (2001). *Conventional dominance and gender: A study of Japanese speakers in first and second language context*. Amsterdam: Benjamins.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development*, *66*, 1–131.
- Kitamura, C., Thanavithuth, C., Chulalongkorn U., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*, 372–392.
- Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion*, *16*, 29–59.
- Li, H. Z. (2001). Cooperative and intrusive interruptions in inter- and intracultural dyadic discourse. *Journal of Language and Social Psychology*, *20*, 259–284.
- Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology*, *79*, 452–468.

- Markus, H. R., & Kitayama, S. (1998). The cultural psychology of personality. *Journal of Cross-Cultural Psychology*, 29, 63–87.
- Matsumoto, D. (1993). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion*, 17, 107–123.
- Matsumoto, D., Consolacion, T., Yamada, H., Suzuki, R., Franklin, B., Paul, S., Ray, R., & Uchida, H. (2002). American–Japanese cultural differences in judgments of emotional expressions of different intensities. *Cognition and Emotion*, 16, 721–747.
- McCrae, R. R. (2000). Trait psychology and the revival of personality and culture studies. *American Behavioral Scientist*, 44, 10–32.
- Mendoza, E., & Carballo, G. (1998). Acoustic analysis of induced vocal stress by means of cognitive workload tasks. *Journal of Voice*, 12, 263–270.
- Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, 112, 197–204.
- Murata, K. (1994). Intrusive or cooperative? A cross-cultural study of interruption. *Journal of Pragmatics*, 21, 385–400.
- Niwano, K., & Sugai, K. (2002). Intonation contour of Japanese maternal infant-directed speech and infant vocal response. *Japanese Journal of Special Education*, 39, 59–68.
- Noor Al-Deen, H. S. (1997). *Cross-cultural communication and aging in the United States*. Mahwah, NJ: Lawrence Erlbaum.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Protopapas, A., & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America*, 101, 2267–2277.
- Ramsey, S. (1997). Double vision: Nonverbal behavior east and west. In A. Wolfgang (Ed.), *Nonverbal behavior: Perspectives, applications, intercultural insights* (pp. 139–167). Seattle, WA: Hogrefe & Huber.
- Scherer, K. R. (1979). Personality markers in speech. In K. R. Scherer, & H. Giles (Eds.), *Social markers in speech* (pp. 147–209). Cambridge, UK: Cambridge University Press.
- Scherer, K. R. (1982). Stimme und Persönlichkeit: Ausdruck und Eindruck. In K. R. Scherer (Ed.), *Vokale Kommunikation* (pp. 188–227). Weinheim, Germany: Beltz.
- Scherer, K. (1997). The state of art in vocal communication: A partial view. In A. Wolfgang (Ed.), *Nonverbal behavior: Perspectives, applications, intercultural insights* (pp. 335–349). Seattle, WA: Hogrefe & Huber.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76–92.
- Schulz von Thun, F. (1999). *Miteinander reden: Psychologie der Kommunikation*. Reinbek, Germany: Rowohlt-Taschenbuch Verlag.
- Searle, J. R. (1969). *Speech acts*. Cambridge, UK: Cambridge University Press.
- Soltis, J., Bernhards, D., Donkin, H., & Newman, J. D. (2002). Squirrel monkey chuck call: Vocal response to playback chucks based on acoustic structure and affiliative relationship with the caller. *American Journal of Primatology*, 57, 119–130.
- Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, 11, 188–194.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.
- Triandis, H. C. (2001). Individualism–collectivism and personality. *Journal of Personality*, 69, 907–924.
- Trull, T. J., & Geary, D. C. (1997). Comparison of the Big-Five factor structure across samples of Chinese and American adults. *Journal of Personality Assessment*, 69, 324–341.
- Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance: Vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, 26, 148–171.
- Uldall, E. (1972). Dimensions of meaning in intonation. In D. Bolinger (Ed.), *Intonation* (pp. 250–259). Harmondsworth, UK: Penguin Books.
- Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38, 253–265.
- Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1968). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes*. London: Faber & Faber.
- Zahn-Waxler, C., Friedman, R. J., Cole, P. M., Ichiro, M., & Hiruma, N. (1996). Japanese and United States preschool children's responses to conflict and distress. *Child Development*, 67, 2462–2477.
- Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, 1, 204–223.



# Part-Time Work

Neil Conway

Birkbeck University of London, London, United Kingdom

1. Introduction
  2. Defining Part-Time Work
  3. Comparing the Work Attitudes and Behaviors of Part-Time and Full-Time Employees
  4. Differences among Part-Time Employees
  5. Theoretical Approaches for Understanding Differences across Part-Time and Full-Time Work
  6. Conclusion
- Further Reading

## GLOSSARY

**contingent work** Work done by part-time employees employed only temporarily (e.g., until a specific organizational goal is met or an operational subtask is finished).

**core jobs** Employment opportunities commonly viewed as permanent full-time contract positions.

**peripheral work** Any job in which the individual does not have an explicit or implicit contract for long-term employment or steady hours of employment.

Part-time work has generally been defined as whether an employee reports himself or herself as working part-time. Although the part-time workforce is acknowledged as being a heterogeneous group, most research has concentrated on comparing part-time employees with full-time employees across a range of criteria, most notably work attitudes and behaviors. Differences in work attitudes and behaviors are frequently found between part-time and full-time employees,

although the findings are inconclusive and very few theoretical frameworks have been put forward to understand or explain empirical differences. Likely explanations for different attitudes and behaviors across the two groups relate to theories of social exchange and motives for accepting employment.

## 1. INTRODUCTION

During the past 30 years or so, organizations have increasingly turned to part-time employees because they provide greater scheduling flexibility, meet market demands more efficiently, and reduce wage and benefit costs. For example, between 1971 and 1996 in Great Britain, the proportion of part-time employees nearly doubled from 15 to 29% of the workforce, increasing by 2.6 million. The proportion of part-time employees is similar in the United States, where part-time employees constituted 24% of the workforce and numbered more than 22 million employees in 2001, with the proportion doubling during the previous 40 years. Part-time employees also constitute the major proportion of the workforce for entire industries (e.g., service, retail), and part-time work is much more likely to be performed by certain demographic groups (e.g., women, older/younger workers).

Although this large part-time workforce is clearly important in terms of the increasing dependence on its productivity as well as its social implications, the part-time workforce is also interesting from a psychological

perspective. For example, some researchers have suggested that part-time workers view work differently than do full-time workers; thus, the way in which the jobs of part-time workers are designed, the human resource practices available to them, and the way in which they are managed all should differ from those of full-time employees.

Organizational researchers have largely neglected researching part-time employment. The majority of studies have simply compared part-time employees with full-time employees across a range of attitudinal outcomes. Too many of these studies have been atheoretical in design and simply documented empirical differences that are of limited lasting scientific value. Moreover, important areas of inquiry (e.g., differences among part-time workers) have been underresearched.

## 2. DEFINING PART-TIME WORK

There are two necessary components to a definition of part-time work. The first concerns the number of hours worked, and the second concerns comparisons with other forms of employment relationships. Fundamentally, part-time work has been defined by the number of hours worked per week. Government workforce data typically define part-time work as less than 30 to 35 hours per week. Part-time work has also been defined by whether the employee reports himself or herself as working part-time. In organizational research, the self-report definition of part-time work status (i.e., whether employed on a part-time or full-time basis) is used in the great majority of cases.

The second definitional issue is more complex and involves the distinction between part-time work and other forms of employment relationships. The key factor here is the extent to which part-time workers can be regarded as “contingent” or “peripheral” workers. Peripheral employment is equivalent, apart from minor differences, to the conceptual notion of contingent employment. These conceptualizations are rooted in core-periphery models of the workforce in which core jobs are commonly viewed as permanent full-time contract positions. Peripheral or contingent work has been defined as any job in which the individual does not have an explicit or implicit contract for long-term employment or one in which the minimum number of hours can vary in a nonsystematic way.

In early studies, part-time employees were synonymous with peripheral employees. However, there is

increasing support for the view that only certain types of part-time employees can be regarded as contingent employees. This view holds that permanent part-time employees are misclassified as contingent simply because they are not employed full-time and that permanent part-time employees are more likely to have contract conditions similar to those of core employees. For example, many permanent part-time positions permit long-term employment. Only certain types of part-time employees, such as part-time employees who are also employed on a temporary basis, are likely to fall within the definition of contingent workers. Distinguishing part-time work from other forms of employment is important because working part-time might not necessarily entail a subjective sense of being peripheral.

In conclusion, part-time work is typically defined by whether the employee reports himself or herself as working part-time. A further definitional issue is sensitivity to differences among part-time jobs, such as the permanence of the positions, because such characteristics are likely to be of substantive theoretical importance.

## 3. COMPARING THE WORK ATTITUDES AND BEHAVIORS OF PART-TIME AND FULL-TIME EMPLOYEES

Over the past 30 years or so, there has been a slow stream of studies examining part-time work where the dominant approach has been to compare the work attitudes and behaviors of part-time employees with those of full-time employees. Only the relationship between work status and job satisfaction and commitment have received any serious level of coverage, although some studies have considered other relationships such as work status and organizational climate, job involvement, and work characteristics.

Empirical research comparing part-time employees with full-time employees consistently finds differences on various work attitudes and behaviors. However, the findings are inconclusive. For example, across studies, part-time employees have been found to report higher, lower, and equivalent levels of job satisfaction compared with full-time employees. When facets of job satisfaction (e.g., satisfaction with pay, satisfaction with supervisor) have been considered, inconsistent findings remain. What these more in-depth studies have shown, however, is that ratings of satisfaction of part-time employees and those of full-time employees

might not act uniformly across facets of satisfaction. For example, part-time employees may report lower satisfaction with job content but higher satisfaction with social aspects of the work. These studies have drawn attention to the fact that nonsignificant differences in overall ratings of job satisfaction across work statuses may mask differences regarding facets of satisfaction.

Similarly, inconsistent findings have emerged from comparisons of the commitment levels of part-time employees and those of full-time employees. If one looks at the much smaller number of studies that have considered the relationship between work status and work-related behavior in the form of employee turnover, absenteeism, and job performance, the findings are again inconclusive.

A few studies have investigated work status as a moderator of the relationships among a range of job-related attitudes and behaviors. This is a potentially important issue because it may indicate whether predictors of attitudes for full-time employees, such as the psychological contract, are of equal utility to understanding part-time attitudes at work. From existing studies, there is little evidence of work status as a moderator, implying that part-time employees will respond in the same manner as will full-time employees in terms of human resource practices, rewards, and management strategies.

The main conclusion to be drawn from this brief review is that although attitudinal and behavioral differences are frequently observed between full-time and part-time employees, these empirical studies yield inconsistent results. Three further points are worth noting. First, findings on whether work status acts consistently across outcomes within the same study are mixed. Some studies have found that part-time employees report higher ratings of both job satisfaction and commitment, whereas other studies have found that part-time employees report higher ratings of job satisfaction but lower ratings of commitment. Collectively, these studies suggest that theories used to explain differences in job satisfaction between part-time and full-time employees might not be appropriate for organizational commitment. Second, some construct measures used in studies comparing part-time employees with full-time employees are designed with full-time employees in mind, raising the question of whether any differences found across work statuses on such measures can be considered as genuine differences on the underlying construct or instead should be deemed as artifacts of the method used. Finally, a narrow range of methods

has been used, with cross-sectional self-report questionnaire surveys being adopted nearly exclusively, where the limitations of such methods are well documented.

#### 4. DIFFERENCES AMONG PART-TIME EMPLOYEES

Until the past decade or so, the heterogeneity among the part-time work group was rarely considered. For example, there were no clear theoretically based rationales for including demographic variables in research designs. Demographic variables were used to refer to age, sex, and marital status as well as to conditions such as organizational tenure and education. The control of demographic variables varied considerably across studies, with some studies not reporting the impact of any demographic variables and others using such unique sets of control variables that generalization was difficult.

In 1990, Feldman attempted to reconceptualize the research on part-time work. He argued that there were also major theoretical distinctions within the part-time work group and identified several key dimensions on which part-time work arrangements can vary, including the following:

- Whether the employee is employed on a permanent or temporary basis
- Whether the employee is recruited directly by the company or through an agency
- Whether the employment is year-round or seasonal
- Whether the job is the employee's main job or a second job
- Whether the individual accepted the part-time job voluntarily or involuntarily due to the unavailability of a full-time position

Feldman argued that these dimensions relate to attitudinal and behavioral outcomes. For example, part-time workers are hypothesized to be more likely to quit when they work part-time on an involuntary basis. Furthermore, there are three key demographic groups within the part-time workforce: students, married women with children, and retirees. These groups are indirectly related to attitudinal and behavioral outcomes in that certain demographic groups will gravitate to certain types of part-time work. For example, retirees are more likely to gravitate to voluntary part-time working arrangements as their main jobs, and these jobs are hypothesized as being more satisfying.

Feldman's research has drawn attention to some of the key dimensions of part-time work, but very few studies have explored whether these dimensions explain differences found in attitudes and behaviors within the part-time work group or, indeed, across part-time and full-time employees. In addition to highlighting important distinctions within the part-time work group, these categories draw attention to the need to control for confounding variables when comparing part-time employees with full-time employees. Researchers are now careful to locate their samples in terms of these categories.

Essentially, research in this area is concerned with what being "part-time" actually means. Some of the main dimensions of part-time work appear to have been identified, but these dimensions have received very little empirical attention and do not appear to be exhaustive.

## **5. THEORETICAL APPROACHES FOR UNDERSTANDING DIFFERENCES ACROSS PART-TIME AND FULL-TIME WORK**

It should be noted that research comparing part-time employees with full-time employees has been largely atheoretical in design. When theories have been put forward to explain differences in attitudes and behaviors, mostly partial inclusion and/or frames of reference, they are applied post hoc to rationalize findings. The theories have received minimal attention and have tended to be applied in a rather cursory way.

### **5.1. "Part-Time" as a Proxy for Other Key Variables**

The composition of the part-time workforce is known to differ from that of the full-time workforce in terms of various demographic (e.g., gender), organizational (e.g., tenure), and labor market (e.g., temporary/permanent) variables. Researchers have argued that these differences explain different attitudes and behaviors found across part-time and full-time employees and that once these differences are controlled for work status, they will no longer be associated with job attitudes and behaviors. In other words, differences in attitudes and behaviors between part-time and full-time employees are not due to the former's "part-timeness." Rather, they are due to the fact that part-time employees are more likely to come

from different demographic, organizational, and/or labor market groups; that is, "part-time" acts as a "proxy" for other key variables that are associated with working part-time.

A major problem with this argument is that proxy explanations constitute neither an adequate theoretical explanation nor an adequate psychological explanation for why there may be differences in attitudes and behaviors between part-time employees and full-time employees, and the use of proxy explanations may mask more meaningful theoretical explanations.

### **5.2. Partial Inclusion**

Part-time employees are sometimes said to be "partially included" in the workplace. By spending less time in the workplace and being more involved in extra-organizational roles, they feel less included in organizations than do full-time employees. A number of limitations to this theory have been documented. The theory has been used in contradictory ways to explain attitudes where, for instance, feeling less included has been used to explain both higher and lower levels of job satisfaction. This theory has never been empirically tested, and it is not clear how the theory could be operationalized given that it is such a poorly elaborated concept.

This theory is perhaps better seen as a broad framework within which better understood theories can be placed. For instance, partial inclusion can be seen as embodying a number of ideas that could be explored separately in their own right, including the psychological contract, theories of the relationship between work and nonwork, and social identity theory.

### **5.3. Frames of Reference**

It has been argued that part-time employees have a "frame of reference" that is different from that of full-timers, where the former evaluate characteristics of their work environment by considering how well they compare with those of another self-selected group. For example, part-time employees may be less satisfied with their pay if it compares unfavorably with that of their full-time counterparts.

This approach has been criticized in a similar way to partial inclusion in that it also has been used in contradictory ways to explain attitudes, is untested, and is too vague to operationalize. Furthermore, frame of reference comparisons can be used in an endless number of ways in terms of to whom, and over what, part-time employees choose to compare themselves.

#### 5.4. Social Exchange and the Psychological Contract

The two exchange-based approaches of the psychological contract and social exchange theory have also been used to explain different attitudes and behaviors across work statuses. It has been argued that there are a number of factors that affect the exchange of contributions for inducements across part-time and full-time employees. For example, part-time employees receive fewer inducements such as benefits, task variety, and opportunities for advancement; part-time employees have lower expectations about what they should get from the organization; and part-time employees are more likely to be subject to “Theory X”-type management. These factors will have the effect of creating a perception of perceived inequity or psychological contract violation across work statuses, and they are more likely to lead to part-time employees developing economic relationships, rather than social exchange relationships, with employers. This, in turn, will affect outcome attitude and behaviors such as job satisfaction and organizational citizenship behavior.

Although some of the assumptions behind the social exchange-type approaches are questionable (e.g., that organizations treat part-time employees less favorably, that part-time employees expect less from their work), these approaches have been empirically tested and found to at least partially explain attitude and behavior differences across part-time and full-time employees.

#### 5.5. Preferred Work Status/Schedule

Employees working their preferred work status (i.e., part-time vs full-time) or schedule (i.e., preferred number of hours per week) are often said to have more positive work attitudes. Previous research has found that a discrepancy between scheduled and preferred work hours for both full-time and part-time employees is related to work attitudes and that preferred work status matters more to part-time employees than to full-time employees.

What is not clear is the precise theoretical explanation for why such a discrepancy affects attitudes and behavior. One possible explanation is that employees involuntarily situated in their work arrangements have

lower levels of attachment. An alternative explanation drawn from social exchange theory is that employees working their preferred schedules achieve a better balance between contributions and inducements.

### 6. CONCLUSION

Despite a long history of comparing the work attitudes and behaviors of part-time employees with those of full-time employees, there have been relatively few studies conducted. Studies frequently find attitudinal and behavioral differences across part-time and full-time employees, but what emerges quite categorically is that findings are not consistent across studies. Many previous studies have been atheoretical in design and, thus, have contributed little to understanding the experience of part-time work. Since the 1990s, research on part-time work has been conducted under a stronger theoretical basis, and two particularly promising approaches to a better understanding of the experience of part-time work are social exchange and motives for accepting part-time employment.

### Acknowledgment

The author thanks the Leverhulme Trust for its support in the preparation of this article.

### See Also the Following Articles

Contracts ■ Cooperation at Work ■ Person–Environment Fit ■ School-to-Work Transition ■ Work Environments ■ Work Teams

### Further Reading

- Barling, J., & Gallagher, D. (1996). Part-time employment. *International Review of Industrial and Organizational Psychology*, *11*, 243–277.
- Conway, N., & Briner, R. B. (2002). Full-time versus part-time employees: Understanding the links between work status, the psychological contract, and attitudes. *Journal of Vocational Behavior*, *61*, 279–301.
- Feldman, D. C. (1990). Reconceptualizing the nature and consequences of part-time work. *Academy of Management Review*, *15*(1), 103–112.



# Pediatric Psychology

Merritt M. Jensen and Anne E. Kazak

Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

1. Definitions: Pediatric Psychology and Pediatric Psychologist
2. Pediatric Psychology: Integrating Psychology and Pediatrics
3. Emergence of Pediatric Psychology
4. Growth of Pediatric Psychology within Medical Settings
5. Major Research Areas within Pediatric Psychology
6. Clinical Practice of Pediatric Psychology
7. Pediatric Psychology Training  
Further Reading

## GLOSSARY

**adherence to medical treatment** The complex interactions among patients, families, and health care providers in ensuring that treatment regimens are followed as intended.

**assessment** The evaluation process to describe domains related to physical, intellectual, cognitive, emotional, social, and family functioning.

**bereavement** The grief process following the loss (typically the death) of a loved one.

**childhood chronic illness** An ongoing medical condition requiring long-term medical care and follow-up (e.g., sickle cell disease, HIV/AIDS, asthma, cancer, juvenile diabetes); it often affects the life of the affected child and his or her family for an extended time period and can interfere with daily functioning.

**cognitive-behavioral therapy (CBT)** Psychological intervention that focuses on understanding how maladaptive thought patterns and beliefs lead to behavior that interferes with optimal social, emotional, and behavioral

functioning; CBT intervention concentrates on altering thoughts and beliefs so as to change behaviors and emotional states toward a more adaptive perspective.

**evidence-based treatment** An intervention for which there is empirical support.

**interdisciplinary** Involving two or more disciplines working toward a common goal; for example, an interdisciplinary team in pediatric oncology might include oncologists, nurses, nutritionists, social workers, physical therapists, occupational therapists, child life specialists, psychologists, and physicians from other medical disciplines.

**intervention** A systematic approach to facilitate change in a desired direction that may focus on individuals or families (e.g., cognitive-behavioral therapy, family therapy, pain management, consultation).

**procedure** A medical treatment (e.g., having blood drawn, undergoing a magnetic resonance imaging [MRI] scan).

**psychosocial** Involves the integration of psychological and social aspects of a person's well-being; sometimes, it is expanded as *biopsychosocial* to include the role of physical (biological) factors.

Pediatric psychology continues to grow and develop as an area of specialty and expertise within clinical psychology. This article highlights the interdisciplinary nature of the field. Research in pediatric psychology, including the emergence of evidence-based interventions, and their use in the clinical practice of pediatric psychology is described. Training requirements for pediatric psychologists are also discussed.

## **1. DEFINITIONS: PEDIATRIC PSYCHOLOGY AND PEDIATRIC PSYCHOLOGIST**

As described by the Society of Pediatric Psychology, pediatric psychology “is dedicated to research and practice addressing the relationship between children’s physical, cognitive, social, and emotional functioning and their physical well-being, including maintenance of health, promotion of positive health behaviors, and treatment of chronic or serious medical conditions.” Pediatric psychologists are experts in the areas of developmental, emotional, behavioral, social, and family aspects of acute and chronic illness and injury. They work with children, adolescents, families, and health care teams around the areas of the physical, emotional, social, and cognitive aspects of illness or injury. Within interdisciplinary pediatric health care teams, pediatric psychologists evaluate and intervene to enhance the well-being of children and families.

## **2. PEDIATRIC PSYCHOLOGY: INTEGRATING PSYCHOLOGY AND PEDIATRICS**

The importance of a close working relationship between psychology and pediatrics has long been viewed as essential to providing quality health care for children, adolescents, and their families. In 1965, Jerome Kagan spoke of the need for a “new marriage” between the fields of psychology and pediatrics to meet the needs of children in health care settings. In 1967, Logan Wright published a seminal article delineating the key elements of the field of pediatric psychology, including a call for empirical research supporting clinical practice.

## **3. EMERGENCE OF PEDIATRIC PSYCHOLOGY**

Pediatric psychology is relatively new, having been founded in 1968 as the Society of Pediatric Psychology. Dortha Ross, Lee Salk, and Logan Wright were early leaders in the field. Interest in working with pediatric patients and their families within medical settings continued to grow during the 1970s. In 1973, the *Journal of Pediatric Psychology* was established and today remains the flagship journal in the field. This journal publishes

articles focused on original research, professional practice, and theory in pediatric psychology. Steady increases in membership and growing interest in working with pediatric patients and their families within medical settings occurred during the 1980s and 1990s, culminating in the Society of Pediatric Psychology gaining division status (Division 54) within the American Psychological Association in 1999.

## **4. GROWTH OF PEDIATRIC PSYCHOLOGY WITHIN MEDICAL SETTINGS**

The numbers of pediatric psychologists are increasing within hospitals and other medical settings. Working together with interdisciplinary medical teams, pediatric psychologists serve as consultants and collaborators within inpatient and outpatient hospital settings, primary care and subspecialty clinics, and academic medical settings and schools.

## **5. MAJOR RESEARCH AREAS WITHIN PEDIATRIC PSYCHOLOGY**

Scientific research is an essential part of pediatric psychology. Childhood chronic illness is a major focus of the field, although treatment of acute illness and attention to health promotion and prevention are also emphasized. Designing and evaluating evidence-based treatments and interventions is an important recent focus.

### **5.1. Chronic Illness**

Frequently studied childhood chronic illnesses include (but are not limited to) asthma, cancer, cystic fibrosis, diabetes, HIV/AIDS, juvenile rheumatoid arthritis, organ transplantation, seizure disorders, pain syndromes, neuromuscular diseases, and sickle cell disease. Childhood chronic illness research focuses on how multiple aspects of diseases and treatments affect children, adolescents, and families, including adjustment to and coping with chronic illness, adherence to medical procedures and treatments, pain management, feeding and sleep disturbances, intellectual and cognitive development, social and emotional functioning with peers and family, family adjustment to chronic illness, and promotion of health and quality of life.

## 5.2. Assessment

Systematic assessment, using standardized psychological tests, can facilitate understanding of the intellectual, cognitive, social, and emotional impacts of illnesses or injuries. The assessment process provides essential information for developing effective treatments and necessary information to health care teams, families, and school personnel. For example, many diseases or illnesses adversely affect the child's central nervous system, causing physical, cognitive, and/or socio-emotional deficits that may influence the child's ability to perform competently at home, with peers, and in school.

## 5.3. Intervention: Evidence-Based Treatments

Carefully designed and controlled studies evaluating the effectiveness of manual-based interventions with children and their families have increasingly become a part of research and practice in the field of psychology. Beginning in 1999, the *Journal of Pediatric Psychology* published a series of reviews describing evidence-based treatments in pediatric psychology, including interventions for treating recurrent pediatric headache, recurrent abdominal pain, procedure-related pain, disease-related pain, pediatric obesity, severe feeding problems, disease-related symptoms, bedtime refusal and nighttime wakings, nocturnal enuresis, constipation and encopresis, and adherence to treatment regimens.

There are several effective treatments that may be used to treat children's pain. For example, cognitive-behavioral therapy (CBT) is an effective, well-established treatment for procedure-related pain, disease-related pain, and recurrent abdominal pain in children and adolescents. Examples of CBT techniques include teaching the child and family relaxation skills such as the use of deep breathing exercises, using guided imagery, and using distraction such as counting out loud or playing with toys. Active coaching by a caregiver and/or medical team member encourages the child to use self-soothing coping skills, which is another effective CBT intervention for pain. Relaxation and self-hypnosis are useful treatments for recurrent pediatric headaches. Biofeedback, whether alone or combined with relaxation or other CBT treatments such as those described previously, is also promising for headaches. In addition, guided imagery is an effective treatment for children with cancer who are undergoing chemotherapy.

Effective treatments for childhood obesity target several behavioral areas at one time, for example, promoting healthy eating patterns while also increasing physical exercise. For severe feeding problems, rewarding the child when he or she engages in desired behaviors (e.g., eating a particular food) and ignoring undesired responses (e.g., spitting out food) are useful interventions.

Evidence-based interventions for bedtime refusal and other childhood difficulties, including enuresis and encopresis, are promising. Educating parents about healthy sleep hygiene during the prenatal and newborn periods is helpful for bedtime refusal and night wakings in young children. Implementing a positive bedtime routine is another promising intervention. Extinction techniques, including parental ignoring of the child's cries so that the child learns to fall asleep on his or her own, is also effective but is often less feasible given that parents might not be willing to attempt this strategy. For nocturnal enuresis (i.e., repeated nighttime urination in places other than a toilet, e.g., in bed), addressing both the biological and behavioral aspects of the problem is most effective. Using a urine alarm combined with the medication desmopressin is a promising intervention. Fewer effective, evidence-based treatments for constipation and encopresis (i.e., repeated bowel movements in locations other than a toilet, e.g., in underwear) have been identified. The most promising treatments combine medical intervention (e.g., fiber, laxatives) with a behavioral component such as positive reinforcement.

Understanding the most effective ways in which to encourage adherence to medical regimens is an essential area of pediatric psychology research. Increased parental supervision and monitoring of medication use, coupled with increased attention and support from a health care team, improves adherence for children and adolescents with asthma. For individuals with juvenile rheumatoid arthritis, treatments using behavioral strategies (e.g., positive social attention, feedback from the child's family, token reward systems) are useful. Multicomponent treatments focused on teaching problem-solving, negotiation, and decision-making skills are promising for helping children, adolescents, and families to improve adherence to complex regimens for juvenile (type 1) diabetes.

## 5.4. Prevention

Often, the best therapeutic intervention is to prevent the occurrence of an illness or injury. Pediatric psychologists are actively involved in promoting healthy attitudes

and behaviors in children, adolescents, and families and in working to prevent disease, illness, and injury. Research targeting lifestyle choices and behaviors during childhood and adolescence, including healthy eating, exercise, drug avoidance, injury prevention, and safe sexual practices, encompasses the many prevention topics studied by pediatric psychologists.

### **5.5. Family Systems**

Increasingly, the focus of pediatric psychology has transitioned from focusing exclusively on the child and adolescent to a more systems-focused perspective emphasizing the important roles that parents, caregivers, siblings, relatives, peers, teachers, the health care team, and other community contacts play in the child's functioning in relation to illness or injury. Families facing illness must balance medical management of the disease with feelings of fear, uncertainty, and grief. How successfully families are at navigating this challenge is influenced by the families' pre-illness functioning and adjustment, levels of support and resources, and the medical aspects of the illness. Families who demonstrate strength-based skills, such as effective communication, clear family boundaries, active coping, flexibility, and the ability to balance the demands of the illness with other family responsibilities, are more likely to demonstrate increased well-being throughout course of the illness.

## **6. CLINICAL PRACTICE OF PEDIATRIC PSYCHOLOGY**

### **6.1. Settings in Which Pediatric Psychologists Work**

Many pediatric psychologists are employed in, or have close affiliation with, hospitals. They work in emergency, inpatient, and/or outpatient facilities in freestanding children's hospitals, university-affiliated medical centers, community hospitals, and satellite clinics. A smaller percentage of pediatric psychologists are based in primary care, and others work in private practice serving children and families. Pediatric psychologists are also employed within schools and in other community settings.

### **6.2. Roles of Pediatric Psychologists**

The professional roles of pediatric psychologists are as diverse as the environments in which they work.

Pediatric psychologists often serve as consultants to interdisciplinary medical teams within inpatient, outpatient, or primary care settings. Assessment of psychosocial, neuropsychological, and/or neurodevelopmental functioning of children with acute or chronic illness, injury, or developmental delay is another important activity. Pediatric psychologists often work directly with children and their families around the adjustment to acute or chronic illness or injury, including how the illness or injury will affect the children's cognitive, emotional, and social functioning. Adjustment to illness or injury may also include addressing adherence to medical regimens, coping with painful or stressful medical procedures, and exploring how the child's and family's beliefs about the adversity of the illness or injury are affecting their thoughts, feelings, and relationships with others. Bereavement work with children and families facing death due to illness or injury is another important activity.

## **7. PEDIATRIC PSYCHOLOGY TRAINING**

Pediatric psychologists are typically licensed, doctoral-level psychologists. Many pediatric psychologists have backgrounds in clinical psychology, whereas others have training in areas such as developmental or health psychology. All have completed 4 to 5 years of graduate-level coursework and clinical training, a 1-year intensive clinical internship in a hospital or other setting related to health care, and 1 or more years of postdoctoral training.

In addition to didactic and experiential learning in basic psychological principles, pediatric psychologists receive specialized training and instruction in health care issues that affect children, adolescents, and their families. Focused training in pediatric psychology often includes instruction in the areas of adherence and health promotion, management of painful medical procedures, social and emotional aspects of pediatric medicine, and prevention of illness and injury. While training within diverse pediatric health care settings, pediatric psychologists gain advanced experience in working with children and adolescents with a broad range of medical illnesses.

Pediatric psychologists continue to use evidence-based research to inform clinical practice within interdisciplinary care environments. With the growth of this

relatively newly emerging field, the role for pediatric psychologists will clearly broaden and diversify as the needs of changing medical environments dictate.

### See Also the Following Articles

Child Development and Culture ■ Diverse Cultures, Dealing with Children and Families from ■ Health and Culture

### Further Reading

- American Psychological Association. (n.d.). *Division 54: Society of Pediatric Psychology*. [www.apa.org/about/division/div54.html](http://www.apa.org/about/division/div54.html).
- Drotar, D. (1995). *Consulting with pediatricians: Psychological perspectives*. New York: Plenum.
- Drotar, D. (2000). *Promoting adherence to medical treatment in chronic childhood illness: Concepts, methods, and interventions*. Mahwah, NJ: Lawrence Erlbaum.
- Holden, E. W., Deichmann, M. M., & Levy, J. D. (1999). Empirically supported treatments in pediatric psychology: Recurrent pediatric headache. *Journal of Pediatric Psychology, 24*, 91–109.
- Janicke, D. M., & Finney, J. W. (1999). Empirically supported treatments in pediatric psychology: Recurrent abdominal pain. *Journal of Pediatric Psychology, 24*, 115–127.
- Jelalian, E., & Saelens, B. E. (1999). Empirically supported treatments in pediatric psychology: Pediatric obesity. *Journal of Pediatric Psychology, 24*, 223–248.
- Journal of Pediatric Psychology* (1975–present). Oxford University Press. <http://jpepsy.oupjournals.org/>.
- Kagan, J. (1965). The new marriage: Pediatrics and psychology. *American Journal of Diseases of Childhood, 110*, 272–278.
- Kazak, A. (2002a). *Journal of Pediatric Psychology (JPP)*, 1998–2002; Editor's vale dictum. *Journal of Pediatric Psychology, 27*, 653–663.
- Kazak, A. (2002b). *Journal of Pediatric Psychology: A brief history (1969–1999)*. *Journal of Pediatric Psychology, 25*, 463–470.
- Kazak, A. E., Rourke, M. T., & Crump, T. A. (2003). Families and other systems in pediatric psychology. In M. Roberts (Ed.), *Handbook of pediatric psychology* (3rd ed.). New York: Guilford.
- Kerwin, M. E. (1999). Empirically supported treatments in pediatric psychology: Severe feeding problems. *Journal of Pediatric Psychology, 24*, 193–214.
- Lemanek, K. L., Kamps, J., & Brown Chung, N. (2001). Empirically supported treatments in pediatric psychology: Regimen adherence. *Journal of Pediatric Psychology, 26*, 253–275.
- McGrath, M. L., Mellon, M. W., & Murphy, L. (2000). Empirically supported treatments in pediatric psychology: Constipation and encopresis. *Journal of Pediatric Psychology, 25*, 225–254.
- McQuiad, E. L., & Nassau, J. H. (1999). Empirically supported treatments of disease-related symptoms in pediatric psychology: Asthma, diabetes, and cancer. *Journal of Pediatric Psychology, 24*, 305–328.
- Mellon, M. W., & McGrath, M. L. (2000). Empirically supported treatments in pediatric psychology: Nocturnal enuresis. *Journal of Pediatric Psychology, 25*, 193–214.
- Mindell, J. (1999). Empirically supported treatments in pediatric psychology: Bedtime refusal and night wakings in young children. *Journal of Pediatric Psychology, 24*, 465–481.
- Mullins, L. L., & Chaney, J. M. (2001). Pediatric psychology: Contemporary issues. In C. E. Walker, & M. C. Roberts (Eds.), *Handbook of clinical child psychology* (3rd ed.). New York: John Wiley.
- Powers, S. (1999). Empirically supported treatments in pediatric psychology: Procedure-related pain. *Journal of Pediatric Psychology, 24*, 131–145.
- Roberts, M. C. (2003). *Handbook of pediatric psychology* (3rd ed.). New York: Guilford.
- Walco, G. A., Sterling, C. M., Conte, P. M., & Engel, R. G. (1999). Empirically supported treatments in pediatric psychology: Disease-related pain. *Journal of Pediatric Psychology, 24*, 155–167.
- Wright, L. (1967). The pediatric psychologist: A role model. *American Psychologist, 22*, 323–325.



# Perception and Culture

J. B. Deręowski

University of Aberdeen, Aberdeen, United Kingdom

1. Introduction
  2. The Notion of Culture
  3. Visual Illusions
  4. Pictorial Perception
  5. Perception of Patterns
  6. Eidetic Imagery
  7. Studies of Perception of Color
  8. Perception of Time
  9. Concluding Remarks
- Further Reading

## GLOSSARY

- cross-cultural** An imprecise term used to describe comparisons between groups differing in culture but also often differing in other respects (e.g., gene pool, environment).
- eidolity** Characteristic of flat figures responsible for perception of illusory depth.
- epitomy** Characteristic of flat figures responsible for their recognition as representing objects.
- horizontal-vertical illusion** Misperception of T-shaped and L-shaped figures.
- Kwengo calipers** A test in which individuals adjust a pair of calipers to an angle presented in perceptually flat or perceptually solid figures.
- Muller-Lyer illusion** Misperception of the length of lines bounded by arrowheads.

Cross-cultural studies of perception show differences among populations. They contribute more to the un-

derstanding of the perceptual processes than do analogous studies of nonhuman species. In spite of this, cross-cultural studies have not flourished greatly and have been concerned mostly with visual perception.

## 1. INTRODUCTION

Cross-cultural differences have long been the subject of travelers' reports. For example, it was reported that somewhere near the present-day Afghanistan, Alexander the Great's army encountered men who had one foot much larger than the other. These feet, in addition to their usual function, served as parasols when the men lay prone on their backs.

Systematic cross-cultural investigations are relatively recent, although not far temporally removed in time from the beginnings of scientific psychology. The first significant cross-cultural psychological study, the scope of which has never been exceeded, was carried out as part of the Cambridge Anthropological Expedition to the Torres Straits. Several psychologists, the most prominent of whom was W. H. R. Rivers, collected data on several perceptual functions such as olfactory and cutaneous sensations. However, much of the data collected, just like the data most ardently sought by research workers of more recent years, concerned visual perception. Later studies became even narrower in the scope of the phenomena investigated. With the notable exception of Reuning's investigations of the visual abilities of

the Bushmen, the studies are generally concerned with only one aspect of vision.

## 2. THE NOTION OF CULTURE

The term “cross-cultural” is vague because there is no definition of culture that would satisfy experimental psychologists, for example, by enabling them to decide whether two individuals come from different cultures with as great an aplomb as the psychologists can decide whether the individuals differ in terms of a perceptual skill or short-term memory. However, it is generally believed that the vectors of genetic endowment and the impact of the natural environment are insufficient to explain behavior and that the manmade components of the environment, such as language, architecture, customary rituals, and social organization, form a third vector, the “cultural” vector.

Therefore, there is a widespread conviction that the vectors that determine individuals’ responses are three-fold: genetic, environmental, and cultural. However, it is also generally accepted that the vectors are not mutually independent. For example, it is accepted that, when a “hunting and gathering” population changes its mode of life (a cultural change), takes to settled agriculture, and clears the land for this purpose (an environmental change), the following changes may follow. Agricultural activity and a settled population may offer an excellent breeding ground for malarial mosquitoes. Therefore, incidents of malaria will increase, as will the proportion of individuals suffering from sickle cell anemia. The latter individuals have certain physical disadvantages but are immune from malaria, and their immunity is genetically transmissible (a genetic change). Thus, culture, environment, and the genetic character of the population change together.

Although the vectors just described cannot be isolated in practice, they are heuristically desirable because they introduce a degree of tidiness into psychologists’ thinking. Such fine distinctions seldom intrude into psychologists’ experimental work. Most studies of perception are not concerned with the genetic vector at all; indeed, only a few attempts to separate the effects of the environment from the effects of culture have been made. This is not to say that results of such studies lack merit but rather to stress that they provide rather confounded data so far as the three vectors are concerned. However, such theoretically impure studies are of immediate practical use because they offer guidance, for example, for selection of personnel when the causes of particular

abilities are of no consequence but the presence or absence of such abilities is.

## 3. VISUAL ILLUSIONS

One of the phenomena investigated by Rivers in the course of the Torres Straits expedition was that of perception of visual illusions. It was a study intended to compare “Western” and “non-Western” populations on a number of illusion figures that, at the time, were thought to constitute a homogeneous category of stimuli. It was not driven by any clearly defined hypothesis. The same phenomena were investigated by Segall and colleagues more than 60 years later. However, their investigation was inspired by specific hypotheses concerning the extent to which environmental experience affects perception. Segall and colleagues sampled a number of distinct populations inhabiting distinct environments and having different cultural artifacts. The conclusions of the two studies were in agreement in that both suggested that common visual illusions do not form a homogeneous body of phenomena and that their effects differ among populations.

Segall and colleagues attributed the differences observed to the differences in visual experience, for example, to openness of vistas (i.e., the extent to which people can see open spaces—savannah dwellers can do so to a larger extent than can city dwellers) in the case of the horizontal-vertical illusion and to “carpentered environment” (i.e., the extent to which artifacts of a culture incorporate straight edges—Western artifacts do so to a larger extent than do those of, say, Zulus) in the case of the Muller-Lyer illusion. It should be noted that the cultural vector is explicitly involved in the latter case. Of course, it is not entirely absent in the former case.

## 4. PICTORIAL PERCEPTION

Surprising responses of individuals drawn from exotic populations to pictures have frequently been mentioned by travelers and anthropologists. These have ranged from reports of the complete inability of certain populations to perceive that pictures represented objects to reports of responses indicating that the depicted objects were seen as real (e.g., that the portrayed people were alive but asleep). These observations, which came from reputable observers, were casual and not based on data obtained in the course of properly conducted studies.

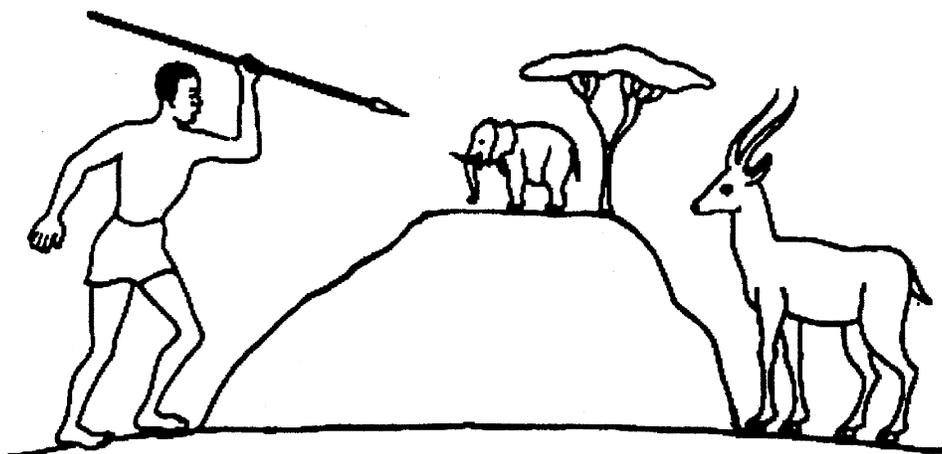


FIGURE 1 One of Hudson's items: A test of pictorial perception.

The problem was first tackled systematically when misperception of pictures was found to be an obstacle to working in industry. It was found that the difficulties experienced were those associated with the failure to perceive the third dimension in portrayals. This was detected by asking individuals to name distinct elements in the test picture, such as the man, the spear, and the duiker in Fig. 1 (taken from Hudson's test), and, if they named them correctly, asking for simple judgments of depicted depth. The questions put to the individuals were "Which animal is nearer to the hunter?" and "At which animal is the hunter aiming his spear?" The answer "elephant" to either of these questions was taken as showing difficulties in perception of pictorial depth.

Following this study, several other measures were developed. These measures did not rely on verbal responses but instead called for reconstructing the pictured scenes or for copying an "impossible figure," taking advantage of the fact that the impossible figure in question is seen as "impossible" only if it is seen as a depiction of a three-dimensional object.

Because pictures are such an important means of communication whenever linguistic communication is not possible, and because the population in such cases is illiterate or has a language that is not widely known, these findings aroused widespread interest. A number of investigations were carried out, intent on defining the phenomenon, identifying its causes, and determining the most efficacious portrayals. It was found that perception of pictorial depth is not an "either/or" phenomenon but rather is greatly affected by the nature of pictorial stimuli, so that certain geometrical figures (notably certain illusion figures) evoke

perception of depth even though they are not reminiscent of any objects and are not familiar to the observers. In contrast, certain other figures (e.g., silhouettes) may readily evoke recognition without evoking perception of depth. An example of the latter is the elephant in Fig. 1, which is recognized as an elephant (a three-dimensional entity) but is not seen as three-dimensional. Indeed, even so-called "pin men" or "stick men," consisting merely of a set of straight lines that depict arms, legs, and trunks, are readily recognized as depictions of men, although they contain no depth cues whatsoever. The term "epitemic" was proposed as a descriptor of such figures that portray objects but lack perceptual depth, whereas the term "eidolic" was proposed for those depictions that evoke perception of depth, although they do not portray anything recognizable (e.g., some illusions).

It was also found that formal schooling is associated with the ability to perceive pictorial depth. Therefore, it has been put forward that pictorial perception is a matter of skill. However, it was also found that otherwise very capable engineering students drawn from certain populations were markedly poor at engineering drawings and that unschooled illiterate Bushmen's performance on the Kwengo calipers test (in which individuals must set a pair of calipers to a depicted angle) varied with ethnic origin and, in the case of one Bushmen population, did not differ from the performance of urbanized and regularly schooled Zulus.

It is noteworthy that intercultural differences were reported on other visual tasks. Differences were obtained on measures of implicit shape constancy (i.e., shape constancy as perceived in pictures). African

individuals were less affected by this constancy than were European individuals. Kwengo Bushmen's children, Zulu children, and White South African children were found to differ greatly in their responses to incomplete figures. The Kwengo performed better than did the two other samples. It is highly probable that they did so because they were better at integrating the stimuli. However, the origin of this skill (if in fact it is a skill) is not apparent; Bentley, influenced by Berry, thought that it might be a facet of the skill of hunting. Therefore, it seems plausible that the difficulties of pictorial perception can arise from failure to integrate perceptual input in the manner intended by the artist. The ease with which young children copy "impossible" figures that contain mutually contradictory elements, and the fact that those individuals from relatively pictureless cultures who build well-integrated three-dimensional models find it difficult to copy drawings containing mutually contradictory cues, supports this view.

The reported investigations skipped over the most striking of pictorial failures, that is, the failure to recognize that a picture portrays an object. There are several anecdotal reports of this but only a paucity of systematic evidence. The evidence in hand suggests that such difficulties are rare and further suggests that an inexperienced pictorial perceiver may attach inordinate weight to the material used in making the picture. For example, if paper is not a common cultural artifact, the paper on which a picture is made, rather than the picture itself, may grasp the observer's attention. Therefore, the picture recognition difficulties may be reduced by substituting a familiar material for an unfamiliar one (e.g., substituting coarse cloth for paper).

The pragmatic implications of these findings are clearly those known to every advertiser: It is imperative that pictures be carefully evaluated on a sample of the population for which they are intended.

## 5. PERCEPTION OF PATTERNS

Another aspect of cross-cultural perception that attracted attention is that of reproduction of geometric patterns. It came to the fore when psychological tests such as Kohs blocks were used and scores attained by some populations were found to be so low that the effectiveness of the tests became questionable. In some instances, the tasks were modified to strengthen their interindividual discriminability. In Papua New Guinea, cubes used to reproduce patterns were replaced by flat plaques and trays that included appropriate recesses in

which the patterns were to be reproduced. In South Africa, a newly designed and interesting test required individuals to render symmetrical, by placement of a rectangular fiche, an arrangement of three identical rectangles. Both of these approaches were used for selection, the former in a scholastic setting and the second in an industrial setting.

Suppression of a recalcitrant variable, although justified by practical considerations, does not elucidate the theoretical problem. A number of investigations concerned factors such as the symmetry of the pattern and its apparent stability as well as individuals' tendency to leave the pattern in whatever orientation it was completed. A persistent reported tendency was to rotate the reproductions of patterns in such a way as to increase both their symmetry about the observers' median planes and their apparent stability. The question of whether the errors of orientation are errors of inattention to orientation (where the observers left the correctly reproduced pattern in whatever orientation they happened to complete it) or result from some lawful perceptual vector was addressed by Biesheuvel and Jahoda.

There are cultural groups, such as the Guugu Yimithirr of Northern Queensland, who judge orientations by referring to the geographical coordinates rather than to those of their own bodies (e.g., left and right). The implications of the two types of judgment are demonstrated in Fig. 2, which shows a stimulus viewed from the distinct stances. Those relying on the geographical notions see the orientation of the stimulus as the same from both stances (it is consistently pointing westward). Those relying on their body coordinates see the orientations as different; from Stance A, it points to the left, and from Stance B, it points to the

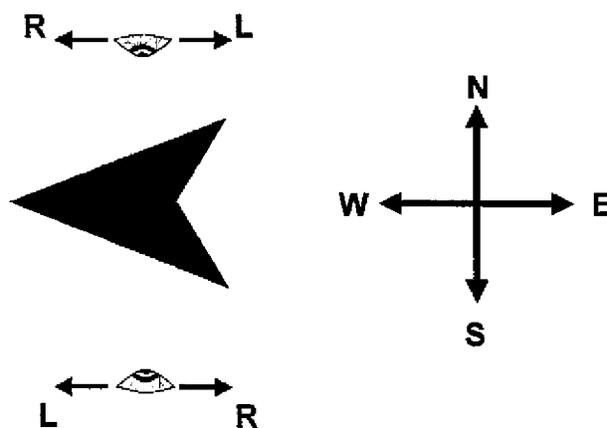


FIGURE 2 Graphical explanation of the effects of the two distinct frameworks on the object's reported orientation.

right. Two issues that ought to be noted here are (a) the justification of describing the phenomenon as perceptual and (b) that arising from the psycholinguistic gloss that would seek the roots of the phenomenon in linguistic use (i.e., “it is consistently pointing westward” vs “it is consistently pointing to the left”).

## 6. EIDETIC IMAGERY

Some people, notably children, experience a special kind of “photographic” imagery, pictorial in nature, that enables them to report with striking accuracy. This eidetic imagery was once believed to be more widespread in illiterate populations. However, empirical data do not support such a notion, and the very phenomenon remains one of the puzzles of psychology.

## 7. STUDIES OF PERCEPTION OF COLOR

Cross-cultural studies of perception of color have been dominated from their very beginning by the issue of the relationship of color vocabularies to perception of color. If the color spectrum can be conceived as a continuum resulting from steadily changing frequency of light waves, such a continuum can be arbitrarily divided in any manner and different languages may present different ways in which to divide it. Speakers of languages dividing the continuum differently will respond to colors differently. For example, speakers of a language that uses the same term for blue and green will respond to these two colors identically. On the other hand, if there is no such continuum, certain frequencies may be perceived as belonging together. That is, if observers, independent of the language, perceive pale green and dark green as mutually related and perhaps perceive a certain kind of green as the epitome of green (called focal green), the color spectrum obviously is not perceived as a uniformly changing continuum and perceived colors themselves impose certain differences in nomenclature. Neither of these polar positions has been sustained empirically, and the fact that pure colors can be blended to evoke percepts identical to those evoked by other pure colors erodes the notion of a perceptual continuum similar to that of, say, weight.

There is empirical evidence that linguistic groups tend to perceive colors differently in accordance with the categorization prevalent in their languages. For example, some Bantu languages use the same word

for green and blue, and speakers of these languages tend to treat these colors as if they were one; however, these effects are not as strong as rigorous application of the linguistic hypothesis would lead one to expect. There is also evidence that when focal colors are sought, these are identical in whatever language is used by the individuals. An additional complication arises because the same color may have quite different significances depending on the object of which it is a color. For example, the focal color detected when using Munsell chips may be quite different from that detected when using toy buffaloes. This effect is likely to be influenced by cultural differences in attention paid to color as opposed to shape of stimuli.

The notion that cross-cultural differences may arise due to differences in the structure of the eyes, namely in the pigmentation of the fundus oculi, has been put forward and investigated and has not yielded convincing data.

## 8. PERCEPTION OF TIME

Travelers' tales about surprising notions of time and of timekeeping are about as numerous as those about perception of pictures, yet the former have not attracted a comparable number of investigators, perhaps because the concepts involved are far more complex. Not only can a distinction be made between “social” time (e.g., as measured by punctiliousness, punctuality, or adherence to time schedules) and “physiological” time (e.g., as measured by the ability to judge and reproduce time intervals), but it also appears that attitude toward time and time-related concepts may vary between urban and rural groups and may affect the memorability of such concepts.

## 9. CONCLUDING REMARKS

It is apparent from this article that the dominant characteristic of cross-cultural studies in perception is their unevenness. Certain areas are pursued with vigor, whereas others are entirely neglected. For example, there are no extensive studies of basic perceptual processes. This is unfortunate because there are issues in relation to which only comparisons of distinct groups of individuals can define the extent of human capacities. For example, the surprising performance of the Bushmen on spatial tasks calls for a more fundamental explanation than does a statement that it may be associated with their hunting practices. These considerations extend beyond the here and now because

cross-cultural studies of perception also embrace studies of cultural artifacts of populations long vanished and sometimes known only by their artifacts. Such are the studies of Paleolithic art, that is, the studies of the early cognitive acts of humans.

It is noteworthy that none of the cultural differences demonstrated suggests that a population has unique perceptual attributes. The differences are in the intensity or frequency of attributes. When reflecting on such differences, one must bear in mind individual differences that are inevitably present in every population.

Probably the most important contribution that cross-cultural studies of perception can make is that of determining the range of human perceptual functions. Determination of the true range is essential both for a full understanding of such functions and for establishing a universal taxonomy of such functions.

### See Also the Following Articles

Conformity across Cultures ■ Cross-Cultural Psychology, Overview ■ Industrial/Organizational Psychology across Cultures

### Further Reading

- Dereęowski, J. B. (1980). *Illusions, patterns, and pictures: A cross-cultural perspective*. London: Academic Press.
- Dereęowski, J. B. (1989). Real space and represented space: Cross-cultural perspectives. *Behavioral and Brain Sciences*, 12, 51–119.
- Levinson, C. S. (1998). Studying spatial conceptualization across cultures: Anthropology and cognitive science. *Ethos*, 26, 7–24.
- Reuning, H., & Wortley, W. (1973). Psychological studies of the Bushmen. *Psychologia Africana* (Monograph Suppl. No. 7).
- Rivers, W. H. R. (1901). Vision. In A. C. Haddon (Ed.), *Reports of the Cambridge anthropological expedition to Torres Straits* (pp. 1–132). Cambridge, UK: Cambridge University Press.
- Russell, P. A., Dereęowski, J. B., & Kinnear, P. R. (1997). Perception and aesthetics. In J. W. Berry, P. R. Dasen, & T. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology* Vol. 2: *Basic processes* (pp. 107–142). Boston: Allyn & Bacon.
- Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception*. Indianapolis, IN: Bobbs-Merrill.



# Performance Slumps in Sport: Prevention and Coping

J. Robert Grove

University of Western Australia, Perth, Australia

1. Slumps Versus Choking and Plateaus
  2. Causes of Performance Slumps
  3. Psychological Consequences of Slumps
  4. Factors Influencing the Reaction to Slumps
  5. Coping Behaviors in Response to Slumps
  6. Can Slumps Be Prevented?
  7. Increasing Awareness and Managing Risk
  8. Maintaining Confidence and Emotional Stability
  9. Performance Recovery Strategies
  10. Summary and Conclusions
- Further Reading

## GLOSSARY

**attribution** The perceived cause(s) of an event; these perceived causes have both emotional and motivational significance, so they can influence an athlete's behavior in response to performance difficulties.

**choking** Temporary unexpected decrements in performance that occur in situations where the perceived importance of good performance is high.

**coping responses** The specific cognitive and behavioral strategies that an individual uses to manage distress; these strategies are sometimes categorized in terms of their general approach or avoidance qualities.

**motivational climate** The psychosocial environment surrounding training and competition; rivalry, punishment, and recognition based on superiority to others are

generally viewed as elements of a counterproductive climate, whereas cooperative interaction and recognition based on improvement are generally viewed as elements of a productive climate.

**performance profiling** Systematic analysis of the essential ingredients for effective performance in a particular setting in combination with an assessment of strengths and weakness along these lines; this profiling can include physical, mental, and/or strategic aspects of performance and is useful for prioritizing skill development needs.

**plateau** Stabilization of performance at a given level, that is, a failure to improve.

**process goals** Specific and measurable objectives that relate to what one will do and how one will do it rather than the desired result or outcome; shooting 100 free throws at the end of practice every day for a month is an example of a process goal related to basketball.

**relapse prevention** A systematic attempt to reduce the probability that one occurrence of an undesirable behavior will lead to long-term debilitation or repeated occurrences; this framework emphasizes awareness of high-risk situations as well as the development of cognitive-behavioral strategies to maintain feelings of self-efficacy and control.

**self-efficacy** Situation-specific confidence in one's ability to accomplish something.

**self-handicapping** Generation of excuses or potential obstacles in advance of performance; although such actions may reduce the emotional impact of poor performances, they can also set in motion a self-fulfilling prophecy that makes good performance less likely.

Slumps are extended periods of below-average performance. The usual level of one's performance provides the frame of reference for defining a slump rather than some idealized or desired level of performance. The nature of one's sport involvement is also an important consideration. For a high-level athlete who trains daily and competes regularly, 2 or 3 weeks of below-average performance could constitute a slump. For a less committed performer with an irregular practice and competition schedule, a similar period of poor performance might not constitute a slump.

### 1. SLUMPS VERSUS CHOKING AND PLATEAUS

Individuals who participate in sport devote considerable time and effort to the development of physical skills, and they derive personal satisfaction from the public display of these skills. Consistency of physical performance is especially important for high-level amateur and professional athletes. Their scholarships or contracts may include performance-related contingency clauses, and their self-identity is often closely linked to physical prowess. Therefore, substandard performances are a source of concern for these individuals, and prolonged periods of poor performance can be highly stressful. Recreational athletes may also experience threats to their sense of self when, for no obvious reason, customary levels of performance are not attained for an extended period of time. The term "slump" is used to describe such a state of affairs.

Slumps can be distinguished from the related but qualitatively distinct phenomena of choking and plateaus. Choking refers to temporary unexpected decrements in performance that occur in situations where the perceived importance of good performance is high. Although choking can occur repeatedly, it is distinguishable from a performance slump because performance levels return to normal when the perceived importance of the event decreases. Factors associated with perceived importance include performance-contingent rewards or punishments and the ego relevance of the event. Performance plateaus are a longer term phenomenon than is choking, and the former are characterized by a failure to improve rather than by a decrement in performance. The athlete's performance improves to a point but then, despite diligent and conscientious training efforts, remains at a given level

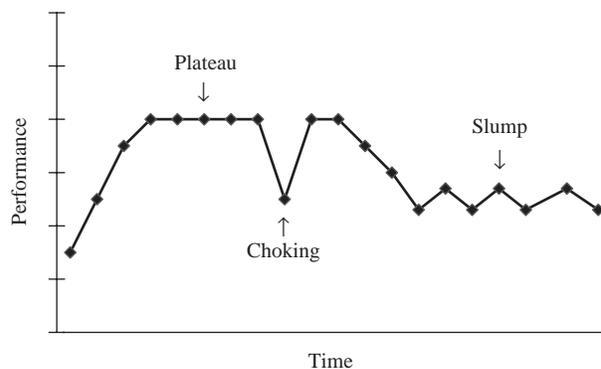


FIGURE 1 Illustration of the distinction among plateaus, choking, and slumps.

for an extended period of time. The distinction among plateaus, choking, and slumps is illustrated in Fig. 1.

### 2. CAUSES OF PERFORMANCE SLUMPS

The inability to pinpoint a specific cause for the extended period of poor performance is an important phenomenological aspect of performance slumps. Frequently (but not always), there is no apparent explanation for the performance deterioration, and this ambiguity creates frustration and a sense of helplessness for the athlete, who feels like he or she is working harder and harder but, despite this additional effort, is making little or no progress. However, from both theoretical and experiential perspectives, performance slumps are typically believed to originate from physical, psychological, behavioral, technical, and/or environmental causes. Table I summarizes these general categories of causal factors and provides examples of each of them. It is important to recognize that these categories are not mutually exclusive. In other words, performance slumps can have multiple causes, and it is entirely possible for factors within one category to have an influence on factors in another category.

Physical causes include lack of fitness, illness, injury, and/or general fatigue. With respect to fitness, it is important to recognize the distinctive purposes of general physiological fitness (e.g., strength, aerobic capacity) and of competition-specific fitness. Athletes sometimes have a tendency to place a heavy emphasis on general fitness and overlook the importance of match-specific fitness. This tendency is especially

TABLE I  
Possible Causes of Performance Slumps

Category	Subcategories	Examples
Physical factors	Fitness	Training not specific to game conditions
	Illness or injury	Lingering cold or flu symptoms
	Fatigue	Suboptimal work-to-rest ratio
Psychological factors	Confidence	Uncertainty about style of play or role
	Concentration	Poor control of distractions and/or emotions
	Expectancies	Unrealistic demands from self/others
Behavioral factors	Routines	Inconsistent preperformance activities
	Self-handicapping	Precompetition behaviors that create obstacles
	Time management	Overcommitment and/or poor organization
Technical factors	Technique/Equipment	Minor changes that disrupt "system integrity"
	Coaching	Too much and/or confusing instruction
	Knowledge	Poor awareness of on-field contingencies
Environmental factors	Motivational climate	Excessive focus on outcomes or superiority
	Facilities	Substandard training/playing conditions
	Nonsport stressors	Family, work, or school problems

strong early in the competitive season when there is a need to make a transition from general off-season programs to performance-specific programs. Fatigue and illness/injury, on the other hand, can become problematic as the season unfolds. The cultural milieu within competitive sport is typically one of "hard work pays dividends," so athletes often push themselves to the brink of exhaustion in an effort to improve performance. A failure to recognize that optimal performance requires a balance between work and rest can produce decrements in performance that persist for an extended period of time.

From a psychological perspective, the ability to play with absorption and certainty of purpose is often viewed as the cornerstone of performance excellence. If athletes do not have a clear idea of their on-field or on-court roles, their performances can suffer due to hesitancy ("lack of conviction"). The heavy emphasis on results in competitive sport also makes it easy for athletes to lose sight of their performance objectives and become overly concerned with outcomes. This loss of focus can be both subtle and debilitating, and it is especially likely to occur when significant others (e.g., coaches, teammates, family) emphasize achievements when interacting with the athletes. Persistent decrements in performance can also occur if athletes allow themselves to become distracted by emotional overreactions and/or the behavior of teammates, opponents, fans, or officials. Tendencies along these lines

are sometimes rationalized as "competitive intensity" and so can persist despite being counterproductive.

Consistent performance in stressful competitive circumstances requires a consistent approach to training and performing. Schedules, plans, and routines help to provide this consistency, and performance disruptions can occur if these elements are missing, underdeveloped, or changed. Seemingly innocuous changes, such as alterations in training times, can sometimes interfere with performance by producing a cascade effect whereby many other activities and commitments must be rearranged to fit the new schedules. Similarly, inadvertent departures from standard precompetition or pre-execution routines can sometimes lead to persistent performance decrements. Self-handicapping behaviors can also interfere with performance and can infiltrate the athlete's preparatory activities in a subtle and seemingly defensible manner. Common self-handicapping acts include paying increased attention to sport-specific issues that are unrelated to performance, training excessively hard close in time to competition, and being preoccupied with minor injuries or illnesses. Failure to manage one's time effectively can also undermine performance by creating a general sense of urgency about whatever the athlete is doing at the time.

Technical factors that can contribute to performance slumps include changes to technique and/or equipment. Although sometimes such changes are not obvious, they can disrupt the integrity of the performance

system by altering finely tuned timing mechanisms. Attempts to execute skills in an ideal or prescribed manner can have a similar disruptive influence on performance that might take considerable time to overcome. Coaches sometimes contribute to this problem by offering too much input, providing overly technical input, and/or by showing a lack of flexibility with respect to methods. A lack of awareness regarding when and how to make on-field adjustments to strategy can also contribute to extended periods of subpar performance despite the presence of objectively good technique. Potential causes of slumps within the environment include substandard training or competition facilities, stressors related to nonsport roles, and a motivational climate that places a heavy emphasis on beating others and winning rather than on skill development.

### **3. PSYCHOLOGICAL CONSEQUENCES OF SLUMPS**

At some level, athletes acknowledge performance slumps as an “occupational hazard.” At the same time, they admit to experiencing considerable distress when confronted with them. Prominent psychological reactions to slumps include feelings of frustration, anger, helplessness, self-doubt, and anxiety. Frustration is known to increase when there is proximity to a desired goal but attainment of the goal is blocked. For slumping athletes, the fact that they are accustomed to performing at a higher level serves as a continual reminder that the goal can be attained. On the other hand, their inability to reestablish these levels of performance despite diligent effort is a source of considerable frustration. This frustration is accompanied by increased emotionality, which may manifest itself in the form of impatience, irritability, and/or hostility. It may also manifest itself in feelings of helplessness, despair, and depression.

Loss of confidence, self-doubt, and anxiety are also experienced during a performance slump. It has been shown repeatedly that the strongest influence on situation-specific confidence (i.e., self-efficacy) is prior performance in similar situations. When the most recent prior performances have been consistently poor, confidence is eroded. At the same time, physical prowess is an important component of self-identity for most athletes, and there is no shortage of physically talented individuals waiting in the wings. Performing at less-

than-expected levels for an extended period can threaten the athlete’s sense of self and lead to pessimistic rumination on things such as “Do I still have what it takes?” and “Will I lose my spot [on the team, in the starting lineup, in the tournament, etc.]?” Such thoughts reflect self-doubt as well as anxiety about the future.

### **4. FACTORS INFLUENCING THE REACTION TO SLUMPS**

Research findings indicate that the amount of stress experienced in response to a slump is related more to how the slump is appraised than to individual difference factors. More specifically, self-reports from athletes in a variety of sports indicate that slump-related stress is slightly lower for those with spouses or partners than for those without spouses or partners but that it is unrelated to age, sex, education level, or years of competitive experience. On the other hand, interpreting the slump as a loss or a threat substantially increases the level of perceived stress. Slumps are also perceived as stressful when they are interpreted as a challenge, but not to the same degree as when loss or threat interpretations are involved. Dimensional properties of causal interpretations also show some relationship to slump-related stress. Specifically, a lack of perceived control over the cause of a slump tends to increase the level of self-reported stress, whereas a high level of perceived control over the cause tends to decrease stress levels. This finding is, of course, consistent with the way in which perceptions of control are viewed within general theories of behavior change as well as within models of stress and coping.

### **5. COPING BEHAVIORS IN RESPONSE TO SLUMPS**

Athletes tend to be concrete thinkers, and the sport environment is one that reinforces hard work and persistence. Therefore, it is not surprising to find that active, problem-focused coping strategies are cited most often when athletes are asked about their behavior in response to a slump. More specifically, findings from both qualitative and quantitative studies using various coping inventories indicate that athletes frequently employ increased effort or resolve, planned action, seeking of technical advice, seeking of social

support, and suppression of competing activities when they are confronted with an extended period of poor performance. Although emotion-focused strategies are used less often, they are sometimes employed. When that is the case, strategies related to conscious control of emotions, self-blame, and wishful thinking appear to be used most often. Avoidance-oriented strategies, such as denial and behavioral disengagement, are consistently the least frequently cited coping behaviors by athletes confronted with performance slumps.

Even though athletes have a general tendency to cope with performance slumps as outlined above, gender and personality influences are also evident in the literature. With respect to gender, it has been shown that female athletes seek emotional social support to a greater extent than do male athletes in response to a slump, but these differences tend to disappear when instrumental social support is examined (or when the two types of social support are considered together). Other studies have shown that male athletes report more use of the avoidance-oriented strategies of humor, denial, and behavioral disengagement than do female athletes. It must be emphasized that the overall frequency of citation for these strategies is low among athletes, but the

findings are interesting because they run counter to findings outside of sport that show a preference for active, approach-oriented coping strategies among males.

Personality factors that have been examined in connection with slump-related coping include dispositional optimism, trait sport confidence, perfectionism, self-consciousness, and achievement goal orientation. Table II summarizes the relationships that have emerged in this research. In general, athletes who are optimistic, possess high levels of domain-specific confidence, and adopt a mastery orientation to achievement tend to use active, problem-focused strategies in an effort to overcome performance difficulties. At the same time, these athletes are disinclined to rely on avoidance-based strategies, such as wishful thinking and behavioral disengagement, in response to a slump. Highly self-conscious athletes, on the other hand, have a tendency to fantasize, become preoccupied with their emotions, blame themselves for their difficulties, and/or disengage from the task. The impact of perfectionism on slump-related coping appears to depend on the particular subcomponent that predominates. More specifically, when perfectionist tendencies are driven by excessive concern about mistakes,

TABLE II  
Personality Correlates of Slump-Related Coping among Athletes

<i>Personality construct</i>	<i>Coping strategies</i>	<i>Nature of relationship</i>
Dispositional optimism	Effort/Resolve	Positive
	Planned action	Positive
	Seeking social support	Positive
	Wishful thinking	Negative
	Behavioral disengagement	Negative
Trait sport confidence	Deliberate emotional control	Positive
	Behavioral disengagement	Negative
Mastery achievement goals	Planned action	Positive
	Behavioral disengagement	Negative
Self-consciousness	Wishful thinking	Positive
	Self-blame	Positive
	Emotional focus/venting	Positive
	Behavioral disengagement	Positive
Perfectionism		
	Concern over mistakes	
Concern over mistakes	Denial	Positive
	Behavioral disengagement	Positive
High personal standards	Suppress competing activities	Positive
	Planned action	Positive
	Self-blame	Positive
	Emotional focus/venting	Positive

avoidance-based coping strategies are employed. On the other hand, when concern over mistakes is accompanied by a focus on high personal standards, a potentially more productive mix of problem-focused and emotion-focused strategies tends to be used.

in connection with performance slumps. Figure 2 outlines a proposed structure for using the relapse prevention model in this way. Specific strategies for increasing awareness, maintaining confidence and emotional control, and facilitating a return to form are discussed next.

### 6. CAN SLUMPS BE PREVENTED?

In general, the answer to this question is no. Individuals who participate in competitive sport on a regular basis will, in all likelihood, be confronted by extended periods of below-average performance from time to time. Acknowledging this possibility from the outset has at least two beneficial consequences. First, it reinforces the need for athletes to understand why slumps occur and to approach their sport involvement in ways that will minimize the frequency of the occurrence of slumps. Second, it suggests that athletes also need to learn about the typical reactions that occur during a slump and need to develop behavioral strategies for “getting back on track” as quickly as possible. To put it somewhat more bluntly, athletes are probably going to be confronted with performance slumps from time to time, so they had better learn how to deal with them productively. In the area of substance addiction, this is the same philosophy that underlies relapse prevention training, and there is merit in using a similar approach

### 7. INCREASING AWARENESS AND MANAGING RISK

The first step toward increased awareness and effective risk management is to become knowledgeable about behaviors and circumstances that increase the potential for slumps to occur. The list of factors in Table 1, although not exhaustive, is helpful in this regard because it highlights some of the more common features of high-risk situations. These “red flags” include depleted physical reserves, information overload, inconsistent training and competition routines, a chaotic and disorganized lifestyle, and the presence of unusually strong nonsport stressors related to work, family, or educational duties. Specific interventions that address these issues are outlined in Fig. 3 and include self-monitoring of physical and mental state, systematic observation of coach–athlete interactions, establishment of regular daily and weekly schedules, rehearsal of preperformance routines, and development of time management skills.

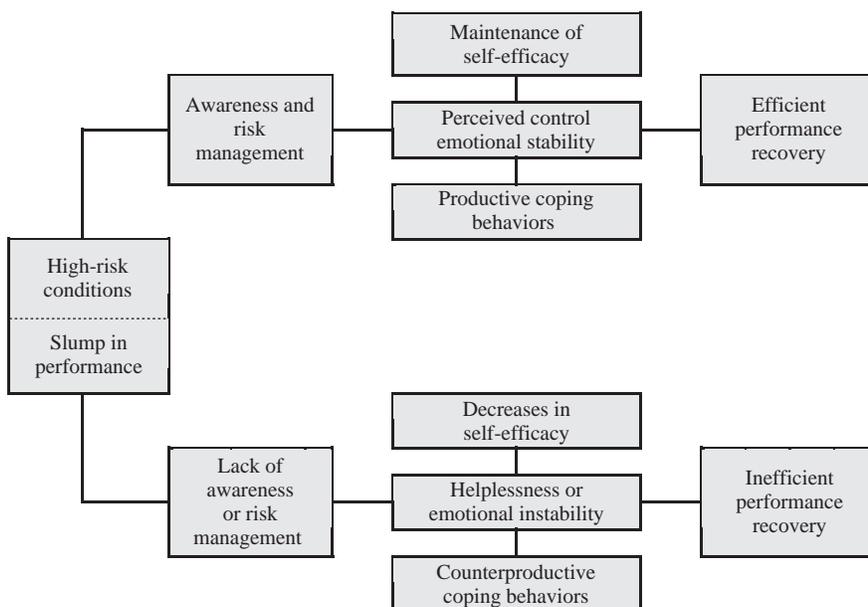


FIGURE 2 Application of the relapse prevention model to performance slumps in sport.

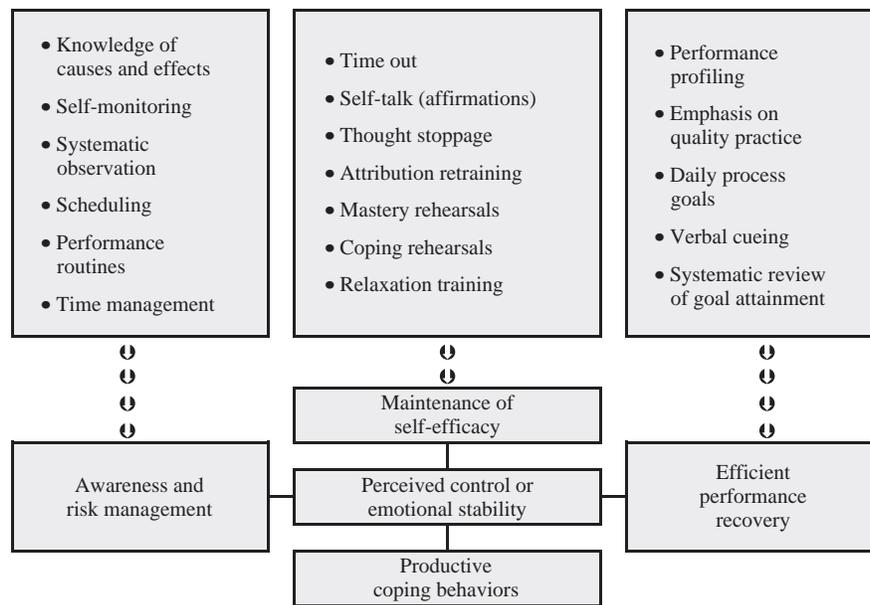


FIGURE 3 Intervention strategies for the management of performance slumps.

## 8. MAINTAINING CONFIDENCE AND EMOTIONAL STABILITY

When athletes are performing poorly, it is natural for them to lose confidence, feel somewhat helpless, and react with heightened emotionality. If not controlled, these negative thought patterns and emotional reactions can lead to a self-perpetuating cycle of counterproductive cognitive-behavioral reactions. To gain control over these reactions, it is sometimes necessary for athletes to remove themselves from the competitive environment for a period of time. This “time out” period, which might last from a few days to a few weeks depending on the level of involvement, gives them a chance to “clear their heads” and stabilize their emotions. There is certainly value in doing so, but it is sometimes difficult to convince committed sport performers that the best thing they can do for their game is to forget about sport for a while. An examination of Table II suggests that this might be especially problematic for highly optimistic/confident athletes with a strong mastery orientation.

Selected thought management strategies can be implemented to maintain feelings of self-efficacy and control as well as emotional stability. These strategies include affirmative self-talk, thought stoppage, and cognitive restructuring of attributions for the slump. Affirmations are simple, declarative, first-person

statements that reinforce positive personal qualities. As such, they are antithetical to the self-doubts that typically accompany performance difficulties. Examples of affirmations that might be relevant during a slump include “I deserve to be here,” “I’ve worked hard at my game,” “I can produce when I have to,” and “I enjoy showing what I can do.” A simple, three-part thought stoppage procedure can be used either by itself or in combination with affirmative self-talk to block negative thoughts and redirect attention in a constructive manner. This procedure consists of using covert verbal triggers (e.g., “Stop”) and/or physical triggers (e.g., snapping one’s fingers) to arrest negative thoughts. Attention is then redirected through a brief focus on the physical self (e.g., “Breathe . . .”) and/or the use of a cue word or phrase that captures a desired performance quality or orientation (e.g., “Let it rip,” “Play your game”). Affirmations can, of course, be used as the cue phrases in this procedure.

Attributions play an important role in the way in which athletes respond to performance difficulties both emotionally and motivationally. In general, it is emotionally comforting and motivationally productive to attribute poor performance to unstable specific factors (e.g., an unacceptably poor first serve percentage in tennis) rather than to stable global factors (e.g., a general inability to serve well). Doing so reduces the emotional impact of the negative outcomes and sets the

stage for continued striving toward improvement of the specific weakness. These explanations must be realistic, of course, because contrived excuse making will certainly not be advantageous. At the same time, consultants should be aware of the link between attributions and feelings of helplessness, and they should remain vigilant for potentially counterproductive attributional patterns when persistent performance difficulties arise.

Mental rehearsal and a variety of relaxation procedures may also be beneficial for maintaining confidence and emotional control. Two different types of mental rehearsal are relevant and can be used to one's advantage at different times. Mastery rehearsal involves visualization of effective performance and is good for maintaining self-confidence. It can be implemented by asking athletes to recall an excellent performance from the past and to reexperience it in as much detail as possible. After doing so, they can be asked to visualize an upcoming performance situation, see themselves performing in a similarly effective manner, and associate the prior feelings of success with the future performance. Coping rehearsal, on the other hand, involves visualization of potential difficulties that might be encountered in the future and the strategies that could be used to manage these difficulties effectively. This form of imagery aids in the recognition of behavioral alternatives and in the development of response flexibility. Relaxation exercises of various kinds can be used in isolation or in combination with mental rehearsal to alleviate some of the tension that typically accompanies a series of frustrating performances. Abdominal breathing exercises and abbreviated forms of progressive muscular relaxation are good choices for athletes because they focus attention on physical processes to produce a relaxation response. Sport performers are physically oriented people, and they are sometimes more accepting of physically oriented relaxation techniques than of cognitively oriented techniques.

## **9. PERFORMANCE RECOVERY STRATEGIES**

On-field performance is the bottom line for sport performers, so taking steps to increase the probability of efficient performance recovery is an important element of slump-related coping. A variety of behavioral techniques are relevant here. For example, performance

profiling can be used to isolate key elements of the physical and/or mental game as well as to identify strengths and weaknesses in these key areas. Skill development prioritization follows quite naturally from such an exercise, so the stage is set for quality practice (as opposed to "mindless" practice). Such an approach to training has been documented as a core element of sporting excellence, and there is considerable merit in modeling performance recovery strategies on these same principles. A central feature of quality practice is the incorporation of daily process goals into the training regime so that athletes have a clear idea of what they will do on any given day, how long they will do it, and how they will evaluate what they have done. Verbal cueing can be used along with daily goal setting to remind athletes of their priorities and to direct attention toward processes rather than outcomes. Cueing can be done covertly by athletes through self-talk, overtly by athletes through written reminders, or overtly by informed others through verbal statements. Structured reviews of the extent to which process goals were attained can then take place following training sessions and/or competitive performances.

## **10. SUMMARY AND CONCLUSIONS**

Slumps are an inherent and unavoidable phenomenon within physical performance settings. They are qualitatively distinct from the related phenomena of choking and plateaus, and they have predictable psychological consequences. These consequences include decreases in self-efficacy, feelings of helplessness and loss of control, and behavioral attempts to cope through approach or avoidance. Individual difference factors that have been shown to influence slump-related coping among athletes include gender and personality. Because it is likely that most sport performers will experience slumps at some points in time, it would be prudent for them to adopt a risk management perspective on the process. Such a perspective involves (a) increasing awareness of the potential causes of and common reactions to slumps, (b) becoming familiar with cognitive-behavioral strategies for the maintenance of confidence and emotional stability in the face of performance difficulties, and (c) adopting a process-focused approach to training and competition. Doing so will help to reduce the frequency of occurrence of slumps and will promote efficient restoration of typical performance levels when they do occur.

### See Also the Following Articles

Coping ■ Intrinsic and Extrinsic Motivation in Sport ■ Self-Confidence in Athletes

### Further Reading

- Baumeister, R. F., & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology, 16*, 361–383.
- Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130*, 701–725.
- Conroy, D. E., & Benjamin, L. S. (2001). Psychodynamics in sport performance enhancement consultation: Application of an interpersonal theory. *The Sport Psychologist, 15*, 103–117.
- Eklund, R. C., Grove, J. R., & Heard, N. P. (1998). The measurement of slump-related coping: Factorial validity of the COPE and modified COPE inventories. *Journal of Sport & Exercise Psychology, 20*, 157–175.
- Goldberg, A. S. (1998). *Sports slump busting*. Champaign, IL: Human Kinetics.
- Grove, J. R., Eklund, R. C., & Heard, N. P. (1997). Coping with performance slumps: Factor analysis of the Ways of Coping in Sport Scale. *Australian Journal of Science and Medicine in Sport, 29*, 99–105.
- Grove, J. R., & Heard, N. P. (1997). Optimism and sport confidence as correlates of slump-related coping among athletes. *The Sport Psychologist, 11*, 400–410.
- Grove, J. R., & Stoll, O. (1999). Performance slumps in sport: Personal resources and perceived stress. *Journal of Personal and Interpersonal Loss, 4*, 203–214.
- Klinger, E., Barta, S. G., & Glas, R. A. (1981). Thought content and gap time in basketball. *Cognitive Therapy & Research, 5*, 109–114.
- Madden, C. C., Kirkby, R. J., & McDonald, D. (1989). Coping styles of competitive middle distance runners. *International Journal of Sport Psychology, 20*, 287–296.
- Madden, C. C., Kirkby, R. J., McDonald, D., Summers, J. J., Brown, D. F., & King, N. J. (1995). Stressful situations in competitive basketball. *Australian Psychologist, 30*, 119–124.
- Maniar, S. D., Curry, L. A., Sommers-Flanagan, J., & Walsh, J. A. (2001). Student-athlete preferences in seeking help when confronted with sport performance problems. *The Sport Psychologist, 15*, 205–223.
- Prapavessis, H., & Grove, J. R. (1995). Ending batting slumps in baseball: A qualitative investigation. *Australian Journal of Science and Medicine in Sport, 27*, 14–19.
- Taylor, J. (1988). Slumpbusting: A systematic analysis of slumps in sports. (1988). *The Sport Psychologist, 2*, 39–48.



# Personal Initiative and Innovation

Fiona Patterson

City University, London, United Kingdom

1. Slumps Versus Choking and Plateaus
  2. Defining Innovation
  3. The Innovation Process
  4. Person-Level Influences
  5. Work Group Innovation
  6. Leadership
  7. Organizational Innovation
  8. Emerging Issues
  9. Summary
- Further Reading

## GLOSSARY

**creativity** The generation of new and original ideas.

**exnovation** When an organization actively discards existing practices associated with a previously implemented innovation, thereby allowing the adoption of a new innovation.

**incremental innovation** An innovation focusing on improvements of existing processes, products, or services to improve performance in an existing market.

**innovation** The process of generating and applying new ideas to benefit individuals, groups, organizations, or the wider society.

**leader-member exchange (LMX)** Where leaders develop relationships with all employees in their work groups; these can be high-quality or low-quality relationships in which a pattern of social exchange emerges.

**path analysis** A set of statistical procedures that allows the researcher to test hypothesized causal relationships among variables (sometimes called causal modeling).

**personal initiative** A class of active behaviors with three facets: self-starting, proactivity, and persistence.

**radical innovation** An innovation focusing on the development of processes, products, and services with unprecedented performance potential to create a dramatic change in existing markets.

**role innovation** Changes in task objectives, responsibilities, methods, materials, and interpersonal relationships.

Innovation is essential for organizational prosperity. It is an iterative process that involves idea generation and idea implementation. There are a variety of influences on the propensity for innovation at work that can be classified into three levels of analysis: individual, work group, and organizational.

## 1. INTRODUCTION

Over the past two decades, there has been an explosion of interest in innovation as organizations have recognized that creating new processes, products, and procedures is vital for productivity and growth in all sectors. With more dispersed and virtual working, role innovation is essential because clearly defined job descriptions no longer exist for many job roles. The research literature on creativity and innovation is immense, both in magnitude and in diversity. Until recently, there has been confusion in terminology regarding these two concepts, particularly in definitions and in criteria for assessment. Many have observed the lack of a cohesive theoretical

understanding of how creativity and innovation operate in organizations.

First, the concepts of creativity and innovation are defined. The various factors that influence innovation in organizations can be classified into three main levels of analysis: individual, work group, and organizational. Each of these levels of analysis is reviewed. Finally, a summary of the emerging issues is presented along with some implications for future research and practice.

## 2. DEFINING INNOVATION

The concepts of creativity and innovation are confused repeatedly, and these terms are often used interchangeably in the literature. Creativity and innovation are overlapping constructs, but the main distinction is with regard to novelty. Creativity is exclusively concerned with generating new and entirely original ideas. Innovation is a broader concept because it also encompasses the application and implementation of new ideas to produce something new and useful (in the context of groups, organizations, or societies). Innovation is often referred to as a process because implementing new ideas necessarily involves influencing others, whereas creativity could be achieved in isolation. Furthermore, an innovation could be the application of something familiar in one organization to another unfamiliar organization.

The most widely accepted definition of innovation is that of West and Farr, emphasizing the positive nature of innovation: "the intentional introduction and application within a role, group, or organization of ideas, processes, products, or procedures, new to the relevant unit of adoption, designed to significantly benefit the individual, the group, the organization, or wider society."

Nearly all authors recognize that innovation in organizations is a complex iterative process. Numerous approaches and process models have been proposed, but two main stages are common to all of them: a suggestion phase and an implementation phase. Innovation is not a linear process given that it involves several cycles of activities such as initiation, reappraisal, adaptation, implementation, and stabilization. In addition to person-level resources, there are various environmental factors that influence this process such as feedback, leadership style, resource availability, and organizational climate. [Table I](#) summarizes the factors that have been shown to influence innovation in orga-

nizations at the three levels of analysis: individual, work group, and organizational.

## 3. THE INNOVATION PROCESS

Early research involved stage model approaches to innovation, where authors attempted to identify the sequence of activities during the innovation process. Zaltman and colleagues' stage model of the innovation process in 1973 is perhaps the most widely cited. The model suggests two main phases, initiation and implementation, with various substages in each. The initiation stage involves the substages of knowledge awareness, formation of attitudes, and the decision to adopt or abandon. During the implementation stage, ideas are tested and eventually the innovation is sustained and routinized. Although this research emphasizes that the innovation process is multifaceted, King and Anderson criticized stage-based approaches because they often lack observable supporting evidence. Innovation processes may share some common features, but these are not easily identifiable as discrete stages occurring in a linear fashion.

During the 1990s, Van de Ven and colleagues studied the innovation process in-depth, using empirically grounded observations. Their studies comprised the Minnesota Innovation Research Program (MIRP), captured in detail in *The Innovation Journey* in 1999. This program encompassed several organization-based longitudinal studies of a diverse collection of innovations observed in real time. The research concluded that the innovation process is best characterized as a "nonlinear dynamic system." In this way, the system involves a complex cycle of divergent and convergent activities that may be repeated over time at various organizational levels and that are shaped by forces such as resource supply, leadership behaviors, and the organizational culture. Although the research concluded that innovation is a nonlinear process, by comparing case histories of various innovations, the MIRP studies identified patterns of commonality in innovation processes. These are the initiation, developmental, and implementation periods of the innovation. Each period is described in [Table II](#).

Research demonstrates that different types of innovation may have varying antecedent factors. Furthermore, the evidence suggests that not all of the elements of the innovation process are the same across different innovations. For example, in exploring differences in incremental and radical innovation processes

**TABLE I**  
**Factors That Influence Innovation at Work**

<i>Level of analysis</i>	<i>Characteristic</i>	<i>Dimension</i>
Individual	Personality	Openness to experience, low conscientiousness, low agreeableness, originality, imaginative, self-confidence, tolerance of ambiguity, nonconforming, high energy, social rule independent, personal initiative (e.g., self-starting, proactive)
	Motivation	Intrinsic (vs extrinsic), personal initiative (e.g., persistent), determination to succeed, self-efficacy
	Cognitive ability	Above average intellect, creative cognitive style, ideational fluency
	Knowledge	Domain-specific knowledge
Work group	Job characteristics	Autonomy, control, support for innovation, job demands, job dissatisfaction
	Team dynamics and structure	Diversity, conflict, constructive controversy, minority influence, organic structure
	Team climate	Participation, shared vision, support for innovation (articulated and enacted), task orientation
	Team processes	Reflexivity, minority dissent, integration skills, decision-making style
Organizational	Leadership style	Transformational, participative, democratic style, delegating responsibility, autonomy, openness to idea proposals and challenges, constructive feedback, expected evaluation, leader-member exchange
	Structure	Centralization, formalization, complexity
	Climate and culture	Support for risk taking, tolerance of failure, organizational values, reward and recognition for innovation, participative management and decision making, flexible employment contracts
	Size	Number of employees, market share, alliances
	Resources	Economies of scale, slack resources
	Work processes	Collaborative idea flow across functions, idea champions, change agents, suggestion schemes
	Physical environment	Technology to support innovation, informal meeting places, stimulating physical environment

Source. Anderson, De Dreu, and Nijstad (2004).

Note. Factors cited represent a summary of variables, but this is not an exhaustive list. Some factors exert a positive influence, whereas others a negative influence. Some factors have a curvilinear association with innovation, whereas others have an indirect association with innovation.

in 1992, King showed that sequential linear models were most applicable for incremental innovations, whereas radical (revolutionary) innovations were characterized more by convoluted iterative patterns of development.

Innovation processes do not occur in a vacuum and can be shaped by current organizational practices, particularly during the implementation phase. To illustrate this perspective, King and Anderson highlighted the concept of “exnovation.” This concept was first introduced in 1981 by Kimberly, who described innovation as a series of processes that combine to define an “innovation life cycle.” The final process of the model is exnovation, where the organization must discard existing practice associated with a previously

implemented innovation, thereby allowing the adoption of a new innovation, where the life cycle starts again.

#### 4. PERSON-LEVEL INFLUENCES

Innovation involves multiple components at the individual level. However, there has been little synthesis of the literature to build a framework within which to explore the many interrelated characteristics involved. During the past decade or so, there has been a shift in research toward more integrative approaches. For example, in 1999, Sternberg and Lubart proposed an “investment theory” suggesting that the propensity to innovate requires six distinct resources: intellectual

**TABLE II**  
**Common Elements in the Innovation Process Identified in the Minnesota Innovation Research Program**

<i>Period</i>	<i>Observation</i>	<i>Examples</i>
<i>Initiation</i>		
Gestation	Initial events might not be intentionally directed at innovation, and gestation might last several years. Coincidental events can trigger a need for change.	Art Fry sang in a church choir for years. Art, like many choir members, put slips of paper in his hymn book to mark each selection. However, his technique was not foolproof, the slips of paper often fell out. Taking his dissatisfaction back to his job, he eventually developed what became 3M's Post-it note pads.
Triggering shocks	Efforts to initiate innovations can be triggered by shocks, whether driven internally or externally to the organization. Without shocks, new ideas might not be acted on.	While on vacation, Edwin Land took pictures of his daughter. When she showed her disappointment that she could not see the results immediately, he set his mind to the task of developing instant photography. In an organization, a shock could be new leadership, a recent product failure, or a change in legislation.
Plans, budgets and exposure	Realistic plans are developed to gain resource commitment (time and finance) to the innovation.	Cost-benefit plans are needed to convince resource controllers and venture capitalists that the risk is worth it. In the organizations studied by the MIRP, during this period, entrepreneurs acted to deflect attention away from the innovations' uncertainty and toward their potential value.
<i>Developmental</i>		
Proliferation	Initial ideas proliferate into diverse multiple pathways, making the process difficult to manage.	In the MIRP studies, most innovations did not consist of a single new product or procedure. Instead, families of related new products were developed in multiple pathways to reduce the risk of failure.
Setbacks	Setbacks can occur frequently because initial plans go wrong or there are unforeseen environmental events.	This period is characterized by budget overruns, unmet deadlines, unsuccessful product launches. Organizations may respond by adjusting plans and resources.
Shifting performance criteria	Criteria for innovation success and failure often change over time, reflecting differing stakeholder views.	Long-term goals could be expressed as return on investment, whereas short-term criteria may be obtaining resources, recruiting appropriate personnel, acquiring equipment, and so on.
Fluid participation of personnel	Personnel often do not work full-time on the innovation. In addition, specific expertise may be required.	There may be high turnover of personnel, or relevant technical expertise might not be available. Part-time involvement means that commitment may vary among personnel.
Investors/Top management involvement	Senior management and/or investors intervene in the process.	The roles of senior management may vary and might not have a unified perspective. Interventions could be to remove impediments or possibly to expand the innovation.
Relationships frequently change	Relationships shape the process, with many interest groups engaging in various transactions to move the innovation forward.	Relationships could be internal to the organization or between external parties through partnerships and joint ventures. With a complex network of interest groups involved, reaching a consensus to move innovations forward might be slow.
Community/ Industry infrastructure	Innovation stakeholders often work within an industry or community infrastructure that is needed to implement an innovation.	The infrastructure could include institutional norms or institutional standards. Organizations seek to establish a distinctive position through innovations but simultaneously must operate within an industrial infrastructure or community.

*Continues*

Continued

*Implementation/Termination*

Integrating the old and the new	The new is linked and integrated with the old rather than substituting or replacing what has gone before.	Because of limited resources and for political reasons, innovations are seldom substituted for existing organizational programs. Instead, the aim is to integrate existing arrangements.
When innovations stop	Innovations terminate when they are institutionalized or when resources expire.	Innovations become routinized and diffused into adopted practice. Reinvention may also occur here when adopters may modify innovations to fit their local circumstances.
Attributions are made	Stakeholders explain causes for the successes or failures of the innovation.	Whether correct or not, senior managers and investors make attributions for success or failure that can significantly affect the career development of the innovation champions and their future likelihood of gaining resources for innovations.

Source. Van de Ven, Polley, Garud, and Venkataraman (1999).  
 Note. MIRP, Minnesota Innovation Research Program.

abilities, knowledge, styles of thinking, personality, motivation, and environment. This reflects the key areas of research at the person level where previous literature can be classified into that concerning associations between innovation and (1) intelligence, (2) knowledge, (3) personality, (4) motivation, and (5) personal initiative.

**4.1. Intelligence**

Numerous researchers have explored the association between innovation potential and intelligence. However, the findings are generally inconclusive. Much of the literature in this area can be classified into four categories conceiving of innovation as (1) a subset of general intelligence, (2) an aspect of genius, (3) a set of cognitive abilities and mental processes, and (4) associated with observer judgments of intelligence.

**4.1.1. General Intelligence**

Early research claimed that creativity was equivalent to high intelligence. The best-known researcher in this field is Guilford. In his theory of the structure of intellect published during the 1950s, he claimed that creative thinking was a mental ability involving divergent production as “thinking that goes off in different directions.” Many researchers followed Guilford’s work by producing evidence that ideational fluency (i.e., quantity of new ideas) underlies divergent thinking test scores. However, review studies show that such scores

often failed to correlate significantly with various indexes of innovation. Other investigations have tested the possibility of a curvilinear relationship between intelligence and innovation where intelligence would potentially become less influential as the level of intelligence increases beyond a certain point. However, tests have proven to be inconclusive, and some authors doubt whether divergent thinking tests measure abilities actually involved in creative thinking.

**4.1.2. Aspect of Genius**

Some have suggested that genius, as the most obvious manifestation of high intelligence, is closely tied to the propensity for innovation. However, there has been a lack of evidence to support a direct relationship between innovation and intelligence. Many have concluded that intelligence is a necessary, but not sufficient, condition for innovation. Recent studies conclude that intelligence and innovation potential are moderately related but that once intelligence quotient (IQ) scores surpass 115, the relationship is near zero. This finding has been described as “threshold theory,” where intelligence and innovation potential, instead of being twin or even sibling constructs, may be more like cousins.

**4.1.3. Cognitive Abilities**

In 1992, Finke and colleagues suggested that to understand the role of cognitive abilities in idea generation, one must draw on current models in cognitive

psychology and use experimentally based observations of the processes that underlie generative tasks. Their work followed a framework called the “geneplore model.” This model proposes that many creative activities can be described in terms of an initial generation of ideas or solutions followed by an extensive exploration of those ideas. Initial ideas are referred to as “preinventive” in the sense that they are incomplete solutions but offer promise in terms of originality and utility. The model assumes that one would alternate between generative and exploratory phases, refining the structures according to the demands or constraints of the specific task. This “creative cognition” approach emphasizes that generative capacity is a property of normative human cognition. Individual differences occur due to variations in the use and application of these generative processes together with the sophistication of an individual’s memory and knowledge in the relevant domain. In simple terms, the capacity for creative cognition is normally distributed. Highly creative people do not have minds that operate in any fundamentally different way from those of other individuals.

#### **4.1.4. Observer Judgments of Intelligence**

Innovative individuals are often perceived and rated by others as more intelligent than less innovative individuals. For example, in MacKinnon’s studies of architects during the 1960s, supervisors rated innovative architects as more “intelligent” than less innovative individuals. MacKinnon described the innovative architects as having high “effective intelligence”, and argued that traditional measures of intelligence (e.g., IQ) do not fully explain this “real-world” intelligence. In 2003, Feist and Barron showed that observer-rated intelligence at 27 years of age predicted lifetime innovation at 72 years of age. Similarly, tested intelligence had much weaker relationships with innovation over this time.

Historically, the literature on innovation and intelligence has lacked clarity. Part of the problem has been that intelligence, similar to innovation, is often viewed as a unitary concept. Previous theories of intelligence have tended to overemphasize cognitive abilities and downplay the role of knowledge-based intelligence.

### **4.2. Knowledge**

Nearly all researchers in this field have assumed that knowledge is a key variable in both generative thinking and innovation. Immersion in domain-specific

knowledge is an essential prerequisite for innovation because an individual must have an accurate sense of domain (i.e., contextual relevance) before he or she can hope to change it for the better. However, the literature highlights that too much expertise in one area can also be a block to innovation within that domain. There is an “inverted U” relationship between knowledge and innovation, where too much or too little knowledge will not lead to new inventions. During the 1980s, Simonton studied the lives of more than 300 eminent people to explore life span development of innovation and found that both a lack of familiarity and an excess of familiarity within a subject domain can be detrimental to innovation. Domain knowledge, like intelligence, is necessary but not sufficient for innovation to occur.

### **4.3. Personality**

From several decades of research on the association between innovation and personality, a consistent set of characteristics has emerged. These include imaginative, inquisitive, high energy, high desire for autonomy, social rule independence, and high self-confidence. The Five Factor Model (FFM) of personality has become a nearly universal template with which to understand the structure of personality. The FFM dimensions are Openness to Experience (e.g., ideas, aesthetics), Agreeableness (e.g., compliance, straightforwardness), Conscientiousness (e.g., order, dutifulness, competence), Extraversion (e.g., warmth, gregariousness, activity), and Neuroticism (e.g., anxiety, depression). Given that the FFM is an appropriate model for charting individual differences among adult populations, it provides a useful structure to review the literature exploring associations between personality and innovation.

#### **4.3.1. Openness to Experience**

There is good empirical evidence of a positive association between various characteristics associated with innovation and those used to depict openness (e.g., imaginative, original, flexible, unconventional). Research suggests that openness enhances an individual’s intrinsic motivation toward novelty and, therefore, works in a multiplicative way to produce innovation. Although there are some inconsistencies in the findings, openness is perhaps the most important personality dimension to predict the propensity for innovation.

### 4.3.2. Agreeableness

Several studies have demonstrated a negative association between agreeableness and innovation. Innovators have high social rule independence. During the 1990s, Eysenck emphasized the potentially negative characteristics of innovators, where innovators are often outspoken, uninhibited, quarrelsome, and/or asocial.

### 4.3.3. Conscientiousness

Research has demonstrated that lack of conscientiousness is associated with innovation. Defined by terms such as fastidious, ordered, neat, and methodical, the evidence shows that individuals high on conscientiousness are more resistant to changes at work and are more likely to comply with current organizational norms. Recent studies have shown that the facets of conscientiousness most closely associated with innovation are being methodical, being ordered, and being dutiful.

### 4.3.4. Extraversion

Meta-analytic studies (i.e., large-scale reviews of quantitative studies) suggest that, in general, introversion is positively associated with innovation. Similarly, many have argued that isolation and withdrawal are necessary conditions for generating new ideas. However, there is little evidence from organizational contexts, and some have suggested that extraversion is a positive predictor of innovation. In meta-analytic studies of occupational work performance in general, extraversion has been shown to be a positive predictor for many occupations. This is particularly the case in large organizations where interpersonal factors are likely to be important for effective job performance (e.g., sales, managers, other professional occupations). Further research is needed to explore the association between extraversion and innovation.

### 4.3.5. Neuroticism

There is relatively little research examining an association between neuroticism (i.e., low emotional stability) and innovation. Of what literature is available, there appear to be some inconsistencies depending on the domain of interest. For example, in a meta-analytic review, Feist observed that artists appear to be more anxious, emotionally labile, and impulsive than the scientists he studied. A more thorough investigation in this area is necessary, particularly in organizational

settings with a broader range of occupations. Some have suggested a curvilinear association between emotional stability and performance, where too much or too little anxiety is detrimental to innovation and where moderate levels of anxiety can enhance innovative potential.

There are various personality measures available that claim to assess the propensity to innovate, including Cattell and Eber's Sixteen Personality Factor (16-PF) Questionnaire and Patterson's Innovation Potential Indicator. Other approaches include cognitive-style instruments such as Kirton's Adaptation-Innovation Inventory.

## 4.4. Motivation

High levels of motivation are required for innovation, and innovators are viewed as displaying a devotion and total absorption in their work. During the 1980s, Amabile suggested a componential model of innovation that involves three components: intrinsic task motivation, domain-relevant skills (i.e., expertise), and innovation-relevant process skills (i.e., cognitive skills and work styles conducive to novelty). The model includes a five-stage description of the innovation process—task presentation, preparation, idea generation, idea validation, and outcome assessment—where the roles of the three components vary at each of the stages. Amabile's model suggests how and where individual skills and motivation affect the progress of the innovation process.

Intrinsic motivation is clearly a prerequisite for innovation, but the role of extrinsic motivators is less clear. In exploring environmental influences on motivation, the evidence suggests that constructive evaluation (e.g., informative, supportive, recognizes accomplishment) can enhance innovation. Amabile suggested that intrinsic motivation is particularly important in tasks that require novelty and that extrinsic motivators may be a distraction during the early stages of the innovation process. Later during the innovation process, when persistence and evaluation of ideas are required, synergistic extrinsic motivators may help innovators to persist in solving the problem within the domain. Any extrinsic motivator that enhances an employee's sense of competence without undermining self-determination should enhance intrinsic motivation and, thus, increase the propensity for innovation.

In related studies, results showed that individuals who received positive feedback given in an informational style, and who worked in a high-task autonomy environment, generated the most innovative solutions. Perceived competence (i.e., an individual's belief that

he or she is capable of doing the task) and self-determination (i.e., an individual's perception that he or she is the main influence behind successful performance of the task) are the two main antecedents of intrinsic motivation. When autonomy is high (in choosing work methods and work schedules) and when an individual receives positive informational feedback, this is likely to optimize innovative production.

#### **4.5. Personal Initiative**

Based on work by Frese and colleagues, the concept of personal initiative (PI) describes a class of behaviors that have been positively linked with innovation and entrepreneurial orientation. PI is defined by three main facets: self-starting, proactivity, and persistence. A self-starting approach is characterized by setting oneself context-specific goals and going beyond formal job requirements. Proactivity implies that one anticipates opportunities and problems rather than reacts and prepares to deal with them before they occur. PI also involves persistence because this will be required to overcome barriers in reaching one's goals. Frese suggested that these three facets of PI reinforce each other and tend to co-occur. Taking an action-based approach, each of the three facets of PI can be used to understand how individuals develop goals, collect information, and make plans for executing them as well as how individuals gather and use feedback.

In 2001, Frese and Fay proposed a complex model of distal and proximal factors that influence performance in organizations, whereby environmental supports (e.g., support for PI, control at work), knowledge, skills and abilities (e.g., cognitive ability), and personality influence individual orientations (e.g., self-efficacy, handling errors, active coping), which in turn influence PI. In this way, PI is conceptualized as a set of active behaviors and is directly linked to effective performance in organizations.

There is a growing body of research examining the association between PI and innovation. The relationship between PI and conservatism in the workplace has been studied. Results show that conservatives showed significantly less PI, were less orientated toward growth and challenge, and were less innovative compared with their less conservative counterparts. In exploring environmental influences on PI, research shows that increases in control and job complexity can help to enhance PI at work.

PI is particularly important during the idea implementation phase of the innovation process because it

involves overcoming barriers and persistence. In a study using path analysis, several predictors were examined, including individual-level variables (e.g., PI, self-efficacy, interest in innovation), work characteristics (e.g., control, complexity), motives (e.g., reward), and organizational system factors (e.g., supervisor support). Results showed that being proactive, being actively involved in one's work environment, and being confident that one is capable of thinking of good ideas were the most important predictors of innovation. Organizational factors were also important variables, and the results suggest that innovation is maximized when organizational climates promote an active approach toward work and interpersonal risk taking.

The concept of PI has been applied at an organizational level. Findings suggest that companies with pro-initiative climates are more innovative and profitable. Climates for initiative may improve organizations' ability to deal with innovation and change by encouraging self-starting, proactivity, and persistence in employees.

### **5. WORK GROUP INNOVATION**

Until recently, there had been little literature available regarding innovation in work groups. Several factors enhance group innovation, and four key issues are considered: (1) operating principles in groups; (2) diversity, dissent, and minority group influence; (3) group development over time; and (4) group climate.

#### **5.1. Operating Principles**

In 2002, King and Anderson highlighted the differences between "organic" and "mechanistic" group structures in their discussion of operating principles. An organic group structure is characterized by loose boundaries of authority and responsibility and by a propensity to work as a group rather than breaking projects down into discrete tasks for individuals. Conversely, mechanistic groups are typified by being rule bound, hierarchical, and formal in operation. Organic group structures tend to be more innovative because autonomy and freedom are enhanced. Work group autonomy is an essential ingredient of innovation, and it is most effective when combined with unambiguous goals and objectives from management. Autonomy is important for idea generation, and mechanistic forms of organizing may have a role in coordinating the implementation of ideas. Research

suggests that “semi-autonomous” operating groups are optimal for innovation.

During the 1990s, West and colleagues described the process of “team reflexivity” as the extent to which team members collectively reflect on the team’s objectives, strategies, and processes, as well as the wider organization context, and adapt them accordingly. When teams regularly devote time to collective consideration of their objectives, plans, and actions, they are in a position to develop continuously and potentially to improve performance. Reflexivity enhances team learning and allows teams to refine shared goals, plans, and actions, all of which are likely to enhance the likelihood of innovation.

### 5.2. Diversity, Dissent, and Minority Group Influence

Groups consisting of people with a wide variety of backgrounds and perspectives are more likely to consider a wider variety of approaches to tasks. However, too much diversity will increase the likelihood of conflict within the group, and this could have negative consequences for productivity. During the 1980s, Tjosvold introduced the concept of “constructive controversy,” which indicates the value of social interaction, and specifically controversy, in decision making. Constructive controversy occurs when team members believe that they are in a cooperative environment (emphasizing mutually beneficial goals rather than a “win–lose” solution), when they believe that their personal competence is acknowledged, and when members use processes of mutual influence rather than attempted dominance. Accepting and encouraging the expression of minority views and dissent in groups and organizations is an important stimulus for creativity and innovation. Conversion to a minority view in groups is most likely to occur when the minority is consistent and confident in presenting arguments. A behavioral style of persistence is most likely to promote attitude change and influence over the majority. The minority group must reflect a visible commitment to the norms, values, and interests of the majority to gain influence.

### 5.3. Group Development over Time

In general, the longer groups are together, the less innovative they become as teams grow to be habit bound and more resistant to change. Work groups do not come together and develop over time in a vacuum. Initiating change in an organization might not be possible if this occurs at a peak time in an organization’s annual cycle of

activity. The concept of “entrainment” has been used to explain that cycles of activities are paced by the numerous other cycles. In applying this to organizations, this “entraining process” provides windows (or barriers) of opportunity for change to take shape and take root. In addition to longevity, the timing of the formation of a project group can be critical to its success.

### 5.4. Group Climate

During the 1990s, West and Anderson proposed a four-factor model of team climate. This model suggests that group innovation is related to (a) participative safety (members feel psychologically safe in proposing new and improved ways of doing things, and all participate in decision making), (b) vision (the team’s goals and objectives are clearly defined, shared, attainable, and valued), (c) support for innovation (there is expectation, approval, and practical support toward group members’ attempts to introduce new and improved ways of doing things in the work environment), and (d) task orientation (the team is committed to achieving the highest possible standard of task performance using constructive progress-monitoring procedures).

There is good evidence to support the existence of these four factors in relation to group climate. Groups with clearly defined and shared visions are more likely to develop new working methods because their efforts are focused and have direction. Participation in decision making is important because it increases the likelihood that group members commit to decision outcomes and will be willing to offer new ideas. In addition, innovation requires group commitment to achieve high task performance and requires members to offer articulated and enacted support for innovation attempts.

## 6. LEADERSHIP

Since the early 1980s, several independent studies have identified a range of leadership behaviors that enhance employee innovation. These behaviors include encouragement of risk taking, an open style of communication, a participative and collaborative style, provision of autonomy and freedom, support for innovation (verbal and enacted), constructive feedback, and optimism about the future. Most research has focused on one or two management behaviors resulting in a lack of integration of results. Furthermore, because many studies have focused solely on either idea generation or the implementation of ideas, clarity has been difficult to achieve.

King and Anderson proposed a contingency model to understand the role of leadership in relation to four phases of the innovation process: (a) initiation (the leader encourages ideas and is nurturing, supportive, open-minded, and nonjudgmental), (b) discussion (the leader obtains opinions, evaluates, and agrees on plans), (c) implementation (the leader sells the solution to all stakeholders, designs plans to include stakeholders, and gains commitment to plans), and (d) routinization (the leader checks effectiveness, modifies, and refines). Therefore, leaders must be flexible in their approaches and must employ different styles according to the different phases of the innovation process.

The quality of the leader–member exchange (LMX) predicts the propensity for innovation. Leaders develop relationships with all members in their work groups, and these can be characterized as high-quality or low-quality relationships, where a pattern of social exchange emerges. In 1999, Tierney and colleagues demonstrated that a high-quality LMX predicts employee innovation because employees who were engaged in more challenging tasks received more resources and rewards for innovation. In recent studies, job self-efficacy has been shown to be the strongest predictor of creative self-efficacy. This implies that if managers are to enhance employee innovation, they must provide the necessary training and feedback to enhance employee confidence and mastery. Recent theories of leadership describe two styles: a transformational style (inspiring and motivating others to achieve a shared vision) and a transactional style (influencing others through contingent rewards and punishments). Research demonstrates that a transformational leadership style enhances employee and group innovation.

## **7. ORGANIZATIONAL INNOVATION**

Several factors facilitate or inhibit innovation at the organizational level of analysis: (1) climate and culture, (2) structure and size, (3) resources and work processes, and (4) external demands and environment.

### **7.1. Climate and Culture**

Climate and culture are important antecedents to innovation to the extent that risk taking and idea generation are encouraged. Climate is the shared perceptions and attitudes that influence the social interaction behavior.

Culture is a set of commonly shared beliefs, values, and assumptions that form the basis for climate behavior. Supportive and challenging climates encourage innovation, whereas environments characterized by distrust, personal hostilities, limited autonomy, and unclear work goals inhibit the implementation of ideas. Organizational support for innovation is characterized by support for members pursuing new ideas and encouragement of innovation through both words and deeds. The espoused expectations and organizational values in relation to innovation (e.g., openness to change, willingness to experiment with ideas) are likely to have a major influence on employee behavior.

Amabile and colleagues noted that organizational encouragement is a vital aspect of a work environment for innovation and includes (a) encouragement of risk taking and a valuing of innovation from the highest level, (b) fair and supportive evaluation of ideas, (c) reward and recognition for innovation, (d) collaborative idea flow across the organization, and (d) participative management and decision making.

Research shows that an organizational climate can build a perceived need for innovation. Setting innovation goals at both the team and individual levels is one way in which to signal that innovation is valued and innovation can be incorporated in performance management systems. In 2001, Csikszentmihalyi argued that innovation must be considered within a complex system composed of the domain, field, and person. Thus, organizations have gatekeepers who can make decisions about which innovations are worth supporting without being too accepting or too conservative with regard to novelty. Organizational practices that positively influence innovation include flexible employment contracts and knowledge management systems, whereas practices that seek to control employees lower innovation supportive cultures.

### **7.2. Structure and Size**

There are three main aspects regarding the influence of organizational structure on innovation: centralization, formalization, and complexity. Centralization refers to the extent to which authority and decision making lie at the top of a hierarchy. Centralization is believed to hinder innovation because it restricts information flow and communication. Decentralization also gives rise to greater participation by allowing more viewpoints to be considered during idea generation. However, decentralization has a negative impact on the adoption of innovation and idea implementation in that

decentralization is necessary for idea generation, whereas centralization is important for idea implementation.

Formalization is the degree of emphasis placed on rules and procedures. High formalization inhibits innovation because it prevents organizational decision makers from seeking new sources of information and from making changes at work. However, procedural guidelines and regulations can enhance the successful implementation of new ideas.

Complexity is the degree of occupational specialization in the organization and is believed to enhance idea generation and inhibit idea implementation. High complexity is likely to provide a diverse range of information sources during idea generation, but such diversity could accelerate the potential for conflict during the implementation phase.

Because innovation is a multifaceted process, some variables that are important to enhance idea generation may be counter to promoting idea implementation.

Some claim that increased organizational size enhances innovation because there is greater access to resources through economies of scale. However, organizational size does not directly predict innovation because it is related to several different variables such as resource availability (which promotes innovation) and integration of communication and knowledge sharing (work processes often become more fragmented as size increases and so innovation is hindered).

### 7.3. Resources and Work Processes

Availability of resources is positively related to innovation because resources are needed to develop ideas. Beyond a certain threshold, more available resources do not significantly increase the likelihood of innovation. Researchers have observed a concept of “escalation”; that is, there may come a point when there is so much invested that it is too costly to stop the implementation process.

Integration across functions and groups promotes innovation through establishing common goals, knowledge sharing, and communication. For example, mergers have been shown to reduce the potential for innovation because resources are focused on managing structural changes to enhance amalgamation. Conversely, alliances and joint ventures between competitors or between customers and suppliers promote innovation and product development where combining resources and knowledge is beneficial.

Some organizations create “idea champions” (i.e., formal job roles to enhance innovation) or “change

agents” (whose specific responsibility is to encourage and enhance innovation at work). Some organizations create an innovation resource fund specifically to encourage development and implementation of new ideas. Idea capture or suggestion schemes are beneficial to enhancing employee involvement in generating and implementing new ideas.

A stimulating physical environment can influence work processes to enhance innovation. Some organizations have innovation laboratories that make use of technologies such as electronic brainstorming software. Using this software in groups can significantly enhance the quantity of ideas produced because it reduces some of the negative social factors in paper-based group brainstorming such as evaluation apprehension (where members withhold ideas because they fear negative evaluation) and free riding (where members rely on others to accomplish the task).

### 7.4. External Demands and Environment

The context in which organizations operate influences the propensity for innovation. For example, environmental uncertainty predicts administrative innovation, and low market share predicts higher levels of product innovations in some sectors. In this way, uncertainty generated by external demands and potential threats can stimulate innovation. However, under extremely threatening conditions, some organizations tend to centralize control, conserve resources, restrict information flow, and rely on traditional work methods. As a result, the propensity for innovation may be reduced in the long term.

The information flow in and out of organizations will influence innovation to the degree that there is an active search of the environment for new ideas. The external reputations of organizations can influence innovation because innovative organizations may be more likely to attract individuals with a high innovation potential.

## 8. EMERGING ISSUES

Innovation is a process with a variety of situational- and individual-level determinants. Each determinant may be more or less important during the idea generation or idea implementation phase. Similarly, the role of the environment in shaping innovation is likely to have a differential impact on individuals. Future research should explore interaction effects where the

individual differences in personality or experience may influence the way in which a person interprets or reacts to extrinsic rewards or constraints. In general, there is a lack of theoretical integration of the variables involved at each level of analysis and relevant interaction effects. In 2004, Anderson and colleagues reviewed previous innovation research and argued for the development of multilevel theories and study designs and for the triangulation of research methods to address these issues. The need to measure negative predictors of innovation, such as job dissatisfaction, task conflict, and home/work strain, was emphasized. Future research could explore innovation as the independent variable.

Historically, there has been less research targeting group-level innovation, but this has expanded recently and there are now clear practical recommendations regarding enhancement of innovation in teams. Future studies on innovation should reflect the changing nature of work and the use of technology. The impact of dispersed and virtual teams (where members are geographically dispersed and in different global

time zones) on team dynamics and innovation has not been explored. Similarly, the impact of international and cultural differences on team working and innovation (where team selection may also be important) has yet to be explored. Future research could examine the impact of technology and the physical environment on innovation by evaluating innovation laboratories to enhance idea generation and implementation at work.

The role of leaders in enhancing innovation has been neglected. Leaders are the gatekeepers in the innovation process and have a significant impact on the organizational climate and culture. Although organizations recognize the need for innovation, some managers may be less equipped to recruit and develop innovative employees.

### 9. SUMMARY

Innovation is a multifaceted and multilevel process. Figure 1 summarizes the main elements of the innovation cycle and the various influences at different levels

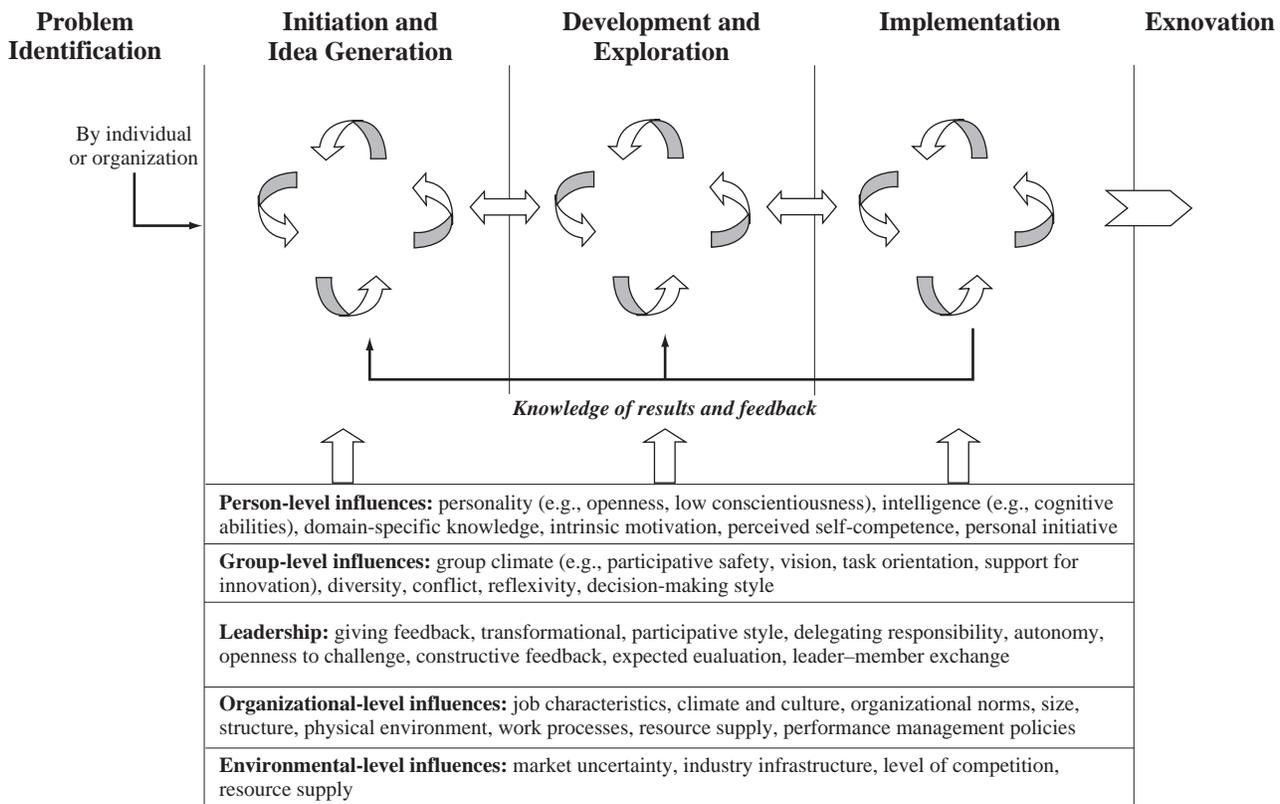


FIGURE 1 Phases in the innovation process and influencing variables.

of analysis. Innovative individuals have a creative cognitive style and are open to new experiences. Innovators are proactive, knowledgeable about the domains in which they are trying to implement their ideas, and persistent in the face of obstacles. A high level of diversity among employees enhances team innovation if there is integration of ideas among employees and team members. A positive team climate can have a significant influence on innovation and is characterized by participative safety, vision, support for innovation, and task orientation. Leaders can enhance innovation and may play different roles during different phases of the innovation process (e.g., encouraging ideas during idea generation, providing adequate resources and selling ideas to others during idea implementation). Many organizational factors influence innovation, including informal characteristics (e.g., climate and culture), formal policies (e.g., performance management systems), and factors external to the organization (e.g., market demand and uncertainty).

### See Also the Following Articles

Agreeableness ■ Intelligence and Culture ■ Intergroup Relations and Culture ■ Leadership and Culture ■ Organizational Culture and Climate ■ Traits ■ Work Motivation ■ Work Teams

### Further Reading

- Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior*, 10, 123–167.
- Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, CO: Westview.
- Anderson, N., De Dreu, C. K. W., & Nijstad, B. A. (2004). The routinization of innovation research: A constructively critical review of the state-of-the-science. *Journal of Organizational Behavior*, 25, 147–173.
- Cattell, R. B., & Eber, H. (1970). *Handbook for the Sixteen Personality Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.
- Csikszentmihalyi, M. (2001). The context of creativity. In W. Bennis, & G. Spreitzer (Eds.), *The future of leadership: Today's top leadership thinkers speak to tomorrow's leaders* (pp. 116–124). San Francisco: Jossey-Bass/Pfeiffer.
- Eysenck, H. (1995). *Genius: The natural history of creativity*. Cambridge, UK: Cambridge University Press.
- Feist, G. J., & Barron, F. X. (2003). Predicting creativity from early to late adulthood: Intellect, potential, and personality. *Journal of Research in Personality*, 37, 62–88.
- Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. Cambridge, MA: MIT Press.
- Frese, M., & Fay, D. (2001). Personal initiative: An active performance concept for work in the 21st century. *Research in Organizational Behavior*, 23, 133–187.
- King, N. (1992). Modelling the innovation process: An empirical comparison of approaches. *Journal of Occupational and Organizational Psychology*, 65, 89–100.
- King, N., & Anderson, N. (2002). *Managing innovation and change: A critical guide for organizations* (2nd ed.). London: Thomson.
- Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology*, 61, 622–629.
- Patterson, F. (2000). *The Innovation Potential Indicator: Test manual and user's guide*. Oxford, UK: OPP.
- Patterson, F. (2002). Great minds don't think alike? Person level predictors of innovation at work. *International Review of Industrial and Organizational Psychology*, 17, 115–144.
- Simonton, D. K. (1991). Emergence and realization of genius: The lives and works of 120 classical composers. *Journal of Personality and Social Psychology*, 61, 829–840.
- Sternberg, R. J. (Ed.). (1999). *Handbook of creativity*. Cambridge, UK: Cambridge University Press.
- Tierney, P., Farmer, S. M., & Graen, G. B. (1999). An examination of leadership and employee creativity: The relevance of traits and characteristics. *Personnel Psychology*, 52, 591–620.
- Tjosvold, D. (1998). The cooperative and competitive goal approach to conflict: Accomplishments and challenges. *Applied Psychology: An International Review*, 47, 285–313.
- Van de Ven, A. H., Polley, D. E., Garud, R., & Venkataraman, S. (1999). *The innovation journey*. New York: Oxford University Press.
- West, M. A. (2001). The human team: Basic motivations and innovations. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology*, Vol. 2: *Organizational psychology* (Vol. 2, pp. 270–288). London: Sage.
- West, M. A., & Anderson, N. A. (1998). Measuring climate for work group innovation: Development and validation of team climate inventory. *Journal of Organizational Behavior*, 19, 235–258.
- West, M. A., & Farr, J. L. (Eds.). (1990). *Innovation and creativity at work: Psychological and organizational strategies*. Chichester, UK: Wiley.
- Zaltman, G., Duncan, R., & Holbek, J. (1973). *Innovations and organizations*. New York: John Wiley.



# Personality and Emotion in Late Life

Nicole Alea, Manfred Diehl, and Susan Bluck

University of Florida, Gainesville, Florida, USA

1. Introduction
2. Personality in Late Life
3. Emotion in Late Life
4. Concluding Remarks: Linking Personality and Emotion  
Further Reading

## GLOSSARY

**affect optimization** Tendency to regulate emotional states toward more positive affective experiences and away from negative emotion.

**cohort** Group of people who are part of the same generation and, as a consequence, share similar life experiences and historical influences.

**cohort-sequential design** A study design that allows researchers to disentangle the effects of cohort and age as it tests different cohorts at the same ages across a period of time.

**cross-sectional design** A study design that allows researchers to draw conclusions about age differences by testing individuals of different ages at one point in time.

**emotion regulation** The ability to self-regulate or manage the internal experience and external expression of emotion.

**emotions** Short-lived negative or positive responses and reactions to events in the environment, or internal cognitions, that involve changes in body physiology, subjective experiences, and expressive behaviors.

**life span development** Development of individuals across the life span characterized by multidirectionality and multidimensionality; that is, developmental trajectories are assumed to include growth, stability, and decline within and across behavioral domains.

**longitudinal design** A study design that allows researchers to draw conclusions about age-related changes by examining the same group of individuals at different ages over a period of time.

**personality** Characteristics of an individual that are relatively stable and endure across time and settings.

**primary emotions** Set of emotions that are fundamental in organizing human thought and action across the life span, including happiness, fear, anger, sadness, surprise, disgust, and interest.

**rank-order consistency** The consistency of an individual's relative placement or rank within a group over time.

This article reviews the literature on personality development and emotional experience across the adult life span, with a particular emphasis on late life. It reviews theoretical developments and challenges, discusses methodological issues, and presents the major empirical findings about personality and emotion in late life. It concludes with a discussion of how future work that better integrates these two areas would be beneficial in advancing our understanding of well-being in late life.

## 1. INTRODUCTION

Advances in health care and public health practices have created a new culture of "old age." But what are old people like? Negative stereotypes suggest that older adults have certain personality characteristics, that is, that they are stubborn, set in their ways, and/or afraid

to take risks. Competing positive stereotypes suggest that older people are full of grace and dignity and are compassionate and wise. How do old people feel? Again, there are competing cultural conceptions. It is often believed that old people are lonely and depressed. On the other hand, old people are often viewed as carefree and satisfied. This article moves beyond cultural stereotypes of personality and emotion in later life to present the theories, methods, and findings gleaned from scholarly research.

Historically, many psychological characteristics (e.g., intelligence, learning ability) were characterized by decline in late life. More recently, however, psychologists have begun to focus on the positive experiences associated with aging. This dynamic tension between the positive and negative aspects of late-life functioning was elucidated in a 1987 article by Paul Baltes of the Max Planck Institute for Human Development in Berlin, Germany. He proposed the life-span approach and postulated that life span development is multidirectional and multidimensional. That is, different psychological characteristics show different developmental trajectories (e.g., growth, stability, decline) and different trajectories may coexist for behavioral functions within the same domain, making growth and positive development possible even into late life. Indeed, Smith's findings from the longitudinal Berlin Aging Study show that successful aging can be defined in terms of the ratio between gains and losses.

Personality and emotion have been at the forefront of research showing positive gains or maintenance of functioning in late life. As such, research on personality and emotion offers a window on facets of successful aging. This article reviews each area of work separately, focusing on major theories, methods, and empirical results.

## 2. PERSONALITY IN LATE LIFE

The term "personality" refers to the relatively stable characteristics of an individual that endure across time and settings. At first glance, this definition suggests that personality must be stable across adulthood; for example, an extraverted 30-year-old is expected to be extraverted when he or she is 65 years old. Evidence supports a good deal of continuity in adult personality, but there are also data showing that change occurs. Whether a researcher favors a view of stability or one of change depends mostly on his or her theoretical perspective.

### 2.1. Theoretical Issues

#### 2.1.1. Personality Theories

Labouvie-Vief and Diehl, in a 1998 review of life span theories, pointed out that two major theoretical perspectives have led the area of personality development. The developmental-level approach is rooted in the psychodynamic tradition, with theorists such as Freud, Jung, and Erikson. This approach is categorized by developmental stages that emphasize qualitative changes in personality during different life periods. For example, Erikson theorized that each stage of development is characterized by a "psychosocial crisis" between opposing tendencies (e.g., generativity vs stagnation during midlife) that must be integrated for the self to fully develop. Although Freud's theory emphasized that personality was set at a fairly young age, later modifications (e.g., by Erikson and Loewinger) and extensions included lifelong processes of personality development.

The trait approach assumes that personality is composed of a number of basic behavioral tendencies (i.e., traits) and has focused on the measurement of these basic traits. McCrae and Costa's "Big Five" model of personality is one prominent example of this approach. According to this theory, there are five broad traits that make up human personality: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Trait theorists emphasize the stability of personality beginning in early adulthood.

A fairly recent addition to the personality literature is the contextual approach, which considers sociocultural and environmental influences that may affect how personality develops across the life span. Researchers such as Helson emphasize that people's environments and experiences influence who they become, at least in part. In contrast to the developmental-level and trait theories, the contextual approach argues for the malleability of personality across a lifetime. What are the mechanisms that lead to changes, or that sustain continuity, in personality?

#### 2.1.2. Mechanisms of Continuity and Change

In a 1999 review, Caspi and Roberts summarized mechanisms that support continuity and those that facilitate change in personality across the adult life span. Genetics are thought to be a major mechanism leading to continuity. Data such as the MacArthur Longitudinal Twin Study suggest that approximately 80% of personality stability can be attributed to genetic

factors. Notable, however, is that genes seem to become less influential in late life. As twins age, they become less similar, suggesting that environmental influences play an increasingly stronger role as individuals age.

Environmental factors, such as differences in life experiences and socialization, are the mechanisms that are thought to produce changes in personality. For example, in a classic study, Helson and Moane showed that interpersonal dominance tends to increase in women who actively pursue a professional career, whereas raising children tends to heighten a woman's level of empathy. Thus, social roles that carry implicit and explicit demands can shape personality during adulthood. Individuals can also learn new behavioral patterns, at any age, by observing others or by getting feedback from others. This type of social learning refutes the old adage that "You can't teach an old dog new tricks." However, social and physical environments can also play a role in maintaining continuity. To the extent that people's environments remain stable, they foster stability in personality. Consistent environments tend to encourage consistency in personality across the life span. Thus, genetics and environment contribute to personality continuity and change in a complex integrated fashion. Researchers' theoretical perspectives, as well as their views of continuity and change, affect the questions they ask and the methods and measures they use.

## 2.2. Methods and Measures

Early theorists, such as Allport, Cattell, and Murray, suggested that personality should be studied over time. During more recent years, Block has advocated studying personality "the long way," that is, using longitudinal methods. Although the majority of personality research is still cross-sectional in nature, the number of longitudinal studies has risen considerably during recent years and invaluable information has been obtained regarding personality continuity and change in late life.

### 2.2.1. Types of Continuity in Longitudinal Research

Everyday conceptualizations of continuity and change in personality are typically thought of at the individual level, that is, involving intraindividual variability in personality (referred to as ipsative continuity). For example, has grandpa changed during the past 10 years, or is he the same person he used to be? Investigating personality

in this way has only recently received adequate attention because doing so requires complicated longitudinal designs of intraindividual change and advanced statistical analyses. Rather, researchers have focused their attention on two other types of continuity. Differential continuity refers to the consistency of an individual's relative placement within a group over time, that is, rank-order consistency. For example, in the case of high rank-order consistency, individuals are expected to preserve their rank over time, whereas low rank-order consistency is indicative of the fact that people are shifting around in terms of their rank. The second commonly studied continuity is absolute continuity, which refers to stability in the level of a particular personality characteristic over time. It is important to note that when most individuals change in the same direction, changes in mean level can occur over time without great changes in rank-order consistency. Thus, there can be various types of continuity, or lack thereof, observed in longitudinal assessments. There are also various factors that affect whether change or continuity is observed.

### 2.2.2. Factors Affecting Whether Continuity or Change Is Observed

Recent meta-analyses conducted independently by Ardel and by Roberts and DelVecchio identified similar factors affecting whether change or continuity in personality is observed in longitudinal studies. Three factors tend to bias results toward stability. The first factor is an individual's age at the initial time of testing. The older people are when personality is first assessed, the greater stability coefficients tend to be. Second, length of retest interval can affect findings. The closer together the times of testing, the more consistent personality appears. Third, the type of measurement instrument can also bias results. Self-report checklists of personality traits, the most commonly used instruments with older adults, tend to overestimate personality stability because individuals are usually biased in favor of reporting continuity. Behavioral observations get around the self-report bias in that they involve observations of how people behave across contexts. Given their contextual nature, behavioral observations are more sensitive to changes in personality, but they have rarely been used with older adults.

Other factors have an effect on whether change in personality is likely to be observed. Life events and transitions (e.g., retirement, widowhood) that individuals go through while participating in a longitudinal study may suggest change in personality. That is, as

individuals respond to new role demands, they may be learning new ways in which to cope. For example, financial strain that can accompany retirement may lead a person to become more worried (with “worrying” being part of the trait of Neuroticism) than they usually are due to the loss of income. This may be a new behavior that emerged as a result of the social transition, and it becomes incorporated into the individual’s personality if reinforced over time or may be a temporary way of dealing with a new life transition. Unknowingly, researchers would observe changes in personality.

Historical experiences and generational effects can also affect whether personality change is observed. In his classic work about the children of the Great Depression, Elder demonstrated how historical events affect people’s personalities in different ways depending on their life stages. Elder found personality change for middle-aged men in reaction to the Depression. Personality continuity was more often found for younger men because these men were able to seize newly emerging opportunities after the Depression. However, conclusions about personality development were confounded with individuals’ birth cohorts; thus, it was unknown whether the findings were a result of maturational change, the time of testing and historical influences, or the participants’ generations. Schaie developed the cohort-sequential design to tease apart the effects of these factors, yet this elaborate design is rarely used. In sum, longitudinal studies, although offering the most comprehensive picture of personality development, can be limited about personality continuity and change. Thus, when reviewing empirical results regarding personality across adulthood in the section that follows, one should pay close attention to the types of continuity found and the methods employed.

### 2.3. Major Findings

Table I summarizes the types of continuity and change evident in research on personality in late life,

the typical statistical methods used, and the general conclusions. This section reviews major empirical results supporting both continuity and change in personality.

#### 2.3.1. Evidence for Continuity

There is quite a bit of evidence pointing toward continuity of personality during adulthood and late life. Costa and McCrae’s Baltimore Longitudinal Study began in 1980 and consists of approximately 1000 participants ages 20 to 96 years. Participants were given the NEO Personality Inventory, which measures the Big Five basic personality traits. In this and other work, they have found moderate to high rank-order consistency for personality traits across time intervals of 5 to 6 years. Costa and McCrae interpreted their results as indicating that personality does not change, except in trivial ways, after the age of 30.

Other researchers espousing the individual differences approach, such as Conley, have also found a great deal of stability in personality across adulthood. Using a different self-report personality measure as part of the Kelly Longitudinal Study, Conley and colleagues found that certain personality traits, such as neuroticism, social introversion–extraversion, and impulse control, demonstrated fairly stable patterns across 25- and even 40-year time intervals. Focusing on late life, Troll and Skaff found continuity in self-concept in the oldest old (85 years of age or over). Participants were asked how they thought they had changed over the years and how they had remained the same. These older adults reported that their core self-concept had remained stable over the years, and their self-descriptors were also relatively consistent. In sum, for different measures used to assess personality, there seems to be differential consistency of personality during adulthood and late life.

TABLE I  
Summary of Personality in Late Life: Types of Continuity, Typical Methods, and Empirical Results

<i>Types of continuity</i>	<i>Typical methods</i>	<i>Empirical results</i>
Differential continuity	Rank-order consistency	Stability: ranks of personality characteristics relatively stable in late life
Absolute continuity	Mean levels	Change: decreases in neuroticism, increases in agreeableness, decreases and maintenance in extraversion in older adults
Ipsative continuity	Intraindividual differences and variability	Stability and change: growth, stability, and decreases in personality characteristics within individuals

### **2.3.2. Evidence for Change**

Advocates of the contextual view, such as Helson, have argued that trait theorists seek to confirm hypotheses about the stability of personality rather than to explain why and how changes in personality occur when they do. In the Mills College Longitudinal Study, Helson found changes in adult women's personalities across a 30-year time period. For example, women became more assertive (i.e., mean-level increase on the masculinity scale) from 20 to 30 years of age, and they became more compassionate (i.e., mean-level increase on the femininity scale) as they approached late life (52 years of age). But what happens in very old age? Field and Millsap analyzed data from the Berkeley Older Generation Study and found consistent age-related trends for the mean level of some personality traits. For example, for this very old sample, they found increases in agreeableness and decreases in extraversion with age. Thus, there seem to be changes in mean levels of personality characteristics, even though rank-order consistency is fairly constant. Studies of intraindividual variability also suggest both change and continuity in personality across individuals' lives.

### **2.3.3. Intraindividual Variability: Evidence for Continuity and Change**

Although few studies of intraindividual variability have been conducted, they provide a comprehensive picture of the variability of personality in late life. A classic study by Jones and Meredith examined individual patterns of personality development over a 30-year time span. For example, they found that some participants' self-confidence was stable until 30 years of age and then increased during middle age, others' self-confidence remained stable across the time period, and others' self-confidence declined with age. In a recent study, Mroczek and Spiro found variability in the rate and direction of change in personality among men ranging in age from 43 to 91 years. Although they found that many men showed declines in neuroticism and stability in extraversion, there were also many men who deviated from this overall pattern, thereby demonstrating individual variability even in old age.

In conclusion, personality in late life is undoubtedly characterized by both differential continuity and changes in mean levels of personality characteristics (i.e., absolute continuity). Studies of intraindividual variability reiterate that continuity and change in personality both occur in late life differentially for

individuals and that group means can be deceiving. However, mean-level changes that occur are relatively positive (e.g., decreases in neuroticism, maintenance in extraversion). Multiple dimensions and directions of change characterize personality in late life. Does a similar variable pattern exist for how people feel in late life?

## **3. EMOTION IN LATE LIFE**

Emotions have been defined as short-lived responses to events in the environment that involve changes in body physiology, subjective experiences, and expressive behaviors. That is, emotions are transient and happen on multiple levels. Some emotions, such as happiness, do not endure over a lifetime but rather are dynamically and subtly tied to specific events and experiences—to cognitions, memories, and personalities. The important questions for emotion research have been whether older adults, given their unique life circumstances, cognitive abilities, and personalities, experience or feel emotions to the same degree as do younger adults, experience the same types of emotions as do younger adults, and regulate and express emotions in the same way as do younger adults. The following theories of emotion in late life address these central questions.

### **3.1. Theoretical Issues**

#### **3.1.1. Emotion Regulation and Affect Optimization**

In 1989, Lawton proposed a theory of the regulation processes of emotion during adulthood, emphasizing positive gains in regulation with age. Furthermore, he postulated that in late life, adults regulate emotion toward "affect optimization," that is, a minimization of negative affect and a maximization of positive affect. He contended that better affect regulation in late life occurs as a function of positive changes in personality as well as changes in social contexts in late life. Examination of social context became a focal area for further theoretical expansion, as seen in the next theory.

#### **3.1.2. Socioemotional Selectivity Theory**

In 1992, Carstensen proposed the socioemotional selectivity theory, which situates emotion in the context of social goals associated with different life phases. This theory focuses on the salience of emotion in older adults' everyday lives. Carstensen, like Lawton, emphasized

regulating emotion toward obtaining positive affective experiences. The core of Carstensen's theory is that older adults show a selective narrowing of social networks to more emotionally close, gratifying social relationships. That is, the regulation of emotion is, in part, a function of social choices; maintaining a positively valenced emotional life is realized through selection of positive social relationships. Carstensen's theory ties emotion regulation to life phase-specific choices regarding social networks. The next theory also ties emotion to life phase, but largely through the relation of emotion to cognition and personality at different points during the life span.

### **3.1.3. Cognitive–Affective Developmental Theory**

Labouvie-Vief proposed a cognitive–affective developmental theory of emotion that delineates the role of emotional functioning in conjunction with the development of cognitive and ego processes (the self) that occur across the life span. She argued that emotion is qualitatively different at various points in the life span. Specifically, with more complex ways of processing information and higher levels of ego development, emotional expression moves from being based on societal standards during adolescence to more sophisticated emotional functioning during adulthood. This includes increases in the complexity of experienced emotion, heightened emotional flexibility, and more tolerance for ambiguous emotions. Labouvie-Vief's theory clearly delineates the connection between emotion and personality, or self-development, as does the discrete emotions theory advocated by Magai.

### **3.1.4. Discrete Emotions Functionalist Theory**

During the 1980s, Magai integrated emotion theory with personality theory by introducing the discrete emotions functionalist theory to research on adult development. Rooted in the differential emotions tradition, this theory emphasizes the functional nature of the primary emotions (e.g., happiness, fear, anger, sadness), focusing on how emotions are used to organize thought and behavior. Magai's work suggested that discrete emotions are hardwired from birth but serve similar functions across the life span. What changes with age is the ways in which emotions are expressed; that is, with increasing age, the complexity and elaboration of emotions increase. In addition, Magai delineated how emotion and personality are intertwined, particularly in late life. She suggested that discrete emotions experienced repeatedly early in

life become traitlike and integrated as “emotion traits” within an individual's personality.

These four theories make unique contributions to the literature regarding how emotion is regulated toward positive affect in late life, how social processes are used to regulate emotion in late life, how life span development of self and cognition encourages more sophisticated processing of emotion with age, and how emotions play a functional role in directing thought and behavior from early childhood but gain increasing complexity and elaboration across adulthood. Despite the differences among these four theories, each theory in some way addresses one of the central issues of how older adults experience emotion, regulate emotion, and express emotion.

## **3.2. Methods and Measures**

Corresponding with the central questions of emotion research, most work has been cross-sectional rather than longitudinal. Emotion researchers use a rich array of methodological tools, including both objective and subjective measures.

### **3.2.1. Objective Measures**

Objective assessments of emotion are psychophysiological measures tied to the basic neurology of emotion (e.g., autonomic nervous system activity, discrete musculature changes during facial expression). Studies assessing the psychophysiological aspects of emotion involve asking participants to look at emotional pictures (e.g., a child playing in a field) or to relive emotional experiences (e.g., a surprising event). Autonomic activity, such as changes in heart rate or breathing, is assessed online during these tasks. Similar stimuli are used when examining facial expressions of emotions. Ekman developed an elaborate system for coding universal, emotion-specific facial expressions. For example, people universally express sadness by pulling down the corners of their lips, and surprise is expressed by a widening of the eyes. These objective measures of emotion inform researchers about the underlying processes and universal nature of emotions.

### **3.2.2. Subjective Measures**

Subjective measures of emotion are more commonly employed. They typically include self-report ratings and, more recently, behavioral observations (e.g., coding the types of emotion present during social exchanges) and

open-ended verbal narratives (e.g., coding emotional content of remembered narratives). Self-report measures (e.g., the Positive and Negative Affect Scale developed by Watson and colleagues) are undoubtedly the most common means of assessing emotion in late life. Adults answer questions about whether, and how intensely, they have felt several negative and positive emotions (e.g., sadness, happiness) over the past month or the past week. Self-reports are subject to response biases and, as Levine and Bluck demonstrated, have a memory component that may lead to greater biasing in older adults. For example, questions about general emotional experience (e.g., “How salient is emotion to you in your everyday life?”) probably require effortful processing if a person actually recalls everyday events to provide an estimate of the general emotional content. Thus, results using such scales are prone to memory error. More naturalistic and ecologically valid measures, such as observing emotional behaviors and collecting emotional memory narratives, may be better suited as subjective measures of emotion in late life. Bluck and Alea recently found that autobiographical narratives and scalar self-reports sometimes provide divergent information about the emotional lives of older adults.

In sum, both objective and subjective measures have been used to examine various aspects of emotion in late life. In terms of subjective measures, there are a number of tools that vary in level of reliability and ecological validity. The next section reviews major findings about emotional functioning during adulthood and late life, paying special attention to the type of measurement employed.

### 3.3. Major Findings

The major empirical findings mirror the chief theoretical concerns in this area: the overall experience of

emotion, the regulation of emotion, and the expression of emotion during adulthood and late life. Table II summarizes the major empirical findings by method of investigation.

#### 3.3.1. Experience of Emotion

Research on the experience of emotion includes both physiological measures and the subjective experience of emotion in daily life. Levenson and colleagues have conducted a series of studies investigating whether older adults experience the same intensity of physiological reactions as do younger adults when they relive emotional experiences. Using online physiological assessments, these researchers have found that, for the most part, older adults exhibit the same physiological responses as do younger adults for discrete emotions. For example, both younger and older adults have increases in heart rate when experiencing fear. The magnitude or intensity of these responses, however, is often lower in late life.

The intensity of experienced emotion has also been examined using self-report measures. For example, Diener and colleagues conducted a large-scale study to assess age differences in the self-reported intensity of emotions in adolescents, younger, middle-aged, and older adults. They used the Affect Intensity Measure, which assesses how intensely individuals experience emotions in various situations (e.g., “I can remain calm even on the most trying days”). Older adults reported less intense emotional experiences than did younger adults. Thus, older adults tend to show the same pattern of physiological reaction as do younger adults, but emotions generally seem to be experienced with less intensity in late life. This holds for both physiological measures and self-reports. But what

TABLE II  
Summary of Emotion in Late Life: Aspect of Emotion, Typical Methods, and Empirical Results

<i>Aspect of emotion</i>	<i>Typical methods</i>	<i>Empirical results</i>
Emotional experience	Psychophysiological	Mixed: no age difference for physiological experience of primary emotions; lower intensity with age
	Self-reports	Lower intensity with age
Emotion regulation	Self-reports	Increase in positive affect and decrease in negative affect with age; better emotion regulation with age
Emotional expression	Facial expression	Mixed: no age differences; more expressive with age; harder to decode expressions with age
	Verbal expression	Express more emotion with age
	Behavioral indicators	Fewer negative behaviors and more positive behaviors

happens if one looks at emotional valence? That is, do older adults experience both positive and negative affect with less intensity than do younger adults?

### 3.3.2. Emotion Regulation

Empirical work investigating the balance between positive and negative affect with age has often been conducted in the context of emotion regulation. Most (but not all) work on emotion regulation is congruent with existing theoretical accounts claiming that older adults tend to regulate their emotions toward affect optimization. Older adults self-report experiencing more positive and less negative affect than do younger age groups. For example, Mroczek and Kolarz analyzed data from more than 2000 adults ranging in age from 25 to 74 years. Using the Bradburn Affect Balance Scale, a self-report scale of positive and negative affect, they found an increase in positive affect after 45 years of age and a linear decrease in negative affect across age groups. Charles and colleagues found similar results concerning negative affect in a 14-year longitudinal study of self-reported positive and negative emotion. Furthermore, Gross and colleagues have demonstrated that increases in positive affect states and decreases in negative affect states also apply to discrete emotions such as happiness and anger. Thus, although researchers have found overall decreases in the intensity of emotion with age, it seems to depend on whether the emotion is positive or negative. In late life, negative emotion is often down-regulated, whereas positive emotion is often up-regulated.

In addition to examining the regulation of emotion through valence, researchers have explicitly asked adults about their ability to self-regulate emotions. In the study by Gross and colleagues, emotion regulation was investigated in a diverse sample (e.g., European Americans, African Americans). Participants were asked how well they could control both their inner and outer expressions of emotion (e.g., "Overall, how much control would you say you have over your emotions?"). They found that across age groups, there was an increase in self-reported ability to regulate emotion. Using a different technique, Labouvie-Vief and colleagues reported similar results as far back as 1989. They asked participants to recall emotional experiences and then conducted semistructured interviews about the types of emotional control or regulation strategies individuals used. They found age differences in emotion regulation, although the primary difference was between adolescents and the two

older age groups (middle-aged and older adults). Thus, using self-reports, there is reasonable evidence suggesting that older individuals are better at regulating emotions than are younger individuals and that older individuals tend to regulate their emotions toward positive affective states. Emotion regulation usually refers to internal experiences of emotion. Does emotional expression differ by age?

### 3.3.3. Expression of Emotion

Emotion can be expressed in three basic ways: facial expressions, verbal expressions, and specific behaviors. Levenson and colleagues examined the facial expressions of young and older adults while they were remembering and describing emotional experiences. The researchers failed to find age differences in how adults expressed the primary emotions of anger, disgust, fear, happiness, and surprise. In contrast, Magai found that older adults were at times more expressive than younger adults but that physical changes in their faces with age (e.g., wrinkles) can mask their expressivity.

Research examining the verbal expression of emotion has also found that older adults express greater emotion than do younger adults. In 1994, Carstensen and Turk-Charles conducted a study in which they asked younger and older adults to remember narrative passages. Compared with younger adults, older adults spontaneously focused more on the emotional aspects of the narratives than on the factual information. Alea and colleagues have found similar age results in their work using autobiographical narratives. Specifically, older adults remembered more emotional information about an autobiographical event than did younger adults.

Behavioral observations of emotional expression, common in the child development literature, have been less common in studies of adulthood. Emotionally expressive behaviors include gestures (e.g., banging one's fists when angry, crying when sad) and body movements (e.g., looking away from something in disgust, shifting when nervous). Carstensen and colleagues conducted one of the few studies examining such behaviors in older adults. These researchers videotaped older and middle-aged married couples discussing positive, negative, and neutral marital issues. They found that, after controlling for level of marital satisfaction, older couples showed less negative emotional behaviors during discussions of unpleasant topics and were more affectionate toward their partners than were middle-aged couples. Thus, work on emotional

expressivity suggests greater expressivity in older adults' verbalizations and behaviors, although the basic facial configurations used to express emotion do not appear to change with age.

In sum, empirical results regarding emotional functioning in late life are centered on the issues of experience, regulation, and expression of emotion with age. Using various measures has provided a rather complex yet comprehensive picture of emotion in late life. In general, emotions sometimes are experienced less intensely with age and at other times are experienced just as intensely in late life as in earlier life; however, for the most part, emotions are regulated toward positive affect and expressed in a variety of ways across the life span. Thus, the pattern of emotional functioning during adulthood is also mostly characterized by gains and maintenance of functioning in late life.

#### 4. CONCLUDING REMARKS: LINKING PERSONALITY AND EMOTION

This article began with a discussion of the competing negative and positive cultural stereotypes about older adults' personalities and emotional lives. Does scholarly work support or refute these different pictures of late life? Theoretical and empirical work on personality development and emotional functioning suggests that although older adults certainly face losses (e.g., widowhood) and challenges (e.g., increases in physical ailments), they also have psychological resources available to compensate for, or moderate, losses in functioning. Personality is mostly stable in late life, and (on the whole) changes seem to be relatively positive. Changes in emotional functioning with age, such as better emotion regulation and decreases in the intensity of physiological arousal levels, may also help individuals to face challenges in late life. Thus, both personality development and emotional functioning represent areas of stability or growth that enhance the experience of growing old.

Consequently, a clear step toward better understanding well-being in late life is further delineating the integral relation between personality and emotion. Although the major personality theories imply that traits have an inherent emotional component, the assessment of personality and emotion linkages has

been rather sketchy so far. Emotion theories have been better at delineating the relation between personality and emotion. For example, Labouvie-Vief's work suggests that emotional development occurs during adulthood in part due to development of the self, and Magai's work postulates that emotions that are repeatedly experienced early in life become salient to an individual and integrated into his or her self-representations across the life span. Recently, with the publication of Block's book *Personality as an Affect-Processing System*, there is a growing awareness that the traditional distinct notions of personality and emotion may limit our understanding. Block suggested that these approaches can be fused in a creative way to examine both phenomena more comprehensively while including a life span developmental perspective relevant to the study of aging. The task for future researchers is to explore empirically how emotion and personality are interrelated and how they act together to influence quality of life in late life.

#### See Also the Following Articles

Aging and Competency ■ Anxiety Disorders in Late Life ■ Cognitive Aging ■ Depression in Late Life ■ Elder Caregiving ■ Emotion ■ End of Life Issues ■ Psychotherapy in Older Adults

#### Further Reading

- Baltes, P. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, 23, 611–626.
- Block, J. (2002). *Personality as an affect-processing system: Toward an integrative theory*. Mahwah, NJ: Lawrence Erlbaum.
- Carstensen, L. L. (1992). Motivation for social contact across the life span: A theory of socioemotional selectivity. In J. E. Jacobs (Ed.), *Nebraska Symposium on Motivation* (pp. 209–254). Lincoln: University of Nebraska Press.
- Caspi, A., & Roberts, B. W. (1999). Personality continuity and change across the life course. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 300–326). New York: Guilford.
- Helson, R., & Kwan, V. S. Y. (2000). Personality development in adulthood: The broad picture and process in one longitudinal sample. In S. E. Hampson (Ed.), *Advances in personality psychology* (Vol. 1, pp. 77–106). Hove, UK: Psychology Press.
- Labouvie-Vief, G., & Diehl, M. (1998). Life-span developmental theories. In C. E. Walker (Ed.), *Comprehensive clinical psychology*, Vol. 1: *Foundations* (pp. 261–296). Oxford, UK: Pergamon.

- Lawton, M. P. (1989). Environmental proactivity and affect in older people. In S. Spacapan, & S. Oskamp (Eds.), *Social psychology of aging* (pp. 135–164). Newbury Park, CA: Sage.
- Magai, C. (2001). Emotions over the life span. In J. E. Birren, & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 399–426). San Diego: Academic Press.
- McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd ed.). New York: Guilford.
- Pasupathi, M., Carstensen, L. L., Turk-Charles, S., & Tsai, J. (1998). Emotion and aging. In H. S. Friedman (Ed.), *Encyclopedia of mental health* (Vol. 2, pp. 91–102). San Diego: Academic Press.



# Personality Assessment

Gian Vittorio Caprara

University of Rome La Sapienza, Rome, Italy

Daniel Cervone

University of Illinois, Chicago, Illinois, USA

1. Modern Approaches and Contemporary Trends
  2. Issues and Strategies
- Further Reading

emphasis on psychoanalysis themes. Like the Rorschach test, the subjects' interpretations of the scenes are considered projections from their subconscious.

## GLOSSARY

**Big Five** A set of personality dispositions (extroversion or energy, agreeableness or friendliness, conscientiousness, emotional stability or neuroticism, and openness to experience) considered by some to be present, to varying degrees, in every individual.

**personality assessment** Procedures designed to identify and evaluate enduring psychological qualities, including modes of thinking, feeling, and acting that characterize the individual and differentiate individuals from one another.

**psychodynamic assessment** Procedures, involving interviews and projective techniques, concerned with assessing the complexity of individual experience, including unconscious motives and conflicts among subsystems of personality.

**Rorschach inkblot tests** Cards with blots of ink (black, black/red, or multicolored) shown to subjects for interpretation. The subjects' interpretations of the abstract designs are considered projections indicative of their subconscious minds.

**temperament** Personality features evident early in life.

**Thematic Apperception Test (TAT)** Similar to a Rorschach inkblot test except that it is produced in shades of black and white and is somewhat more concrete, with an

Personality assessment refers to procedures designed to identify and evaluate enduring psychological qualities, including modes of thinking, feeling, and acting that characterize the individual and differentiate individuals from one another. Assessment techniques address the cognitive, affective, motivational, and moral components of individual functioning, including cognitive abilities and styles, temperament and mood, motives, attitudes and values, habitual behaviors, coping strategies, and self concepts and regulatory mechanisms. Through a great variety of quantitative and qualitative techniques, personality assessors serve both basic and applied research needs in settings in which descriptive, predictive, and explanatory information about people is required. Personality assessments may be conducted for a variety of purposes. One may aim to describe normal variations in any of the previously mentioned individual characteristics in the population at large, to diagnose psychopathology suffered by a small subset of the population, or to obtain a detailed portrait of the psychological structure and dynamics that characterize a particular individual. The existence of widely varying techniques also reflects the presence in the professional field of competing theoretical views that one may endorse when pursuing an assessment of personality.

## **I. MODERN APPROACHES AND CONTEMPORARY TRENDS**

Modern approaches to personality assessment can be traced to late 19th- and early 20th-century efforts to assess intelligence, temperament, and vulnerability to mental illness. While Francis Galton was an advocate of measurement of both intellectual abilities and character, Alfred Binet set the premises for the long-lasting Intelligence Quotient tradition, and Emile Kraepelin and Carl Gustav Jung pointed to word association as a common technique to detect and analyze mental disorders and complexes. The Woodworth Personal Data Sheet, which appeared in 1918, represents the earliest model of a paper-and-pencil self-report inventory to assess personality, whereas the Rorschach inkblot test, which appeared 2 years later, remains the most famous exemplar of projective techniques, which were designed to circumvent the psychological defenses that, in principle, might invalidate responses to self-report paper-and-pencil test items.

Two main lines of inquiry set the stage for subsequent developments in personality assessment. One primarily focused on interindividual genetic differences, whereas the other was mostly focused on the internal dynamics of individuality. Investigators who assessed phenotypic individual differences, namely traits, motives, and attitudes, conceived personality as an architecture of tendencies and habits. In contrast, theorists focusing on internal personality dynamics perceived personality as a system in service of adaptation and pointed to internal forces, external pressures, and defense mechanism that underlie individual personality development and shape its distinctive features of functioning.

During the subsequent decades, self-report trait inventories were commonly used in a variety of contexts to describe and predict performance and adaptation. Projective techniques such as the Rorschach inkblot tests and the Thematic Apperception Test (TAT) were more frequently used in clinical settings. Both were largely endorsed by scientists and practitioners, regardless of their theoretical orientation. Although there was not yet a unique science of personality, most progress came from work in psychometrics, including the extension of factor analytic methods to study the components of personality.

Thus, by the mid-20th century, personality assessment could count on not only a large variety of instruments but also a consistent body of knowledge regarding sources of data and the quality of measures.

It was made clear that one should rely on different sources of data, such as behavioral observations, self-reports, psychophysiological indices, and narratives, to tap aspects of personality, and that alternative instruments may be differentially appropriate to capture general versus unique features of the individual, to assess different constructs, or to apply to different populations or in different settings. In this regard, the seminal contributions of Cattell to psychometrics and statistics and of Cronbach and Meehl to the quality of personality measures and to the various methods to ascertain their validity and reliability were long-lasting and influential.

However, sophistication of methods could not solve all disputes at the theoretical level. Today, although personality psychology is a well-established domain of scientific inquiry, the desirable idea of a unique science of personality in which multiple contributions from social, cognitive, psychodynamic, and dispositional orientations converge synthetically in the service of cumulative knowledge is far from being fully achieved. Instead, one finds a number of distinct broad trends in theory and assessment, each of which can claim to have made scientific progress.

One such trend addresses the subject as an active interpreter of his or her world. The cognitive and affective revolutions ultimately led to better appreciation of the role of the self system in granting unity, continuity, and coherence to individual experience. A new conception of personality as both an agent capable of making things happen according to its anticipations and standards and a social construct embedded in a network of interpersonal relationships has gained broad consensus. This widened the traditional view of personality as a mediated system, further highlighted the importance of continuous individual—environment interactions, reoriented the study of individual differences from global dispositions to contextualized behavioral strategies, and pointed to the systems that mostly account for the attribution of meanings and regulation of action as they operate and develop over the course of life.

Another area of advancement involves the assessment of personality features evident early in life. Research on temperament has begun to assess specific biological systems involved in the generation and regulation of affective experience. Much of this work has demonstrated that the assessment of temperament must consider not only the biology of personality but also the social contexts in which the person acts, especially because the nature and meaning of biological response measures may shift from one context to another.

Studies on emotional, social, and practical intelligence have brought closer the traditions of research on intelligence and personality that had often developed along separate pathways, the former focusing on cognitive abilities and problem solving and the latter on interpersonal relations and social behavior. The discovery of the cognitive unconscious, namely the variety of phenomena associated to automatic and implicit forms of knowledge, gave new impulse to the study of informational and motivational processes that operate beyond individual awareness.

The study of self systems has led to a focus on self-representations, self-beliefs, and self-regulatory mechanisms as they operate in concert to grant unity and coherence to individual experience and behavior. These changes in personality theorizing inevitably have carried new assessment aims, such as the assessment of skills and knowledge that underlie overt behavioral competencies, and have given impulse to the design of new procedures and instruments, including cognitive techniques to measure individual differences in knowledge accessibility and biopsychological techniques for tapping brain systems in personality functioning.

This progress has greatly benefited from major advances in the neurosciences that have yielded enhanced knowledge of brain functioning, as well as from the technological devices that allow closer access to the biological mechanisms underlying thought, affect, and action. In addition, access to large and diverse populations and computer-aided psychometrics further widened the scope of personality assessment process.

## 2. ISSUES AND STRATEGIES

Throughout much of its history a great part of personality psychology has been concerned with individual differences in observable variations in styles of behavior, affect, and cognition. Self-report and observer ratings have been the main methods to assess habitual behaviors, attitudes and beliefs, motives, and values.

A number of authors have posited traits to account for stable patterns of experience and action that people exhibit and that distinguish them one from another across tasks and situations. Within this line of thinking, personality has been conceptualized as a hierarchical organization with high-level traits (e.g., extroversion) that organize lower level tendencies (e.g., sociability), which in turn supervise lower level behavioral habits (e.g., talkative). Within this tradition, a broad consensus

has been achieved on a central question, namely, the number and content of the high-level traits. Five global dispositions are seen as necessary to individual-difference assessment: extroversion or energy, agreeableness or friendliness, conscientiousness, emotional stability or neuroticism, and openness to experience. These or similar individual-difference dimensions have been consistently identified across different cultural contexts. These dimensions have proven to subsume many previous taxonomies, with a variety of specific traits being shown to correspond to facets or variations of each main trait or combination of main traits. Among the most commonly employed instruments to assess the so-called Big Five traits are the Revised NEO Personality Inventory of Costa and McCrae and the Big Five Questionnaire of Caprara *et al.*

During the past 20 years, many studies have corroborated the concurrent and predictive validity of the Big Five in a variety of educational, organizational, and health settings. Yet this approach meets only part of what personality assessment should aim to achieve.

The Big Five provide a sufficiently comprehensive description of personality at the level of surface behavior tendencies. They may serve to tap interindividual differences by pointing to features that may significantly impact on others as well as predict performance in a variety of settings. However, it is beyond the reach of the Big Five to explain the underlying mechanisms that govern behavior and grant its continuity and coherence. The constructs and their associated assessments indicate little about essential characteristics of being a person, such as self-awareness and intentionality. They meet the need of the bystander psychologist, but they miss the perspective of self-reflective agent.

In contrast, psychodynamic theories have long been much more concerned with assessing the complexity of individual experience. In this tradition, interviews and projective techniques such as TAT or the Rorschach continue to be used by clinicians as well as by researchers. Unfortunately, work in this tradition has encountered numerous empirical obstacles since these instruments are hardly amenable to standard criteria. Reviews indicate that in many important applications, projective tests fail to display adequate predictive validity.

Other investigators have sought alternatives to trait and psychodynamic assessment approaches that are able to capitalize on both quantitative and qualitative methods and to meet the requirements of objective research seeking for general laws without sacrificing what is unique of individual experience. One

theoretical grounding for such efforts has been social cognitive theories. In conceiving of personality as an open, dynamic, unifying, and integrating system, social cognitive theorists point to the emerging properties of mind and focus on the processes and mechanisms conducive to knowledge and motivational structures that enable personality to function as a proactive self-regulatory system. Persons and socio-cultural environments are viewed as reciprocally interacting systems, with the core person variables being cognitive and affective mechanisms through which people interpret the world and regulate their own behavior. Thus, social cognitive theorists focus on the self-construction of personality as an integrative and coherent system as it takes place over the course of life, on the processes that enable the system to function proactively in continuous and reciprocal interaction with the environment, and on the self-structures that orchestrate these processes. Where assessment issues come to the fore, the primary targets are self-regulatory structures, including self-representations, self-efficacy beliefs, and personal goals and standards, namely the structures that are involved most centrally in the self-reflective and self-evaluative properties of mind that operate at the individual level to meet environmental opportunities and constraints and to maintain personally valued courses of action.

### See Also the Following Articles

Extroversion–Introversion ■ Psychological Assessment, Standards and Guidelines for ■ Psychometric Tests ■ Traits

### Further Reading

- Aiken, L. R. (1996). *Personality assessment methods and practices*. Kirkland, WA: Hogrefe & Huber.
- Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26.
- Caprara, G. V., Barbaranelli, C., Borgogni, L., & Perugini, M. (1993). The Big Five Questionnaire: A new questionnaire for the measurement of the five factor model. *Personality and Individual Differences*, 15, 281–288.
- Caprara, G. V., & Cervone, D. (2000). *Personality: Determinants, dynamics, and potentials*. New York: Cambridge University Press.
- Cattell, R. B. (1946). *Description and measurement of personality*. New York: World Books.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Cronbach, L. J. (1970). *Essentials of psychological testing*. New York: Harper.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327.
- Kagan, J. (2003). Biology, context, and developmental inquiry. *Annual Review of Psychology*, 54, 1–23.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1(2).
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Wiggins, J. S. (2003). *Paradigms of personality assessment*. New York: Guilford.



# Personal Space

Robert Sommer

University of California, Davis, California, USA

1. Personal Space Concept
  2. Measurement of Personal Space
  3. Research Findings
  4. Applications
- Further Reading

## GLOSSARY

*individual distance* The amount of space between organisms and their conspecifics.

*personal space* The emotionally tinged zone around the human body that people feel is “their space.”

*proxemics* A term introduced by Edward Hall for the study of spatial relationships; Hall identified four interaction zones: intimate, personal, social, and public distance.

*territory* A fixed geographical space marked and defended by an organism and used for life-sustaining activities.

Because the concept of personal space originated in ethology, some of the most promising theories to explain why humans maintain predictable distances from one another employ a Darwinian (functional) framework. Researchers have investigated differences in interpersonal spacing associated with culture, gender, age, personality factors, and situation. Applications of the concept are found in design, communications, and legal cases.

## 1. PERSONAL SPACE CONCEPT

The term “personal space” was introduced into the social psychological literature to describe the emotionally tinged zone around the human body that people feel is “their space.” It has been described using analogies to a soap bubble, a snail shell, and an aura around the human body. The dimensions are not fixed but rather vary according to internal state, age, culture, and context. The concept has its roots in animal studies, particularly the work of ethologists and zoologists, because most animals in the wild maintain defined distances from conspecifics, and these distances are influenced by an animal’s age, size, gender, social rank, and other factors.

## 2. MEASUREMENT OF PERSONAL SPACE

Methods for measuring personal space in human interactions include both field studies and laboratory simulations (Figs. 1 and 2). A favorite laboratory technique is the invasion, which produces tension-reducing responses such as gaze aversion and withdrawal. In the stop distance technique, a researcher approaches the participant, who tells the researcher to stop when he or she comes uncomfortably close. In the approach distance technique, the participant is asked to move toward another person and to stop at a comfortable interaction distance. In figure placement tests, the participant places manikins, dolls, or other human



FIGURE 1 Field observation of seating in a college library.



FIGURE 2 Laboratory simulation of a space station environment.

surrogates in various social arrangements such as having a conversation. Other measurement techniques include questionnaires and physiological recording.

### 3. RESEARCH FINDINGS

Interpersonal distance is decreased by similarity between participants, acquaintanceship, attraction, cooperation, and cohesion. It is increased by threat, anxiety, stigma, smoking, and mental disorder. Children have smaller interaction distances than do adults, female pairs generally sit and stand closer than do male pairs, and people from Latin cultures use smaller interaction distances than do people from Anglo-Saxon backgrounds.

Environmental variables such as noise, crowding, and room size affect interaction distance.

## 4. APPLICATIONS

There is a flourishing research literature on interpersonal spacing in human services fields such as social work, nursing, counseling, and family therapy. Guidebooks and self-help books describe appropriate spacing in various interpersonal encounters. Some books dealing with sales, management, and dating specifically recommend the aggressive appropriation of space (i.e., standing excessively close) as a form of impression management.

As personal space enters the popular culture, airlines advertise more of it in their seating, residents of homeless shelters complain that they have too little of it, and corporate training manuals warn employees to respect each other's personal space to protect the company against lawsuits.

### 4.1. Design and Architecture

The personal space concept has been applied in the design of offices, stores, banks, and other building types, but its greatest applicability is in mass transit and institutional settings with fixed seating and little opportunity for personal mobility. The U.S. space agency, the National Aeronautical and Space Administration (NASA), used the results of personal space research to improve habitability in the space station.

### 4.2. Communication Skills

Diplomats and corporate executives being posted to different cultures are taught about the different ways in which people around the globe use space and time. These ideas have also been applied to interactions between salespeople and customers, police interrogations, nurse-patient and therapist-client relationships, and interactions among family members.

### 4.3. Legal Uses

In the United States, the concept has had nearly as much application in the courtroom as in design. It has figured prominently in lawsuits on prison crowding and became a pivotal issue in cases of sexual harassment as unwanted closeness was interpreted as a form of harassment toward individuals who lacked power. Judges have used the concept in establishing "no intrusion zones"

between protesters and their adversaries. In high-profile trials, space consultants advise lawyers during jury selection, a time when potential jurors are being screened by both the prosecution and the defense for signs of bias. Nonverbal behaviors such as posture, gesture, and interpersonal distance are observed as potential jurors respond to lawyers' questions.

### See Also the Following Articles

Environmental Stress ■ Privacy ■ Residential Satisfaction and Perceived Urban Quality

### Further Reading

- Aiello, J. R. (1987). Human spatial behavior. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (Vol. 1, pp. 505–531). New York: John Wiley.
- Bechtel, R. B. (1997). *Environment and behavior*. Thousand Oaks, CA: Sage.
- Sommer, R. (1969). *Personal space*. Englewood Cliffs, NJ: Prentice Hall.
- Sommer, R. (2002). Personal space in a digital age. In R. B. Bechtel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 647–660). New York: John Wiley.



# Person–Environment Fit

Cheri Ostroff and Kerstin A. Aumann

Columbia University, New York, New York, USA

1. Overview and Definitions
  2. Creation of and Importance of Fit
  3. Assessing and Measuring Fit
  4. Selected Applications of Fit
  5. A Potential Paradox of Fit
  6. Conclusion
- Further Reading

*person–job fit* The extent to which an individual's skills and abilities match the skill and ability requirements of the job.

*person–organization fit* The extent to which an individual's attributes (e.g., goals, values, personality) are consistent with those of the culture or climate of the organization.

*supplementary fit* The extent to which an individual's attributes are congruent with or similar to those of the group or organization.

## GLOSSARY

*attraction–similarity–attrition (ASA) cycle* Cycle whereby individuals are attracted to, selected by, and remain in organizational environments that have characteristics similar to their own.

*complementary fit* The extent to which an individual's attributes provide something that is needed or lacking in the environment (e.g., the group, the organization).

*levels of analysis* Delineation of organizations into interrelated levels of analysis—the individual, the group or unit, and the organization as a whole.

*occupational fit* The extent to which an individual's vocational interests and skills match the opportunities and requirements of the occupation.

*person–environment fit* The extent to which the characteristics of an individual (e.g., skills, abilities, goals, needs, personality, values) match the characteristics of the work environment (e.g., job requirements, work group climate, organizational goals, organizational culture).

*person–group fit* The extent to which an individual's attributes (e.g., skills, goals, values, personality) are similar to or complement those of others in the work group.

Person–environment fit theory addresses the extent to which the characteristics of an individual are similar to, or are complementary to, the characteristics of the work context, group, or organization. The degree of “fit” has implications for individual attitudes and behaviors in organizations, as well as for group and organizational effectiveness. Fit can be created through naturally occurring processes such as a cycle of being attracted to, selected by, and remaining in an organization, or through deliberate interventions such as socialization and training. Fit has a broad range of applications in organizations, including culture and climate, selection and socialization, and team composition and diversity.

## 1. OVERVIEW AND DEFINITIONS

Consider the hiring practices of two different organizations in the same manufacturing industry. In the first organization, the selection process entailed the

screening of each applicant's résumé to assess relevant background and work experience, a written test to assess job knowledge, and a short interview. This type of hiring process is not unusual in many manufacturing organizations. Now consider the hiring process used at Toyota when it opened a new plant in Kentucky. The process entailed a general knowledge test, a manufacturing exercise to provide a realistic job preview of the type of work to be performed, surveys of attitudes toward work, extensive interviews, and a physical exam. Each applicant spent at least 18 hours engaged in the selection process. Why would an organization devote so much time, money, and resources to hire employees from the lowest to highest job levels?

In the first case, the organization's primary concern was to achieve person–job (P–J) fit, that is, to hire employees with the relevant skills and abilities for specific jobs in the organization without specific consideration of the extent to which job applicants' characteristics are congruent or compatible with the characteristics of the organization in which the job is embedded. In the second case, the intensive selection process was geared toward hiring "whole" persons, that is, those who not only would fit the job but also would fit well into the organization's culture and the organization as a whole. Organizations that engage in such intensive hiring practices maintain that P–J fit alone is not sufficient to ensure that an employee will perform well in the organization. Rather, they advocate that employees who fit the job and the environment of their work group and the organization overall will be better performers, more satisfied employees, and less likely to leave the organization.

### 1.1. Definition: What Is Fit?

The fit or match between the characteristics of the individual, on the one hand, and the characteristics of the environment, on the other, is what is generally referred to as person–environment (P–E) fit. On the person (P) side, these characteristics may include skills, abilities, personality, personal needs, goals, and interests. On the environment (E) side, these characteristics may include specific job requirements, reward systems, goals, culture, climate, and resources of the unit or organization.

The underlying assumption is that a good fit between the characteristics of people and those of the environment will result in better performance, higher job satisfaction, enhanced loyalty to the organization, more positive attitudes, and a lesser desire to leave the

organization for another job. Furthermore, it is assumed that fit is a dynamic and developmental process such that a lack of fit will prompt a change in careers, jobs, groups, or organizations in the hope of achieving fit. Although this notion may appear to be intuitive, understanding how fit operates in organizational settings is complex. For example, what processes help to create fit? How can an organization best identify and attract those individuals who will better fit the organization? How can an organization assess which job applicants will ultimately be a good fit for the job, the work group, and the values and culture of the organization? Is fit always desirable? These types of questions represent the core of fit research and theory and are addressed in the following sections.

### 1.2. Brief History: From Interactional Psychology to Fit

Researchers in organizational behavior are concerned with understanding and predicting people's responses in organizational settings. Historically, researchers debated whether individuals' responses are primarily a function of their personal attributes (e.g., personalities, knowledge, past experiences) or the characteristics of the situation (e.g., organizational climate, rewards system, organizational structure). Interactional psychology theories assume that individuals and their situations are interdependent. The characteristics of the situations constrain and affect the responses of individuals, whereas individual responses simultaneously influence characteristics of their situations or organizations. In other words, people react to situations and also create or enact them.

Although the interactionist perspective emphasizes the importance of both personal and situational characteristics, little theoretical explanation as to how person and environment variables interact to affect behavior was provided. The principle of congruence or fit is one explanation that has gained considerable recognition, and it highlights that person and environment factors should combine through "matching." The optimal combination is one in which the personal and situational characteristics are matched or compatible. Conditions failing to reach this kind of match will lead to a state of incongruence that will have negative effects on outcomes such as performance and satisfaction.

This notion that a better fit between people and their environment results in more positive behaviors and responses has a long history in various fields of psychology, particularly social psychology, personality, group

dynamics, and educational psychology. For example, in social and personality psychology, laboratory experiments over the past 40 years or more have suggested that for each individual, there are environments that more or less match the characteristics of the individual's personality. In educational psychology, seminal studies beginning in the 1950s and 1960s demonstrated that larger discrepancies between self-ratings and ideal ratings of the college environment and between self-ratings and other students' ratings of the environment were related to greater dissatisfaction with college life. In vocational psychology, a substantial amount of research has been devoted to Holland's seminal theory of vocational choice, indicating that greater career success and satisfaction with work purportedly results from a match between individuals' interests and the tasks and activities associated with broad classes of occupations (as opposed to specific attributes of a particular job or a particular organization).

The early work in industrial and organizational psychology research domains addressed (implicitly if not explicitly) the notion of fit to job characteristics. In particular, a great deal of emphasis was placed on personnel selection, that is, devising selection procedures and tests to maximize the probability of an organization hiring individuals whose skills and abilities matched the skills and abilities required for the particular job. Through training, attempts were made to maximize fit by upgrading the skills of individuals to better match the job requirements. Furthermore, in the area of job design, emphasis was placed on redesigning jobs to fit the abilities and limitations of the people performing them. This early work focused largely on fit as it pertained to skills and abilities or the specific type of attributes and tasks inherent in the job itself. Surprisingly little attention was directed at understanding fit in the broader social, work group, and/or organizational contexts.

Starting in the 1980s, theory and research on fit in organizational psychology domains became more complex and integrative. There was a shift from a predominantly skill-oriented P-J fit perspective to a broader perspective of addressing fit between a wide range of individual attributes and a wide range of job, work group, and organizational characteristics across a wide range of organizational areas.

### **1.3. Fit to Different Levels of Analysis**

The primary emphasis in organizational fit research has been on the extent to which individuals fit the attributes

of their job, their group, or their organization. As a result, three broad types of fit have been proposed. P-J fit focuses on congruence or fit between an individual's skills and abilities and the skills and abilities required to perform a specific job in the organization. Person-group (P-G) fit addresses the extent to which an individual's characteristics (e.g., personality, skills, values, goals) are congruent or aligned with the characteristics of other members of his or her work group or work unit. Fit to the broader organizational context is addressed through person-organization (P-O) fit focusing on the extent to which an individual's personality, goals, values, and attitudes fit in with the organization's broader context characteristics such as organizational climate, goals, values, policies, and systems.

Theory suggests, and some research has begun to illustrate, that aggregating across multiple types of fit (e.g., P-J and P-G together) is related to more positive responses (e.g., job satisfaction, reduced stress) than when fit is assessed by only a single type of congruence. On the other hand, some have argued that certain types of fit may be more important than others. For example, when evaluating the extent to which they fit, individuals tend to place a heavier emphasis on fit with salient environmental aspects than on fit with less salient environmental aspects. To date, little research has explored the relative importance of different types of fit.

### **1.4. Supplementary and Complementary Fit**

The preceding conceptualizations define fit broadly as the congruence between individuals' characteristics and those of the job, group, and/or organization. However, compatibility or alignment is further delineated into supplementary and complementary fit. A situation in which an individual's attributes are similar to those of other individuals or to the characteristics inherent in the setting is referred to as supplementary fit. In contrast, when an individual's characteristics enhance or supply what is missing in the work environment, it is defined as complementary fit. For example, consider fit at the group level. Harmonious relations among group members might be achieved when supplementary fit on personality exists because each individual's personality is similar to the personalities of others in the work group. On the other hand, to function effectively in performing their task, a group might require a blend of people with different interpersonal strengths and personalities, for example, one

person with leadership qualities, one with skills at conflict resolution, and so forth. Thus, fit can exist when one entity provides what the other needs (i.e., complementary fit), when they share similar characteristics (i.e., supplementary fit), or both.

## 2. CREATION OF AND IMPORTANCE OF FIT

A number of theories explain how fit is achieved over time and why fit occurs. This section first addresses broad theories that explain the overall process that creates fit before considering current theories on why fit occurs and why it is important for both individuals and organizations.

### 2.1. Processes That Create Fit: The Attraction–Selection–Attrition Cycle and Socialization

Schneider's attraction–selection–attrition (ASA) cycle seeks to explain how persons and the organizational environment become naturally integrated over time. Individuals are attracted to and want to join organizations that have characteristics similar to their own, organizations restrict the range of the types of persons in them through the recruitment and selection of people who have the desired attributes and who are able to comprehend what the organization's context and goals should be, and, individuals who do not fit the organization leave, either voluntarily or involuntarily. At the same time, the kinds of persons in the organization determine the structures, processes, and general climate and culture of the organization.

The effects of the ASA cycle can be enhanced through socialization processes. Through mechanisms such as orientation programs, training, mentoring, and informal discussions, new employees learn the predominant values, attitudes, goals, social knowledge, and expected behaviors in the organization. Socializing new employees can lead to changes in their values or attitudes in the direction of the organization's values, or it can result in their leaving if the organizational characteristics continue to be incompatible with their personal attributes. Over time, through the ASA cycle and socialization, individuals within an organization tend to be homogenous and share many common personal attributes. Indeed, some research has shown that the personality characteristics of individuals within

organizations are more similar to each other than those between organizations. Furthermore, individuals with longer tenure in organizations tend to have personalities that are more consistent with the climates in the organizations.

### 2.2. How Fit Affects Responses in Organizations

A number of different theories have been proposed to explain why fit occurs and how it affects employees' responses and reactions. One perspective is "needs–supplies." Individuals have a variety of needs, desires, and preferences (e.g., achievement, safety, nice working conditions, affiliation, pay, growth and development). When these needs are met or satisfied by the job, group, and/or organizational contexts, individuals will perceive a good fit and are likely to feel fulfilled. As a result, they will be more motivated and satisfied and will perform better. A related perspective is based on "abilities–demand," which suggests that fit occurs when an individual has the abilities required to meet what the organization demands (e.g., time, effort, commitment, skills from employees). Here, when the demands of the organization are met, employees fit the requirements of the organization, allowing for more effective functioning.

Refinements and additional explications of these broad theories have been offered to further explain why fit occurs and how fit guides behavior, reduces stress, and reduces cognitive demands. First, the similarity–attraction paradigm proposes that the perception of similarity to others in background, attitudes, and lifestyles is the most important determinant for attraction dynamics. Individuals who perceive more similarity to others are more willing to pursue further interaction, work together, share knowledge, and form productive work relationships. Second, fit purportedly helps to guide behavior. The environment of the organization is satisfying and reinforcing when it is consistent with individuals' personalities. This consistency or congruence increases the stability of behavior because employees are reinforced for their behavior, whereas incongruence is not reinforced and stimulates a change in behavior. Third, when individuals' needs are met by the organization, or when ability–demand fit is achieved, stress and anxiety are reduced, thereby allowing for greater satisfaction, attitudes, and performance. Finally, when people fit the environment, cognitive demands on employees are reduced. They can more

readily interpret the work environment correctly and discern what is valued as important. Less learning, socialization, and adjustment are required on the part of employees, allowing them to adapt to the organization more quickly and to perform more easily. Internal conflict is also likely to be reduced due to employees having more similar or better alignment of values, competencies, and attributes.

### **2.3. Relationships between Fit and Outcomes**

Fit has been shown to play a significant role in how job applicants choose organizations and how recruiters select applicants. A higher degree of fit has also been consistently related to higher job satisfaction, increased commitment to the organization, more citizenship or helping behaviors, greater career success, reduced stress, and less likelihood of leaving the organization. Although theory suggests that positive outcomes will emerge for the group and organization as a whole, very few studies have examined the impact of fit on group or organizational effectiveness. It is important to note, however, that some P-E fit theorists argue that fit can also have a negative impact on organizations. This potential paradox of fit is addressed in a later section.

## **3. ASSESSING AND MEASURING FIT**

A variety of different techniques, methodologies, and analytical procedures have been used in studies of fit, although debate continues with respect to which conceptualizations and procedures are superior. This section highlights some of the fundamental conceptualization and measurement issues in conducting fit research.

### **3.1. Commensurate Dimensions**

What classes of variables constitute the P and E factors? Although this list is not exhaustive, the key variables on the P side include values, personality, needs, desires, goals, values, interests, skills, abilities, and personal resources such as time and effort. On the E side, core variables include organizational culture, values, organizational climate, values, goals, norms, skill and ability requirements, pay and rewards, and job characteristics. In studying fit, it is recommended

that the P and E dimensions be commensurate or assessed in the same language and with the same classification scheme. For example, in examining P-O fit on cultural values, the P variables are often a list of 30 to 50 value statements, and individuals are asked to assess the personal importance of each value. The E variables contain the same list of 30 to 50 value statements, and assessments of how important these values are in the organization are made. Similarly, in examining the extent to which needs are met, the P factor represents a list of dimensions that individuals might desire (e.g., high pay, friendly coworkers, challenging work), whereas the E factor consists of the extent to which the work environment provides each of those same characteristics (e.g., employees receive high pay, work environment is friendly, jobs are challenging). In this way, direct comparisons of the degree of fit can be made on the same set of P and E variables or dimensions.

### **3.2. Direct Fit Versus Indirect Fit**

Direct fit and indirect fit represent two different conceptual and operational approaches to fit. Direct fit focuses on individuals' own perceptions of the degree to which they fit into the work environment. It is derived by asking individuals the extent to which they believe they fit into the work environment, either generally (e.g., the extent to which they believe their skills fit the job) or for some particular dimension (e.g., the extent to which their values fit those of the organization). Here, for example, the focus is on how much individuals personally believe that their values fit those of the organization, but it is not known what the individuals' values or the organization's values actually are.

In contrast, indirect fit is based on separate measurements of the P and E factors, with fit being assessed statistically by jointly comparing the two P and E measures. For example, individuals' reports of their own values are compared with a separate assessment of the cultural values in the organization. In this case, the interest is in determining whether fit between individuals' values and organizational values is related to attitudes and behaviors, regardless of whether or not the individuals actually perceive this fit. Research is only beginning to address the relative importance of these different conceptualizations of fit. For example, some research has indicated that direct fit is more strongly related to individuals' behaviors and attitudes than is indirect fit.

### 3.3. Sources and Focus of Measurement

The source from which the P and E factors are derived can vary (e.g., individual, other employees, manager, independent assessor, objective data) depending on the purpose and research question. In many studies, both the P and E factors are measured from the same individuals. For example, individuals are asked to report on their desired characteristics of a work environment and then to report on their perceptions of their current organization's work characteristics. The degree of similarity or discrepancy between these two is taken as an index of fit. Same-source measurement is appropriate when the research question is to assess individuals' perceptions. However, the disadvantage is that relationships might be artificially inflated due to common method variance.

When attempting to address fit to the "actual" environment rather than to individuals' own perceptions of the environment, the E factors are typically derived from a separate source. For example, if the research question is to determine the extent to which an individual's values fit the cultural values of the work group, the individual might report his or her values while other members report on the values in their work group. Here, the individuals' values are compared with the aggregate of other employees' perceptions of the organization's values. Likewise, if the interest is in evaluating the effects of fit between individuals' goals and managers' goals, individuals' attributes would be compared with assessments from managers.

### 3.4. Analytical Issues

In testing fit, a number of different statistical indexes have been used, including difference scores (i.e., the algebraic or absolute difference between the score on the P measure and the score on the E measure),  $D^2$  measures (i.e., the square root of the sum of squared differences between P and E across variables), and profile comparisons (i.e., the correlation between the profile of P dimensions and the profile of E dimensions). These indexes were criticized by Edwards for their conceptual and methodological shortcomings, and polynomial regression procedures have been touted as a more appropriate methodology for assessing congruence. Polynomial regression procedures allow for testing specific functional forms of congruence corresponding to a more specific conceptual model of interest. For example, one can assess

whether a lack of fit results in more negative outcomes when individual values are stronger than organizational values compared with the reverse, or one can assess whether congruence when both the individual and organization values are evaluated highly is related to more satisfaction than is congruence when both the individual and organizational values are rated low. Although polynomial regression procedures have increasingly gained acceptance in fit research, this technique is still the subject of some controversy because it requires very large sample sizes and is best for examining the dimensions of fit one at a time rather than simultaneously.

## 4. SELECTED APPLICATIONS OF FIT

Fit has been examined across a wide range of topic areas, including selection, training, job design, motivation, leadership, goals, diversity, and teams. This section focuses on three specific applications of the fit concept in the areas of organizational culture and organizational climate, selection and socialization, and team composition and diversity.

### 4.1. Organizational Culture and Climate

Organizational culture and organizational climate are two key attributes that differentiate organizations from one another. Furthermore, both culture and climate focus on how employees interpret, experience, and make sense of their organizations. As a result, both culture and climate have become important elements of P-O fit.

An organization's culture is composed of its fundamental ideologies underlying assumptions and core values. It is deeply embedded in the organization and is collectively held. Climate is based on the ways in which individuals experience and describe what is happening at work. Employees' perceptions of what the organization is like in terms of its practices, policies, procedures, and routines are the foundation of psychological climate (individuals' own perceptions) and organizational climate (shared perceptions of the situation across individuals). Culture and climate are aspects of the "supplies" side of the needs-supplies theory underlying P-O fit. To the extent that these organizational supplies are congruent with and

meeting employees' individual needs, supplementary P-O fit will exist.

In terms of fit to organizational culture, much research has assessed congruence between individuals' values and the cultural values espoused in the organization. Studies have been directed at understanding the implications of both direct and indirect fit for cultural values and have demonstrated that greater value congruence is related to greater satisfaction, commitment, adjustment, performance, career success, and tenure. Furthermore, fit between employees' preferred organizational culture and their employer's actual organizational culture has been shown to predict task performance as well as to enhance organizational citizenship behaviors such as helping other employees and volunteering for tasks. Similarly, with respect to climate, some work has indicated that congruence between individuals' needs, personalities, and work environment preferences, on the one hand, and the organization's climate, on the other, is related to higher individual satisfaction at work, commitment to the organization, involvement, performance, and adjustment to work as well as to lower levels of stress, turnover intent, and absenteeism.

Unfortunately, very little empirical research has been devoted to examining the impact of P-O fit to culture and climate on group or organizational outcomes. However, some research suggests that organizational effectiveness is greater to the extent that employees' personalities overall are congruent with certain dimensions of the climate of the organization. Likewise, although theory suggests that hiring employees whose personal values are consistent with future desired organizational values can be an effective tool to promote change in the organization, research is lacking in this area.

The implication of this body of work is that individuals' well-being and quality of work life are enhanced to the extent that individuals fit, or perceive that they fit, into their organization's climate and culture. Organizations desiring to maximize these employee attitudes can foster such P-O fit through practices such as management, rewards, mentoring, hiring, and socialization.

#### 4.2. Selection and Socialization

Organizations traditionally focused on P-J fit in their selection tactics and training and development processes because high P-J fit (i.e., abilities-demands fit) has been demonstrated to positively affect task performance, tenure, and attitudes. More recently, attention

has been devoted to understanding the role of P-O fit in recruiting, selection, and socialization.

Values represent a starting point for P-O fit because an organization's cultural values affect what types of applicants are likely to seek, and find, employment there. Fit plays a significant role in how job applicants choose organizations and in how recruiters select applicants. Recruiters are more likely to make hiring recommendations to applicants who they perceive as having a stronger P-J fit, a stronger P-O fit, and values that are consistent with those of the organization. At the same time, job seekers' perceptions of value congruence are related to perceptions of their general P-O fit perceptions (i.e., how they would fit into the organization overall), which in turn positively affects attraction to the organization. Moreover, new hires whose values are more congruent with those of the organization are more likely to have higher job satisfaction and commitment to the organization and are more likely to remain in the organization 1 to 2 years later.

During socialization, new employees acquire an understanding of the attitudes, values, behavior, and knowledge required to participate as organizational members. Socialization practices, such as social activities that foster interaction, mentoring, and sets of formal and collective (group) learning experiences, can enhance alignment between newcomers' values and organizational values over time. The underlying assumption is that such tactics lead to a change in individuals' values. Although individuals' values are often thought to be relatively stable and difficult to change, there is also some evidence suggesting that employees' work-related values may be malleable if the organization's processes are strong.

#### 4.3. Team Composition and Diversity

Fit in the group context has received the least amount of research attention to date. The distinction between supplementary fit and complementary fit appears to be particularly important when considering fit issues in teams or groups. Individuals' attitudes are likely to be more positive when the individuals fit well into the group and have attributes (e.g., skills, abilities, goals, values, personalities) that are similar to those of other group members (i.e., supplementary P-G fit). Likewise, in terms of the demographic composition of the group (e.g., age, race, sex, education), research has consistently demonstrated that individuals who are

more similar to others are more likely to have positive attitudes, better performances, and longer tenures.

However, it is also important to consider the composition of the group as a whole, that is, how the attributes of the group members complement one another to influence group functioning and group effectiveness. Effective group processes and performance may depend on supplementary fit, as in the case of a customer service team where all roles are shared by all group members and, hence, homogeneity in skills and interpersonal style may result in better team performance. Alternatively, a complement of more heterogeneous skills and styles might be needed, as in the case of retail teams dealing with diverse customers where a range of personality types may be linked to better team performance. Similarly, some research suggests that greater diversity in terms of demographic characteristics is also likely to have a positive impact on group creativity and performance. The underlying notion here is that some heterogeneity in characteristics and attributes is desirable. Complementary fit helps to ensure that all of the various skills and roles needed in the group are available.

The issue of achieving the appropriate balance between supplementary fit and complementary fit is a difficult one that has yet to be explored in detail. The understanding of the optimal profile or configuration of attributes across individuals for different types of work settings, tasks, and roles is in its infancy.

#### 4.4. Implications

Although P-J fit can result in better task performance, research findings increasingly suggest that task performance alone is not sufficient for an organization's overall effectiveness. In particular, contextual performance or prosocial and organizational citizenship behaviors and commitment to the organization have been shown to contribute to organizational effectiveness. Furthermore, there is evidence that contextual performance and organizational commitment increase when P-O fit is high. As a result, all three types of fit should be attended to in making selection decisions and in designing practices such as training and socialization. More specifically, an organization should first consider (a) P-J fit as a match between the applicant's skills and abilities and the requirements and skill demands of the job to promote better job proficiency and performance; (b) P-O fit between the applicant's and organization's values, needs, and goals to enhance satisfaction, commitment, citizenship behaviors, and retention; and (c) interpersonal

attributes and broad proficiencies of the applicant to promote appropriate compatibility and alignment among group members to enhance group performance and intragroup relations. The relative emphasis placed on each will likely differ depending on the organization's overall goals and strategies.

Some researchers suggest that P-J fit may be more important for jobs with greater technical skill requirements, whereas P-O fit may be more important when the organization has a distinctive culture. Others suggest that a multiple-hurdle approach may be the most effective method for considering fit in the selection and recruiting process. For example, one suggestion is to first focus on identifying candidates with high P-J fit and then, from those candidates, to select individuals who are higher on P-G and P-O fit. Alternative sequences might be equally plausible. For example, when the organization provides a great deal of job training to new employees, it may be more appropriate to first identify those with high P-O fit and P-G fit and then, from these candidates, to select those with stronger P-J fit.

In terms of fit to groups, the needs that are most critical to team effectiveness should be identified and prioritized based on the various roles within the group. Supplementary fit on attributes such as values, goals, and personality can help to foster cooperation among team members. Yet members of the group also need to have complementary proficiencies so that the weakness of one group member can be offset by the strengths of another so as to enhance overall group performance. Thus, teams need to identify the unique attributes (e.g., specialized skills, interpersonal proficiency, problem-solving ability) that are not prevalent in current group members and that need to be brought into the group to enhance group performance and effectiveness. Once individuals are hired, additional socialization and group relations training is likely to be important due to the diversity in skills and other characteristics among group members.

The majority of fit research has been conducted on new hires, with the underlying assumption that newcomers who do not fit their job, their group, or the organization will adjust poorly and will be more likely to leave the organization. In turn, tenured employees are expected to have greater fit. However, employees who do not fit may remain in the organization for a variety of reasons such as a lack of market opportunities and personal reasons. Furthermore, given increased competition in the marketplace, the changing nature of jobs, and increased technology, job

demands often change after an individual has joined the organization. Likewise, group cultures are likely to change as managers and employees transition in and out of the work environment. This implies that an individual's fit may decrease over time even if fit was initially high, and as such, continued attention should be devoted to fit after the hiring and socialization period. For example, training can be used to foster greater fit as job demands change, formal and informal activities (e.g., information sharing, social activities, mentoring) can aid in enhancing P-O fit, and group development interventions (e.g., continual feedback, interpersonal skills training, leader training) can further enhance fit to the group and group relations.

Finally, targeted interventions can be implemented to help employees deal with conflicts and problems that arise from diversity. Although some research shows that greater diversity on demographic variables is related to higher levels of creativity and performance for the group, greater diversity has also been related to lowered attitudes, more turnover, and poorer perceptions of group relations. To offset the potential negative impact on individuals' responses and to capitalize on the diversity of viewpoints, interventions such as training targeted at interpersonal relationships, communication, and understanding diversity may be warranted. As a result, harmonious group relations can be achieved more quickly, and people may perceive a greater degree of fit sooner, thereby enhancing both individual satisfaction and group effectiveness. Furthermore, over time and with sustained contact, the negative impact of diversity tends to dissipate, allowing positive group-level outcomes to emerge.

## 5. A POTENTIAL PARADOX OF FIT

The majority of research studies have consistently shown that a greater degree of fit is related to more positive outcomes for individuals (e.g., greater attraction to organizations, higher satisfaction and commitment, better adjustment, less turnover). The ASA cycle and socialization processes can create a situation whereby those who remain with an organization are a more homogenous group than are those who were initially attracted to and selected by the organization. However, several authors have argued that too much homogeneity within an organization can lead to stultification. When individuals have similar personalities, values, goals, and perspectives, new ideas and innovation may be stifled. The conformity in outlook that

arises from a similarity of perspectives does not allow for alternative interpretations and makes it difficult for the organization to adapt to environmental changes. Information about the environment is incomplete, and the organization is unable to make appropriate and necessary changes to adapt to the environment.

Therefore, there is a potential paradox in fit. More positive attitudes and increased satisfaction of employees in an organization is a worthwhile goal because satisfaction is related to the quality of life for individuals as well as to greater productivity for the organization. Yet focusing on creating fit for individuals can lead to too much homogeneity in the organization, and this can stifle innovation, creativity, flexibility, and the ability to adapt, ultimately endangering the organization's competitiveness and long-term viability. How does one reconcile these conflicting findings on the relationship between fit and outcomes at different levels?

### 5.1. Possible Solutions

Although it is a topic of debate among fit theorists and researchers, there are several ways in which this paradox might be resolved. First, achieving a high degree of P-O fit among middle- and lower level employees can promote greater satisfaction and other outcomes for the majority of individuals within the organization. Achieving more diversity in perspectives, values, and skills among key decision makers or the top management team can foster problem analyses and consideration of alternatives so as to adapt to the environment and implement strategic initiatives as the organization moves forward.

Second, attention can be given to achieving similarity for core attributes that are critical for the organization while allowing variation for less critical attributes. Those attributes for which there should be similarity and those that may differ will vary from organization to organization depending on the organization's purpose, mission, strategy, and industry. Homogeneity in terms of the organization's key goals or most fundamental values is important so that employees perceive and understand the same goals and behave accordingly to promote the effective operating of the organization. Yet employees can vary in their attributes and perspectives around other less core goals and values. In such cases, there will be sufficient similarity to promote more positive attitudes and effective functioning but also sufficient diversity to allow new ideas and innovation to emerge.

Third, when the organization desires to change in some way (e.g., culture change, move to new markets), selection should focus on hiring people who have specific attributes that fit with the new goals. Both complementary fit (i.e., attributes that add what is needed) and supplementary fit (i.e., similarity) can be achieved simultaneously if the new attributes that are needed to effect the change are delineated and individuals with those attributes are selected while still ensuring that these individuals are similar to current employees on other attributes.

## 6. CONCLUSION

Insights from fit theory and research can be used to enhance organizational life at all levels: At the individual level, it can help to understand how to increase satisfaction, commitment to the organization, and task and contextual performance as well as how to raise individuals' quality of life at work by decreasing stress and anxiety. At the group level, fit can help to identify ideal team compositions based on both supplementary fit and complementary fit so as to enhance individual satisfaction as well as to promote harmonious group relations and effective functioning. Furthermore, fit can contribute to a better understanding of the processes and impact of demographic diversity, which in turn can lead to better, more effective diversity practices in organizations. Finally, at the organizational level, fit can be helpful in thinking about issues such as building a distinct culture and promoting change. It can also assist the organization in addressing issues such as turnover, absenteeism, and low productivity. In addition, P-O fit implies that organizations should shift away from the traditional focus on P-J fit in their

selection processes in favor of practices geared toward hiring "whole persons."

## See Also the Following Articles

Executive Development and Coaching ■ Organizational Culture and Climate ■ Personnel Selection ■ Recruitment ■ Vocational Psychology, Overview ■ Work Teams

## Further Reading

- Bowen, D. E., Ledford, G. E., & Nathan, B. R. (1991). Hiring for the organization, not the job. *Academy of Management Executive*, 5, 35-51.
- Chatman, J. A. (1991). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, 36, 459-484.
- Edwards, J. R. (1991). Person-job fit: A conceptual integration, literature review, and methodological critique. *International Review of Industrial and Organizational Psychology*, 6, 283-357.
- Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, 49, 1-49.
- Pervin, L. A. (1989). Persons, situations, interactions: The history of a controversy and a discussion of theoretical models. *Academy of Management Review*, 14, 350-360.
- Schneider, B., Kristof-Brown, A., Goldstein, H. W., & Smith, D. B. (1997). What is this thing called fit? In N. Anderson, & P. Herriot (Eds.), *International handbook of selection and assessment* (pp. 393-412). Chichester, UK: Wiley.
- Tinsley, H. E. (2000). The congruence myth: An analysis of the efficacy of the person-environment fit model. *Journal of Vocational Behavior*, 56, 147-179.
- Werbelt, J. D., & Gilliland, S. W. (1999). Person-environment fit in the selection process. *Research in Personnel and Human Resources Management*, 17, 209-243.



# Personnel Psychology

David Chan

National University of Singapore, Singapore

1. Recruitment
2. Selection
3. Training
4. Performance Appraisal and Feedback
5. Compensation
6. Current Themes in Personnel Psychology  
Further Reading

Personnel psychology is a subfield within industrial, work, and organizational psychology that is concerned with the application of psychological theories, methods, and research findings to the personnel functions in organizations. Primary functions include recruitment, selection, training, performance appraisal and feedback, and compensation. This article focuses on the research applications to these functions and outlines the current themes in personnel psychology.

## GLOSSARY

**compensation** The entire process of designing and administering the package of returns that an organization provides to an employee in exchange for his or her work input.

**performance appraisal** The evaluation associated with a rater's evaluative judgment of the effectiveness of an employee's job performance.

**performance feedback** The information an employee receives concerning a rater's evaluative judgment of the employee's performance and the way in which the information is communicated.

**recruitment** The organizational activities and processes aimed at providing information to a pool of potential candidates to attract them to apply for job openings in the organization.

**selection** The process of testing and evaluating job applicants for the purpose of determining the subset of applicants to whom job offers will be made.

**training** The systematic and typically formal process of employee acquisition or learning of job competencies—knowledge, skills, abilities, and attitudes—that would enhance job performance.

## 1. RECRUITMENT

Recruitment refers to the organizational activities and processes aimed at providing information to a pool of potential candidates to attract them to apply for job openings in an organization. In the development of recruitment strategies, personnel practitioners must address important factors such as features of the relevant labor market for the job, the recruitment message and information to be communicated, and the specific recruitment methods to be used for the communication. The bulk of the recruitment research in personnel psychology is concerned with candidate perceptions of recruitment processes. These perceptions are of practical importance because candidates with negative perceptions may decide to withdraw from the application and selection process. Findings from previous studies suggest that although job attributes constitute the most important factor in affecting candidates' reactions and job choice behaviors, candidate perceptions of recruitment activities

and the recruiter may also affect reactions and associated behavioral intentions such as intent to apply for employment in the organization or to recommend it to others. There is some evidence that candidates tend to make inferences about the organization (e.g., leadership style) by extrapolating from perceived recruiter behaviors.

Another interest among personnel psychologists doing recruitment research is in the area of realistic job previews. The purpose of realistic job previews is to provide applicants with “true” information about the job and work environment (i.e., as the job incumbents actually experience them) rather than presenting the job and organization in “public relations” terms to maximize job and organizational attractiveness. The rationale is that the failure to present realistic job previews will lead to unrealistic, and hence subsequently unmet, expectations among newcomers (i.e., applicants who accepted the job offer) that, in turn, will have a negative impact on satisfaction, commitment, and motivation.

## 2. SELECTION

Personnel selection refers to the process of testing and evaluating job applicants for the purpose of determining the subset of applicants to whom job offers will be made. The majority of personnel selection research is concerned with the extent to which different selection predictors (i.e., instruments) can predict job performance and other job-relevant outcomes such as job satisfaction, organizational commitment, and intent to quit. The selection predictors that have been widely researched and used in practice include cognitive ability tests, personality measures, biodata measures, interviews, and assessment centers. Recently, there has been increasing interest in the research and use of situational judgment tests, which is one type of work simulation. Comprehensive reviews of the validity evidence of selection predictors in predicting job performance and other job-relevant outcomes were provided by Salgado and colleagues in 2001 and by Schmitt and Chan in 1998. The search for valid selection predictors needs to be driven by the nature of the criterion of interest. For example, with respect to job performance, the criterion space can be decomposed into specific areas of task requirements and specific knowledge/skill/ability requirements. In addition, the job performance criterion can be decomposed into distinct performance constructs such as typical versus maximum performance, task versus contextual performance, and routine versus adaptive performance. Each specific

performance construct can, in turn, be construed in terms of multiple dimensions; for example, contextual performance can be construed in terms of interpersonal facilitation and job dedication. The validity of a selection predictor is dependent on the specific performance criterion variable in question.

In making selection decisions, it is possible to use more than one predictor, that is, to use a battery of selection tests. One major issue that needs to be addressed when using selection predictors in a test battery concerns incremental validity, which refers to how much gain in prediction power is obtained by adding the new predictor to the existing predictor(s). Issues of incremental validity have focused on effects of predictor–criterion correlations and intercorrelations among predictors. All other things equal, incremental validity is increased when the new predictor is highly correlated with the criterion but is lowly correlated with the existing predictor(s). Another major issue concerns the joint effects of the combined use of predictors on subgroup differences in overall selection score. Research has shown that adding a new predictor with a small subgroup difference (or no difference) to an existing predictor with a large subgroup difference is unlikely to lead to any substantial reduction in subgroup difference on the composite selection score.

Special topics in personnel selection include test bias (e.g., whether a test predicts similarly for applicants from different demographic groups), applicant reactions (e.g., what characteristics of the selection system would influence the degree to which applicants perceive the selection as fair), selection for staffing teams (e.g., what competencies can be tested to predict how applicants would function when working in teams), and validity changes associated with dynamic criteria (e.g., how the validity of a selection test is affected when the components of effective job performance change over time). Many of these topics were summarized by Guion in 1998 and by Schmitt and Chan in 1998.

## 3. TRAINING

Training refers to the systematic and typically formal process of employee acquisition or learning of job competencies—knowledge, skills, abilities, and attitudes—that would enhance job performance. The training environment may range from artificial or contrived settings (e.g., classroom lectures, role-play exercises) to highly realistic settings (e.g., on-the-job training in which the learning environment is nearly identical to

the actual work environment). Regardless of the setting, the underlying assumption is that job competencies are malleable and alterable through training. The goal of training research is to identify appropriate interventions that can positively affect employees' target attitudes and behaviors so that the employees can meet the demands of the target job. Training research typically follows the framework provided by the instructional systems design (ISD) model that emphasizes three key interrelated components of the training process: training needs assessment, training design, and training evaluation. In training needs assessment, the primary goal is to provide information on what, where, when, and who. The information gathered also serves as input for the subsequent stages of the training process, that is, training design and evaluation. The strategies to gather information may include organizational analysis to determine where training is needed; job analysis to identify the nature of the job tasks, the underlying job competencies, and the performance standards required for successful task performance; and person analysis to determine performance gaps between potential trainees' performance and expected performance levels. Training design is concerned with specifying instructional objectives, sequencing training materials, incorporating learning principles, and identifying effective training methods. There is an increasing consensus among training researchers that a training program is more effective when trainees are given the opportunity to produce the capability, when the training environment enhances trainee self-efficacy and expectations that the training will be effective in leading to acquisition of valued learning outcomes, and when performance feedback given to trainees is credible, accurate, constructive, and timely. Training researchers also appear to agree that the effectiveness of training methods is a function of the properties of the method, the nature of the training content, and the person characteristics of the trainees. For example, declarative knowledge is better acquired through classroom instruction, whereas procedural knowledge is better acquired through work simulations. However, more empirical research is needed, and advances in training research are likely to be led by rigorous studies that establish empirical interactions linking training method, training content, and trainee characteristics. Training evaluation refers to the assessment of the degree to which trainees achieved the anticipated gains in job competencies from training and effectively applied these gains to the actual job. To measure these gains and conduct an adequate training evaluation, training objectives and the intended training

outcomes must be explicitly specified in ways that allow their valid measurement. An excellent overview of the typical problems in evaluating training and the alternative ways in which to assess gains was provided by Goldstein and Ford in 2002.

The ISD model, which is based on design principles derived from research findings on performance on simple tasks that require little complex cognitive processing, is somewhat deficient as a theoretical framework for training in jobs that are made up of ill-defined problems and cognitively complex tasks. Recently, several researchers have proposed more holistic approaches. These approaches involve applying cognitive constructs and principles to the training process and recognizing the multifaceted nature of the learning process by expanding the criterion space to include cognitive, skill-based, and affective learning outcomes. This cognitive-oriented and holistic context, grounded in the theory and empirical findings from cognitive psychology research, led to significant conceptual and measurement advances in training research. Many of these advances are related to the conceptualization of expertise and the measurement of metacognitive skills and knowledge structures.

#### 4. PERFORMANCE APPRAISAL AND FEEDBACK

Performance appraisal refers to the evaluation associated with a rater's evaluative judgment of the effectiveness of an employee's job performance. Performance feedback refers to the information an employee receives concerning a rater's evaluative judgment of the employee's performance and the way in which the information is communicated.

In performance appraisal, the performance measure can be either objective or subjective. Examples of objective measures include dollars sold and number of units produced or contracts secured. Some objective measures, such as number of complaints received and days absent from work, are negatively scored. Examples of subjective measures include qualitative judgment of the creativity of work products and supervisor or peer ratings of overall job performance. The vast majority of early research on performance appraisal has focused on subjective measures and examined issues such as the various types of appraisal rating scales, rater accuracy, rater errors, and the cognitive processes underlying appraisal judgments made by raters. Despite the rater training implications of these findings, there is no clear evidence that rater

training based on these streams of research has in fact contributed to increased effectiveness of actual performance appraisal systems in organizations. One explanation for the apparent futility of early studies is that the researchers' focus on rating scales, rater accuracy and errors, and rater cognitive processes is somewhat misplaced because the researchers have neglected the central elements of the appraisal system, including the rater's motivational context and the employee's (i.e., ratee's) perspectives and reactions. This gap was filled by several subsequent streams of research focusing on motivational structures and rater–ratee relationships such as effects of rating purpose, accountability, and rater–ratee similarity on ratings and effects of organizational climate, perceived organizational support, and leader–member exchange relationships on appraisal reactions. One productive stream of research came from Greenberg's application of organizational justice research to performance appraisal in 1986. Studies by Greenberg and other researchers have identified a variety of aspects of the appraisal process and outcome (e.g., soliciting and using input prior to appraisal, ability to challenge appraisal) that affect justice perceptions about the appraisal. Most attention has been directed toward employee participation or voice in the appraisal and feedback process. In addition to identifying the antecedents of justice perceptions, researchers have examined the influence of justice perceptions on a number of attitudinal and behavioral outcomes (e.g., organizational commitment). Current research has adopted a broader perspective by locating the appraisal process in the larger context of the performance management system in an organization. This system refers to the interrelationships linking appraisal system development, appraisal processes, and feedback processes.

There has been increasing interest in "360-degree" assessment/feedback systems, where the central idea is to obtain appraisals on the same employee from multiple sources (typically supervisors, peers, and subordinates). Such systems have unique features involving multiple sources and perspectives that raise important research and practical questions regarding the impact of expanding the performance construct space on appraisal outcome and reactions. However, insofar as these systems consist of multiple rater–ratee relationships, many of the cognitive and motivational elements present in conventional appraisal are likely to remain applicable. For example, prior research findings on conventional appraisal that found that employees are more likely to react favorably to appraisal feedback when there is prior discussion on performance

expectations at the beginning of the performance period, opportunity for employees to provide input to the appraisal and discuss raters' rationales for the appraisal judgments, and consideration of employees' feelings.

Performance appraisal and performance feedback have the ultimate common goal of increasing job performance effectiveness. Important conceptual and practical relationships exist between performance appraisal and performance feedback. For example, accountability effects in performance appraisal are likely to be related to the type and specificity of performance feedback information, and justice perception effects in performance feedback are likely to be related to the accuracy and transparency of the performance appraisal process. Research findings with more theoretical value and higher external validity are likely to be obtained from studies that integrate performance appraisal and performance feedback.

## 5. COMPENSATION

Compensation refers to the entire process of designing and administering the package of returns that an organization provides to an employee in exchange for his or her work input. The package, which may include financial remuneration, services, and medical and other benefits, is typically divided into direct pay and indirect pay/benefits. Direct pay may consist of the fixed base pay, base pay supplements (e.g., market value adjustments, cost-of-living allowances, overtime pay), and performance-related pay (e.g., merit increments, performance bonuses, performance commissions). Indirect pay and benefits may consist of any services or benefits provided to the employee (e.g., vacations, medical benefits, discounts for goods and services).

In human resources management, most attention on compensation design and administration is focused on the strategic mix of the various components of returns to align the compensation system with the resources, constraints, priorities, and strategic goals of an organization within the labor market context of the need to attract and retain suitable employees. This typically revolves around issues of decisions concerning the value or grade level of the various jobs in the organization, the level of compensation relative to that of external referent organizations (which are typically competitors for the same employees), the design of the compensation package for individual employees, and the administration of the package. Compensation research in personnel psychology, on the other hand, is more concerned with the psychological relationships at

the individual level (although it may also involve team, organizational, or higher levels of analysis) linking compensation to employee performance and reactions. Specifically, these relationships are concerned with the motivations, perceptions, and attitudes of employees who are affected by the various types of compensation (and ways in which it is administered) and how these motivations, perceptions, and attitudes, in turn, affect employee performance and reactions.

Compared with issues of selection, training, and appraisal, compensation issues have largely been ignored by personnel psychologists, despite the fact that the idea of linking rewards to job performance has been central in empirically established theories of work motivation such as expectancy theory, goal setting theory, and equity and justice theories. Although compensation typically refers to a contractual (legal or semilegal) or formal package of returns, rewards is a more generic term used to refer to any positive reinforcement or inducement provided (explicitly or implicitly) by an organization to employees that is assumed to positively correlate with past, current, and/or future performance. During recent years, the rapidly changing nature of work has led to an increased use of new pay systems by employers. This has, in turn, activated research interest among some personnel psychologists. Not surprisingly, the majority of the recent personnel psychology research on compensation is on new forms of pay systems and has found considerable evidence that some of these new pay systems, such as variable pay and competency- and skill-based pay, are related to improved employee and organizational performance. However, it remains unclear why or how these new forms of pay lead to increased performance, although a number of conceptual models based on various psychological theories (e.g., social exchange theory, theory of planned behavior) have been proposed to account for the observed relationships. Instead of applying grand theories to commonalities across various new and effective pay systems, midrange theories that relate the specific form of pay system to the pay context, employee characteristics, and employee reactions are more likely to enhance our understanding and provide practical recommendations for specific compensation contexts.

## 6. CURRENT THEMES IN PERSONNEL PSYCHOLOGY

There are several discernible themes in current research in personnel psychology across the areas of recruitment,

selection, training, appraisal, feedback, and compensation. One theme concerns the importance of focusing on the multidimensional and dynamic nature of job performance. In early research, job performance was construed nearly solely in terms of core technical proficiency or task performance. Current research has expanded the performance construct space to include organizational citizenship behaviors, counterproductive behaviors, and adaptive behaviors. This expansion has implications for the various areas of personnel psychology (e.g., identifying and selecting candidates with dispositions toward citizenship behaviors and against counterproductive behaviors, appraising employees on these additional performance dimensions, designing training that enhances the ability to adapt to changing situations). Current research has also focused much attention on the dynamic nature of job performance with respect to changes in performance level over time and changes in performance dimensionality over time. These changes over time raise complex but important conceptual and measurement questions that provide a springboard to productive future research. Another theme concerns person–environment fit, where environment may refer to the job, the work group, or the organization. The basic idea is that maximizing person–environment fit often (but not always) leads to positive outcomes in recruitment, selection, training, appraisal, feedback, and compensation. Research in person–environment fit brings with it several challenging methodological and data-analytic issues such as the need to distinguish between subjective fit perceptions and objective fit assessments. Finally, running across virtually all current research in personnel psychology is the central theme that the changing nature of work has major implications for the conceptualizations and measurement of the various focal constructs as well as the relationships (conceptual and empirical) between constructs.

### See Also the Following Articles

Employment Interviewing ■ Industrial/Organizational Psychology, Overview ■ Occupational Psychology, Overview ■ Personal Initiative and Innovation ■ Person–Environment Fit ■ Personnel Selection ■ Recruitment

### Further Reading

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal means and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, 1, 421–483.

- Chan, D. (2000). Understanding adaptation to changes in the work environment: Integrating individual difference and learning perspectives. *Research in Personnel and Human Resources Management*, 18, 1–42.
- Gilliland, S. W., & Chan, D. (2001). Justice in organizations: Theory, methods, and applications. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 2, pp. 143–165). Thousand Oaks, CA: Sage.
- Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth.
- Greenberg, J. (1986). Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology*, 71, 340–342.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum.
- Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods, and techniques. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 1, pp. 165–199). Thousand Oaks, CA: Sage.
- Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.



# Personnel Selection

Deniz S. Ones

University of Minnesota, Minneapolis, Minnesota, USA

Chockalingam Viswesvaran

Florida International University, Miami, Florida, USA

1. Introduction
  2. Predictors Used in Personnel Selection
  3. Fairness in Selection
  4. Combining Predictors
  5. Emerging Issues in Selection
  6. Conclusion
- Further Reading

Personnel selection is concerned with identifying personal characteristics of applicants that relate to performance on jobs, measuring applicants on these characteristics, and hiring those who have the required characteristics.

## GLOSSARY

**adverse impact** Disproportionate hiring of different groups; in the United States and the United Kingdom, adverse impact is said to occur when the selection ratio for a legally protected group is less than four-fifths that of the majority group.

**criteria** Measures of how well individuals behave or perform; these are the behaviors, performances, and outcomes that the organization is interested in predicting using a selection system.

**criterion-related validity** The association between the predictor and criterion scores; the criterion-related validity coefficient indicates the extent to which individual differences in a particular predictor are related to individual differences in a criterion (e.g., job performance).

**predictive bias** The systematic over- or underprediction of a criterion.

**predictor** Any measured applicant trait or characteristic that is used to predict performance and other behaviors on the job; a predictor assesses applicant characteristics deemed to be job relevant using a specific measurement method (e.g., test, application blank, interview).

**selection ratio** Proportion of job applicants hired to those who apply for a job.

## 1. INTRODUCTION

Personnel selection has been, and continues to be, one of the central functions of industrial, work, and organizational (IWO) psychology. Selecting the right people for the right jobs constitutes a source of competitive advantage for organizations. Individuals differ in various characteristics such as skills, abilities, personality traits, and interests. Similarly, jobs differ in what is required of the incumbents. The objective in personnel selection is to identify the personal characteristics that relate to performance on jobs, measure applicants on these characteristics, and hire those who have the required characteristics. When done properly, selection can enhance organizational productivity as well as provide individual benefits to workers themselves (e.g., increased satisfaction).

In personnel selection, one assesses the applicants on characteristics relevant to performance on the job and may want to select those who are likely to perform the best on the job. There are several characteristics on which individuals differ. Abilities, skills, personality traits, and physical abilities are some examples. These individual differences, which are relevant to

performance on the job under consideration, must be measured reliably and accurately before hiring decisions can be made. The method of such measurement could be an interview, a test, biographical information, or even physiological measurement.

A predictor can be any measured applicant trait or characteristic that is used to predict performance and other behaviors on the job. That is, a predictor assesses applicant characteristics deemed to be job relevant using a specific measurement method (e.g., test, application blank, interview). This article provides an overview of several commonly used and emerging predictors used for personnel selection.

Given several predictors, how should one choose the most desirable or optimal one for personnel selection in a particular situation? The answer to this question depends on the behaviors one wishes to predict. Such work behaviors can include both positive behaviors (e.g., job performance, task performance, teamwork) and negative behaviors (e.g., absenteeism, employee theft). Any work behavior that one is interested in predicting is referred to as the criterion.

Predictors assess variables on which individuals differ, and such differences are relevant only to the extent that they reflect individual differences in relevant criteria of interest (e.g., job performance, other work behaviors). Whether or not a predictor is a good prehire indicator of work behaviors that one is interested in can be determined by examining the relationship between a given predictor and work behaviors of interest. Criterion-related validity refers to the association between the predictor and criterion scores. The criterion-related validity coefficient indicates the extent to which individual differences in a particular predictor are related to individual differences in a criterion (e.g., job performance). Thus, the criterion-related validity coefficient can be used to evaluate the job relevance of a predictor for selection. Once the criterion one wishes to predict is identified, one can choose the predictor or combination of predictors with a high criterion-related validity coefficient for that criterion.

In addition to criterion-related validity, one may want to consider other factors that could affect the utility of the selection procedures. The utility of a selection system depends on the selection ratio and variability in job performance across individuals. The selection ratio is the percentage of applicants who are hired and can range from 0 to 1.0. When an organization hires all applicants (selection ratio = 1.0), the usefulness of a predictor is nonexistent because no "selection" takes place. The smaller the selection ratio, the greater the benefits obtained

from personnel selection. When there is no variability in job performance across individuals, the usefulness of a selection system is also nonexistent; it does not matter who gets selected. Thus, the exact utility (i.e., value and organizational benefits) of a predictor for personnel selection depends on the context in which it is used.

Another important consideration in evaluating the various predictors used for personnel selection is applicant reactions. For example, managerial applicants might not accept that scores on a verbal ability test reflect individual differences in managerial job performance. Undoubtedly, there are also cross-national differences. For example, personnel selection in the North American context is more reflective of the organizational perspective in that research and applications have emphasized the predictive validity of selection tools. In contrast, European research not only has been more eclectic in its assessment of the predictive validity of selection tools but also has taken a multilateral approach to evaluation, including evaluations from multiple stakeholders. It is more common in the European context to view selection as a negotiation among various parties such as the organization, unions or employee associations, and governmental agencies. Perhaps this broader focus is reflected in academic discussions of social constructivism in inferring meaning of organizational and individual acts. The study of applicant reactions, although in its incipient stages, has the potential to bridge this gap between the organizational perspective and multiple stakeholder perspectives.

Applicant reactions have important consequences for an organization. If top applicants turn down job offers, the utility of the selection effort is reduced. Furthermore, dissatisfied applicants are more likely to litigate against an organization and boycott its products. This article reviews research on the variables that affect applicant reactions and discusses different predictors used in personnel selection, both in terms of their criterion-related validity and in terms of applicant reactions.

Selection affects not only organizations and the individuals applying for jobs but also society. Society has a stake in promoting fair selection procedures, maximizing productivity, and ensuring opportunity to all individuals. The issues become complicated when there are discernible differences across demographic groups in predictor scores. Group differences result in greater proportions in hiring of the higher scoring group, and this could result in an adverse impact for the lower scoring group. Therefore, in discussing the different predictors used in personnel selection, this article also

reviews issues of fairness, bias, adverse impact, and group differences in predictor scores. Attempts made to reduce the impact of group differences (e.g., banding procedures) are also summarized. Finally, methods of using multiple measures in selection and research on emerging topics in personnel selection are presented.

## 2. PREDICTORS USED IN PERSONNEL SELECTION

Selection occurs after a pool of qualified applicants has been identified. Recruitment refers to the activities undertaken to attract a qualified pool of applicants, and several factors affect the effectiveness of recruitment. The content of the recruitment message, the characteristics of the recruiter, the medium used, and the like have been researched. Although this article does not elaborate on these topics, it is important to note that some prescreening (including self-selection) will have occurred in the normal flow of events in organizations prior to application of the selection methods discussed in what follows.

Several predictors are used in personnel selection. Some predictors assess whether individuals have the knowledge, skills, and abilities to perform a job task (i.e., “can do” predictors), whereas others assess whether individuals will typically perform well on the job (i.e., “will do” predictors primarily assessing motivation). In classifying predictors, it is possible to view them as determinants of declarative knowledge (i.e., knowledge of facts) needed to do the job (e.g., job knowledge tests), procedural knowledge (i.e., knowledge of how to get the job done), and motivation. A multitude of predictors have been used in personnel selection. This section

provides a summary of some of the most popular predictors used in personnel selection. It organizes the predictors into three main categories: (a) predictors based on individual differences (cognitive ability, physical ability, personality, and biographical information); (b) method-based predictors that can be used to assess various knowledge, ability, and trait combinations (job simulations and work samples, assessment centers, and interviews); and (c) predictors used primarily in the prediction of undesirable work behaviors (integrity tests and drug testing).

The coverage in this article is not comprehensive given space constraints, but it is intended to familiarize the reader with some of the popular selection tools. In fact, at times, popular predictors might not be the most valid ones. Table I summarizes the use and validity of the predictors discussed in this section. The data on predictor use vary across surveys; therefore, the table indicates a general categorization of use as high, medium, and low. The validity data are culled from meta-analyses and are also summarized into broad categories (high, medium, and low). The data in the table are suggestive and not definite in any sense because averaging percentages across surveys, where the numbers of respondents differ, is not indicative of any specific trend. The table is provided only as a heuristic to enable the reader to appreciate that some organizations may be employing less valid methods and incurring large utility losses. In a way, the table suggests that society may be incurring forgone opportunity costs to the extent that popularity and validity diverge.

### 2.1. Cognitive Ability Tests

There are more than a dozen definitions of what constitutes cognitive ability. Common threads across most

TABLE I  
Summary Status of Personnel Selection Predictors

Predictor	Criterion-related validity	Extent of use	Applicant reactions
Cognitive ability	High	Medium	Moderate
Physical ability	Medium	Low	Positive
Personality tests	Medium	Medium (but increasing)	Moderate
Biodata	Can be high	High	Moderate
Job simulations	High	Low	High
Assessment centers	Can be high	Low	High
Interviews	Can be high	High	Moderate
Integrity tests	High	Low (except in the United States)	Moderate
Drug testing	Medium	Low (except in the United States)	Negative

of the definitions are the capacity to process information and learn. IWO psychologists have been using cognitive ability tests for more than 80 years. Furthermore, cognitive ability is one of the most widely researched traits in personnel selection. IWO psychologists have developed several tests of cognitive ability, and meta-analytic cumulation of the vast bodies of criterion-related validity studies have shown that cognitive ability is one of the best predictors of job performance. The relationship between cognitive ability test scores and job performance averages approximately .50. This is higher than what is found for other predictors. Furthermore, the validity of cognitive ability tests has been demonstrated in nearly all situations. This predictor is perhaps the best and most applicable one in personnel selection for nearly all jobs. All jobs require some cognitive ability to learn the jobs and perform effectively. The predictive validity of cognitive ability tests tends to be higher for high-complexity jobs (.58) than for low-complexity jobs (.23). A high-complexity job (e.g., nuclear physicist) involves more processing of information than does a low-complexity job (e.g., truck driver).

In addition to tests developed to assess general cognitive ability, there are ability tests tailored for specific purposes such as assessment of accounting ability and knowledge of mathematics. Tests have also been developed to assess perceptual ability and knowledge of mechanical principles. All of these tests measure general cognitive ability (*g*) to some extent, but they also have a specific ability component. It has been hypothesized that these specific abilities could be more predictive for specific jobs. For example, it has been argued that perceptual ability may be more important in clerical tasks. However, empirical research suggests that the predictive validity of specific abilities stems from their assessment of a general information-processing ability (i.e., the general cognitive ability). If the general cognitive ability is partialled out (i.e., held constant), most of the predictive power of specific ability tests vanishes. However, specific ability tests may have more face validity and may generate positive applicant reactions in a selection context.

## 2.2. Physical Ability Tests

Physical ability tests have also been used in personnel selection, especially in certain jobs such as firefighting and police work. Although there is a vast empirical literature on cognitive ability, the research on physical ability in selection contexts is more limited. Researchers have identified several dimensions of physical ability. One

classification conceptualizes physical ability as being composed of nine dimensions: static strength, explosive strength, dynamic strength, trunk strength, extent flexibility, dynamic flexibility, gross body coordination, gross body equilibrium, and stamina. Tests have been devised for these nine dimensions based on common physical tasks. Other researchers have categorized physical ability assessments into two major dimensions: endurance and maximum strength measures. The validity and usefulness of physical ability as a predictor of job performance has been empirically investigated. One study presented the validity of physical ability tests in personnel selection as .32 for predicting supervisory ratings.

## 2.3. Personality Tests

IWO psychologists have used personality variables for personnel selection since the beginning of the 20th century. Walter Dill Scott devised measures for selection of salespersons that assessed aspects of personality. The Woodworth personality datasheet was used to select or identify individuals who are likely to suffer from war neurosis. During early days, IWO psychologists hypothesized and assessed different variables that they considered to be promising. Variables such as reliability, dependability, sociability, and ego strength were assessed. Sometimes different names were given to the same trait, and at other times the same label was used to assess different traits. Given this state of affairs, the validity and usefulness of personality variables in personnel selection were questioned during the 1960s. Furthermore, the behaviorist revolution in psychology contributed to personality variables falling out of favor during this time.

Research conducted over the past decade has been more positive. Meta-analysis was used to cumulate the validity and usefulness of personality factors for personnel selection across studies. Results suggested that the dimension of conscientiousness is valid across jobs and situations, with a criterion-related validity of .23. Openness to experience was found to be correlated with the criterion of training performance, and extraversion was found to be correlated with performance in certain jobs such as sales. Emotional stability was found to be a good predictor of performance in security-related jobs (e.g., police work). Salgado summarized the validity evidence based on European samples and found that conscientiousness and emotional stability are useful predictors in Europe as well. Newly emerging research suggests that emotional stability may also be an important predictor in the United States. Although concerns have been raised that applicants could fake

their responses, the criterion-related validity is substantial for personality inventories despite the presence of any potential response distortion among job applicants.

Assessments of personality have employed structured paper-and-pencil inventories. The Hogan Personality Inventory (HPI), the California Psychological Inventory (CPI), the NEO Personality Inventory–Revised (NEO PI-R), and the Personal Characteristics Inventory (PCI) are some of the paper-and-pencil inventories. The Minnesota Multiphasic Personality Inventory (MMPI), although developed for clinical assessments of psychopathology, has also been used in personnel selection, especially in jobs where emotional stability is important for job performance (e.g., police work, firefighting).

#### **2.4. Biographical Information (Biodata)**

The use of application blanks is prevalent in personnel selection. The underlying principle in the use of application blanks is that past behavior is the best predictor of future performance. However, the question of what weights to assign to each response has been much debated. Weights reflect the relative importance assigned to each response. In general, larger weights should be assigned to responses that characterize individuals who perform better on the job. For example, if individuals with bachelor's degrees are likely to be better performers than those with high school diplomas, those applicants who check the bachelor's degree response in the application blank should receive higher scores than should those who check the high school diploma response. This is an example of how empirical scoring is done. Scoring of biodata items can be done either empirically or based on rational grounds.

In a weighted application blank (WAB), weights are assigned to several items and the scores are summed. Once a WAB has been developed, a score can be computed for each prospective applicant by weighting the response provided by that particular applicant. Sometimes a biographical information blank (BIB) could be employed. A BIB is similar to a WAB, but the former includes more experiential and personal information. An accomplishment record is a structured inventory of experiences that an individual considers to be a reflection of his or her talents.

Biodata items can be used to infer several individual differences characteristics. Decision makers can use education-related biodata items to make inferences about math and language abilities, sports-related items

to make inferences about physical ability, and items on social positions held to make inferences about personality characteristics. Research indicates that biodata instruments that are transportable across organizations can be developed. That is, scores developed in one organization can be used in another organization, so it is not essential to develop biodata items specifically for each organization. Validities of .48 have been reported for such biodata instruments. In general, applicant reactions have been favorable, especially when items have focused on events under the control of individuals. Applicant reactions are more negative when invasive items that raise privacy concerns for applicants are included on biodata measures.

#### **2.5. Job Simulations and Work Samples**

One of the major criticisms of the paper-and-pencil assessment method is that it does not capture the rich, complex, and subtle ways in which a trait is manifested in the real world, especially as performance on the job. Researchers have attempted to build realism into the assessment methods by simulating as much as possible the actual conditions under which individuals have to perform their work and testing how individuals perform in those simulations. Simulations differ in the degrees of realism captured. At one extreme are simulations that have very high psychological and physical fidelity to reality. At the other extreme are assessment methods in which the individuals are presented with paper descriptions of actual situations and asked to indicate how they would respond if they encountered the situations described.

Situational judgment tests are simulations of low fidelity. As the name indicates, individuals are presented with descriptions of situations and asked to indicate their likely responses. Situational judgment tests are typically high on face validity and have been found to display high levels of predictive validity. However, it has been difficult to ascertain the traits assessed in such simulations. Work sample tests have been employed to assess both job knowledge and procedural knowledge. Work sample tests have demonstrated high criterion-related validity (averaging .54). Work sample tests are relatively costly to develop. Also, in some instances where specialized knowledge is required or where tasks pose a danger to individuals assessed or to others, it is not possible to use work sample tests in personnel selection. Applicant reactions

tend to be favorable toward work samples, probably due to high face validity.

### **2.6. Assessment Centers**

Related to simulations, but more comprehensive in coverage of the traits assessed, are assessment centers. Assessment centers typically involve multiple assessments of applicants using various tools and methods, including tests, interviews, and several types of exercises such as leaderless group discussion, in-basket exercises, and other problem-solving approaches. Assessment centers may also include role-playing and/or oral presentations. As such, the assessment center method of job candidate evaluation can span several days. A meta-analysis of the assessment center literature has found that the average validity of assessment center studies reported in the literature is .37 for predicting overall job performance. Recent research indicates that much of this validity is derived from traits such as cognitive ability and personality characteristics assessed. The use of assessment centers has also been found to elicit positive applicant reactions due to the apparent linkages between exercises and managerial work functions. Job applicants, especially for higher complexity jobs, consider assessment centers to be more acceptable predictors of job performance.

### **2.7. Interviews**

Interviews are by far the most widely recognized personnel selection tools used. Interviews that are structured (i.e., the same set of questions and standardized scoring system is used for all applicants) have been found to have better validity than unstructured interviews. Another issue in the research on interviews is the use of multiple interviewers or panel interviews. Researchers have explored the composition of the interview panel in terms of demographic characteristics. The demographic characteristics (e.g., race, gender) of the interviewers and interviewees, and potential interactions between the two, have also been investigated. Typically, demographic variables do not influence the interview outcomes significantly.

During recent years, IWO psychologists have introduced two new types of interviews: the situational interview and the patterned behavior description interview. The basic rationale for the two types of interviews is that the questions asked of individuals should be reflective of actual situations that people face in the real world (where one wants to predict performance). Thus,

individuals are given examples of workplace situations and asked how they would respond if confronted with such situations. Validities of approximately .45 have been reported for structured interviews.

### **2.8. Integrity Tests**

Employers have attempted to screen out applicants who are likely to engage in theft or related counterproductive behaviors. This is especially true in some industries such as retail, fast food, and banking. Physiological measures have been used to assess dependability. Since 1988, the use of polygraphs in preemployment testing has been limited to security-related jobs in the United States. Furthermore, most psychologists believe that the validity of polygraphs is doubtful. Another option is the use of paper-and-pencil tests designed to assess integrity of applicants. Some of these tests ask applicants about their thoughts about theft, their admissions of theft, and their estimations of theft prevalence rates. The rationale is that individuals who are likely to engage in theft are likely to rationalize their behaviors by reasoning that most people will engage in such behaviors. Thus, they are likely to overestimate the prevalence rates. The emphasis is on assessing applicants' attitudes toward theft. However, recent research indicates unequivocally that integrity tests assess primarily well-known personality traits of conscientiousness, agreeableness, and emotional stability. As for usefulness in prediction of work-related behaviors, meta-analytic findings suggest that integrity tests have a validity of .39 in predicting general counterproductive behaviors. The criterion-related validity for predicting actual detected theft is lower (.13), and the validity for predicting overall job performance is .41.

### **2.9. Drug Testing**

Screening for substance abuse has increased over the years. Integrity tests can be used to screen applicants for potential theft and for counterproductive behaviors on the job, including substance abuse. However, many employers are skeptical that responses to a few items on a paper-and-pencil test will identify substance abuse. In fact, such personality assessment may be more useful for detecting proclivity for substance abuse, whereas physiological measures may be more applicable for detecting current use. Urinalysis, blood samples, and hair testing are some of the physiological measures that have been employed in testing for substance abuse. Many employers provide job offers to applicants

conditional on passing a drug-screening test. Some studies report substantial predictive validities for preemployment drug screening on a variety of job performance criteria. However, drug testing cannot distinguish between the casual or weekend user and the workplace abuser. A person who smokes one marijuana “joint” on Saturday night will have the same result as a person who is smoking marijuana continually on the job. This inability to discriminate is a major point of controversy in the drug-testing literature. Arguments have been made that personality tests, such as integrity tests, can assess the proclivity to use drugs and so can make the distinction between a casual user and a regular user.

### 3. FAIRNESS IN SELECTION

In discussing fairness in selection, one needs to clearly distinguish among culture-fair assessment, group differences, adverse impact, discrimination, bias, and fairness. These are related but distinct concepts. Any assessment is grounded in a cultural context to some extent. The degree to which responses are unique to a particular culture constitutes the culture-loadedness of the test. If one uses culture-loaded tests in selection, members of cultural groups for which the items were developed are likely to score better than are members of other cultural groups. However, there is a long history of test developers attempting to identify and delete such items from tests. The essential strategy is to equate individuals belonging to various cultural groups on the trait being measured and to test whether the difficulty level of items (i.e., percentage of individuals getting the items right) is significantly different across groups. If differences are found, a conclusion of differential item functioning is reached. Several statistical procedures have been developed to assess differential item functioning. However, empirical studies indicate that, at least for well-developed professional selection tests, there is no evidence of differential item functioning across racial groups.

Culture-fair assessment does not imply lack of group differences in traits. Just as there are discernible individual differences, there can be discernible group differences in traits. Groups can be defined based on race, gender, or age. When there are group differences, it is possible that there will be more members of the high-scoring group being selected. When the selection ratios (i.e., ratios of those who were selected to those who applied for a job) differ substantially across groups, adverse impact results. That is, the low-scoring group has been adversely affected by the selection procedures.

The issue of bias, in contrast to issues of fairness, is a statistical concept with a precise definition. Predictive bias is defined as the systematic over- or underprediction of a criterion. Two related concepts should be mentioned here: single-group validity and differential validity. Single-group criterion-related validity refers to the idea that the predictor is valid for one group but not for the other group. Differential validity indicates that the predictor is valid in both groups but that the magnitude of the validity differs. Finally, fairness depends on the value system of the individuals, and there are perhaps no right or wrong answers. Fairness perceptions are reflected in applicant reactions.

Applicant reactions to a personnel selection system are important for several reasons. First, applicants might not apply for jobs with an organization if they have negative reactions to the selection system implemented by that organization. Second, when applicants are offered jobs, those applicants with negative reactions might turn down the job offers. This may force the organization to offer the jobs to less qualified applicants and to increase testing and recruitment costs, all of which would lower the utility (i.e., overall value) of the selection system. Third, applicants who have negative reactions to selection procedures employed may be more likely to sue the organization for unfair discrimination. Finally, applicants who have negative experiences with an organization's selection system may boycott the organization in the marketplace and influence others to do likewise.

Of course, it is not feasible to avoid every factor that someone, somewhere, and at some time will find offensive in a selection system. Individuals who score low on a cognitive ability test are likely to blame the test. IWO psychologists have studied various factors that affect applicant reactions based on theories of organizational justice. Research has shown that applicants have positive reactions when the selection system is based on a thorough job analysis, provides multiple opportunities for applicants to demonstrate their skills, is valid, and is administered in the same manner to all applicants. Applicants care about both the outcomes (i.e., the scores obtained whether they are selected or not) and the procedures used (e.g., whether the procedures and rules are applied consistently to all job candidates). Finally, there is evidence that applicant reactions are influenced by the extent to which all applicants were treated with dignity and respect.

Empirical data have been gathered about applicant perceptions of various predictors. In general, most commercially developed predictors have been shown to be

free of bias, although group differences may exist. Organizations must consider these issues before choosing the appropriate predictors to use. When faced with group differences (even if there is no predictive bias), an organization may have to confront the potential for adverse impact for the low-scoring group when that predictor is used in personnel selection. The extent of adverse impact will depend on factors such as the magnitude of existing group differences and how selective the organization is; that is, the lower the selection ratio, the more choosy the organization is and the greater the adverse impact will be on the low-scoring group.

Organizations have tried different strategies to address this potential for adverse impact. One strategy is to specify a low score above which all applicants are considered to be equally qualified. This strategy ignores the linear relation between predictor and criterion variables for most predictors and results in substantial loss of the utility of the selection system. Another option is to search for alternate predictors where there are fewer group differences. However, alternate predictors might not capture all of the individual differences initially identified as critical for successful job performance.

Another option is to consider the use of banding. Banding refers to a set of procedures that groups some scores as not being significantly different from one another. The rationale advanced for the use of banding is the well-known fact that there is a measurement error associated with each observed score. The idea is that scores within a band all are the same; differences among them are probably due to measurement error. Bands are usually formed using the top score in a group of applicants and the standard error of measurement (SEM). Selection can be at random within the bands, or some form of preferential hiring based on other variables can be employed. However, banding is controversial in personnel selection, and several criticisms of the technique question its scientific basis. For example, the SEM is dependent on the reliability of the test. If an organization uses an unreliable test, it can generate broad bands; that is, the SEM increases with unreliable tests. Thus, an organization intent on ensuring enough selected individuals belonging to a low-scoring group can achieve its political goals by using unreliable tests.

#### 4. COMBINING PREDICTORS

Organizations rarely rely on a single predictor for personnel selection. When individuals are assessed on several traits, how should the scores be combined? There

are two options. The first is a compensatory system in which high scores on one predictor can make up (i.e., compensate) for low scores on another predictor. In a compensatory system, the scores from different predictors are combined in one of three ways. The scores may simply be added (i.e., unit weights given to predictors), different weights may be assigned to each predictor based on rational arguments, or (for optimal prediction) regression weights can be used.

The second option is a noncompensatory system in which applicants are expected to obtain a minimum score on each predictor. The minimum score is referred to as the "cut score," and extensive technical literature exists on how to set cut scores. Noncompensatory systems can be further classified as multiple-hurdle and multiple cutoff systems. In multiple-hurdle systems, applicants are assessed on the traits one trait at a time, and only those who meet the requirements on a trait are assessed on the subsequent trait. This procedure is useful when the costs of the various assessments vary. One can administer the least expensive assessments first, and the more costly assessments may be limited to only those who meet the minimum qualifications of the least costly assessments. In multiple cutoff systems, individuals are assessed on all of the traits, and those who meet the minimum cut score in all predictors are selected. Multiple cutoff systems are used when it is costly to bring in applicants for testing on each predictor.

#### 5. EMERGING ISSUES IN SELECTION

An increased emphasis on teamwork has been witnessed during recent years. Some issues, such as whether teams should be staffed with individuals who are similar in trait levels or with individuals who complement one another in traits, need to be explored more comprehensively. The increasing globalization of business has raised the issue of expatriate selection. What variables should be considered in selecting individuals for overseas assignments? Finally, research is needed on whether organizations should select for particular jobs or for qualified applicants in general. This is important when the boundaries among jobs fade. However, such a strategy will raise a whole host of issues such as the definition of "qualified in general." The rapid advances in computer and other multimedia technology have definitely altered, if not revolutionized, how assessments are made. The role of technology is an important and evolving topic in personnel selection.

## 6. CONCLUSION

Personnel selection is one of the central functions of industrial–organizational psychology. This article has reviewed the various predictors used for personnel selection and how predictors are combined to form a selection system. It discussed the relationship between selection and society at large. It considered issues of fairness in selection and applicant reactions. The article underscored distinctions between compensatory and noncompensatory systems. It noted emerging issues of team selection, expatriate selection, and selection for an organization rather than for specific jobs. Personnel selection is a dynamic field in which research and practice go hand in hand in building a science around how to best select and hire individuals for jobs.

### See Also the Following Articles

Achievement Tests ■ Assessment and Evaluation, Overview ■ Employment Interviewing ■ Industrial/Organizational Psychology, Overview ■ Intelligence Assessment ■ Job Analysis, Design, and Evaluation ■ Personality Assessment ■ Psychological Assessment, Standards and Guidelines for ■ Psychometric Tests ■ Psychophysiological Assessment

### Further Reading

Anderson, N., & Herriot, P. (Eds.) (1997). *International handbook of selection and assessment*. New York: John Wiley.

Anderson, N., Ones, D. S., Sinangil, H. K., & Viswesvaran, C. (Eds.) (2001). *Handbook of industrial, work, and organizational psychology*. Thousand Oaks, CA: Sage.

Campbell, J. P., & Knapp, D. (Eds.) (2001). *Project A: Exploring the limits through personnel selection and classification*. Mahwah, NJ: Lawrence Erlbaum.

Cook, M. (1993). *Personnel selection and productivity* (2nd ed.). New York: John Wiley.

Eder, R. W., & Harris, M. M. (1999). *The employment interview handbook*. Thousand Oaks, CA: Sage.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum.

Heneman, H. G., III, Judge, T. A., & Heneman, R. L. (2000). *Staffing organizations* (3rd ed.). Middleton, WI: Irwin McGraw–Hill.

Kehoe, J. F. (Ed.) (2000). *Managing selection in changing organizations: Human resource strategies*. San Francisco: Jossey–Bass.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679–703.

Schmidt, F. L., & Hunter, J. E. (1983). New research findings in personnel selection: Myths meet realities in the '80's. In K. Pearlman, F. L. Schmidt, & W. C. Hammer (Eds.), *Contemporary problems in personnel* (3rd ed., pp. 98–103). New York: John Wiley.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43*, 627–670.

Schmitt, N., Borman, W. C., & Associates. (1993). *Personnel selection in organizations*. San Francisco: Jossey–Bass.

Stokes, G. S., Mumford, M. D., & Owens, W. A. (Eds.) (1994). *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction*. Palo Alto, CA: Consulting Psychologists Press.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8*(4), 216–226.

Viswesvaran, C., & Ones, D. S. (Eds.) (2002). Roles of general mental ability in industrial, work, and organizational psychology [special issue]. *Human Performance, 15*(1/2).



# Play

Cindy Dell Clark

Pennsylvania State University, University Park, Pennsylvania, USA

1. Introduction
2. Play: Developmental Processes
3. Play: Sociocultural Contexts
4. Therapeutic Play  
Further Reading

## GLOSSARY

**bibliotherapy** Therapeutic intervention making use of books and reading.

**catharsis** Process by which unresolved emotional distress is relieved and resolved through expressive activity.

**imaginal coping** Coping through the use of imagination and play.

**language play** Play in which children transform and practice linguistic forms, including sounds, words, sentences, and narratives.

**metacommunication** Communication between players that serves to shape the play process rather than serving as the actual content of play; it serves to mutually negotiate and to maintain the play frame.

**play frame** Parameters for sociodramatic play, as negotiated between players; this includes which roles are to be played and which actions and objects will serve symbolic purposes in the play.

**sociodramatic play** Social pretend play in which children enact dramatic roles and fantasy scenarios.

**zone of proximal development** The gap between a child's independent ability and the ability gained when the child is helped by others who are more capable; this zone may

be manifested by adult assistance in enabling a child to function on a higher level as well as through interaction with more capable peers.

Play, an activity in which children apply nonliteral flexible behavior that is motivated by pleasure rather than external goals, has been theorized to have positive developmental advantages for children's cognitive, affective, and sociocultural growth. In Euro-American society, play has been exploited for therapeutic purposes. Children make use of play for self-healing, both in their own spontaneous play and when they are guided by a professional therapist in formal play therapy. Play entails the re-representation of experience, which enables a flexible remaking of experience that has both therapeutic and normal developmental advantages.

## 1. INTRODUCTION

Straightforward attempts to define play have been elusive. Play, which encompasses both human and animal behavior, evades obvious definition. The difficulty traces partly to the broad and varied forms of play, ranging from rule-governed games, to improvised role-play, to rough-and-tumble mock wrestling. Play, for definitional purposes, is recognized mainly by how

TABLE I  
Four Criteria Defining Play

Criterion	Definition	Example
Nonliterality	Takes an “as if” stance toward reality, transforming actions and meanings	The child lifts and shakes his or her head, roaring as if he or she were a lion
Intrinsic motivation	Activity is done for its own sake rather than for external motivation	The child pulls a wagon for the fun of it rather than to take things somewhere
Positive affect	Pleasure or enjoyment	The child squeals with glee when riding a swing
Process flexibility	Actor flexibly makes use of an object rather than conforming to an action dictated by the object	The child uses a large box for many varied playful uses: turning it upside down as if it were a rock to hide under, turning the opening to the side as if it were a “store,” getting inside the box and rolling down a hill, and/or using it to build a pretend machine with fantasized powers

it contrasts with other dispositions or orientations. A workable approach for defining play involves assessing a behavior based on four criteria. The more the four criteria apply to the behavior, the more the behavior can be categorized as constituting play. These criteria, summarized and illustrated in Table I, are (a) nonliterality (a transformational “as if” stance such as riding on a stick as if it were a horse), (b) intrinsic motivation (an action pursued for its own sake rather than driven by external goals), (c) positive affect (pleasure), and (d) process flexibility (action taken in a free and flexible manner rather than constrained by an object). Activity that fits these criteria can be observed across the human life span.

In addition to occurring in many forms, play has ambiguous or paradoxical characteristics. For example, play is an activity set apart from ordinary reality, yet it often makes reference to ordinary social reality in its content and form. Play encompasses creative innovation by the individual, yet it concurrently is shaped by sociocultural norms. Play, an activity that is not explicitly goal directed, has also been used with therapeutic effect for problems ranging from support during hospitalization, to trauma intervention, to grief therapy among children. Despite (or perhaps in response to) the intriguing ambiguities of play, scholarly attention to play has blossomed over the transition into the 21st century.

This article discusses accumulated theories and findings about play among children, including the developmental trajectory, sociocultural context, and therapeutic application of play.

## 2. PLAY: DEVELOPMENTAL PROCESSES

Early in the 20th century, Groos offered the speculative theory that play served as practice for young organisms who used play to develop their survival skills. Modern-day observers of play have recognized Groos’s insight as appealing. For example, when a child and a playmate pretend to go shopping for groceries, they seem to be practicing social behaviors that prepare them for actual adult roles.

Contemporary theories of play share an underlying assumption that play is beneficial and contributes positively to children’s development. Influential theorists of play who wrote during the first half of the 20th century, including Freud, Piaget, and Vygotsky, all argued that play serves childhood mastery through the transformation of themes within a nonliteral domain. In other words, the representational symbolism of play helps children to master issues related to affect, cognition, language, and/or social understanding. Play is a way in which to advance children’s competence and mastery and has pronounced developmental significance.

In Freud’s book *Beyond the Pleasure Principle*, he argued that play is significant for children’s emotional mastery. Freud asserted that play could have a cathartic effect for children, facilitating the discharge of negative feelings. Building on Freud’s ideas, Erikson posited in his classic work, *Childhood and Society*, that play provides a domain in which children can find self-healing. As Erikson described, children use play to create a re-representation, that is, a model situation

flexibly manipulated to achieve emotional mastery. For example, a child who witnessed a violent crime may repeatedly reenact the scene in improvised play scenarios as if to reencounter affective issues precipitated by the violence. Play, according to Klein's *The Psychoanalysis of Children*, is a suitable medium for children to express complex ideas and feelings that they cannot yet verbalize. Therefore, for young individuals, play is a more appropriate vehicle for therapy than is conversational discourse.

Piaget focused on children's cognitive functioning in his work on play. Piaget conceived of learning as essentially a balancing of two complementary processes: assimilation (i.e., modifying reality to fit one's cognitive structures) and accommodation (i.e., changing one's cognitive structures to conform with reality). In Piaget's book *Play, Dreams, and Imitation*, he viewed play as an unbalanced state in which assimilation dominates over accommodation. Piaget believed that children achieve new learning not through play but rather through practiced and consolidated recently acquired cognitive abilities. In contrast, Vygotsky viewed play as having a more central role in cognitive development, for example, by furthering children's abstract representational thoughts through playful creation of imaginary scenarios. For example, substituting a banana for a telephone in a make-believe phone conversation would precipitate growth in children's ability to use meaning abstractly and symbolically. Vygotsky also referred to pretend play as constituting a zone of proximal development, a framed and imagined experience in which peers can perform activities together beyond a particular child's level of competence outside pretend play. Vygotsky also noted that play has affective components in that children act out desires that have not yet been fulfilled.

Much of the empirical work on children's play has focused on sociodramatic or pretend play in which children enact pretend roles and situations. Such play has been observed as early as the end of a child's 1st year of life, although true social pretend play normally emerges at around 3 years of age. In his *Play, Dreams, and Imitation*, Piaget reported an observation of his own daughter at 15 months of age where she treated a piece of cloth in the same manner as she usually treated her fringed pillow, that is, by laughing and pretending to close her eyes to signal her playful intent. Infants from varied cultures have been shown to produce pretend gestures from roughly 12 to 13 months of age. As sociodramatic play develops, children's activity gradually becomes separated from specific contexts

and outcomes, contains more references to others, and becomes less reliant on realistic props or materials. For example, early pretense might entail using a cup and pretending to drink, but later pretense could involve serving a pretend beverage to a menagerie of pretend animals from an imagined cup filled with a potion having fantastic effects.

Pretend play, as studied in middle-class Euro-American children from intact families, rises in incidence from the 3rd year of life to approximately 6 or 7 years of age. Sociodramatic play declines during the late kindergarten and elementary school years. During the grade school years, another form of play, games with rules, increases.

### 3. PLAY: SOCIOCULTURAL CONTEXTS

Established play theorists, including Piaget and Vygotsky, made presumptions that play, its forms, and its trajectory are universal such that social pretend play might, in theory, emerge at 3 years of age in all societies. Yet children's play is cultural activity and is subject to variation within distinctive cultural milieus distinguished by particular physical, economic, and social contexts.

Play is encouraged more in some cultural settings than in others. Play covaries with economic and social factors, which may affect the opportunity to play in the face of necessary work, as well as with the overall value placed on play by a community. Goncu, who examined cultural variations in children's play in his 1999 edited book *Children's Engagement in the World*, argued that in small villages with subsistence economies, parent-child play is rare due to the pressing demands of work. Children, who are expected to contribute to the economic well-being of families in these villages, might not have time to play. In the same volume, Gaskins described Mayan children as having access to few props for pretend play. Pretend play is not assumed to be a valued activity by Mayan adults; rather, it is tolerated at best. Playful use of language, or language play, is also less frequent among Mayan children than among American children. In another study, Carlson and colleagues found that in American Mennonite society, pretend play is rarely discussed or embellished and is sometimes actively discouraged. Play themes among Old Order Mennonite children during recess adhere closely to real-life roles.

If the amount and content of pretend play is subject to cultural variation, so too are the metacommunications by which play participants interact to maintain and transform play. Martini, in an ethnographic study of peer interactions among Polynesian Marquesan 2- to 5-year-olds, reported that these children avoided lengthy one-to-one negotiations about play activities, preferring to maintain consensus by playing in ways that were familiar to all players, that is, without much change or elaboration made to play events. Marquesan children similarly avoided leadership roles by engaging in games where all participants acted together in similar ways. In contrast, American children cooperated less often so as to negotiate and individually innovate during play. In Haight's study of caregiver-child pretending, she reported on naturalistic home observations of child-mother pretend play among 1- to 4-year-olds, tracing differences in the metacommunications about play between Chinese families and U.S. families. Although both Chinese and American caregivers sanctioned pretend play and initiated pretending for purposes of socialization, their metacommunications about play diverged. Chinese caregivers initiated pretend play in a more hierarchical and didactic manner, whereas American caregivers facilitated a more mutual engagement in which children were allowed to lead the interactions. Table II offers specific examples from Haight's research.

Although developmental scholars once viewed children in social groups who engaged in relatively little pretend play as if they were disadvantaged (by the so-called "play deficit"), this view has been dismissed by culturally sensitive researchers (e.g., Schwartzman). Variations in play reflect larger differences inherent to the particular cultures of children, representing sociocultural distinctions rather than deficits. Healthy functioning in play is best understood by keeping in mind the sociocultural context of a given time and place and by assessing play within the sociocultural confines of children's own environments.

It is important to note that culture is not an all-determining factor of the form or quantity of play. Individual differences occur within a culture in children's fantasy predispositions and play styles. When society has not provided materials explicitly meant for play, children have been known to use discarded items and places (e.g., abandoned buildings, construction sites) for play. Moreover, gender-related differences have great significance, both in the form of parental discouragement of cross-gender playthings and in the form of peer interactions that reflect and shape gender

**TABLE II**  
**Examples of Caregiver-Child Pretending: Girls and Mothers at Play**

Taipei (Chinese)	<i>Caregiver:</i> Stand up, bow down. Teacher is going to deliver a lesson.
	<i>Child:</i> [Child smiles and moves closer to caregiver.]
	<i>Caregiver:</i> Teacher is coming to the classroom. What should the class monitor say?
	<i>Child:</i> Stand up! [Child stands and bows. Play continues as they read a story.]
Age of child: 2½ years	<i>Caregiver:</i> We have finished the story. [She claps her hands.] Before we dismiss the class, the class monitor should say 'Stand up! Bow! Sit down!' Stand up!
	<i>Child:</i> [Stands.]
	<i>Caregiver:</i> Bow!
	<i>Child:</i> [Bows.]
	<i>Caregiver:</i> Class is dismissed. Go play on the slide [She points to imagined slide.] Class is dismissed and you are happy.
Chicago (American)	<i>Child:</i> [Child gets toy, a plastic dashboard with steering wheel and seatbelt, and positions herself as if to drive.]
	<i>Caregiver:</i> Be sure to fasten your seat belt, right? Buckle up for safety!
	<i>Child:</i> [Child fastens seatbelt. Child turns to the steering wheel and pushes the horn.]
	<i>Caregiver:</i> Where are you going? Are you taking a trip?
Age of child: 3 years	<i>Child:</i> To Havana. [Child steers and pretends to drive.] You drive a car like this. Like this, okay? [Play continues as child explains that Havana is far and discusses with her mother how many stops will be needed en route and how old one must be to get a driver's license.]

*Source.* Haight, W. (1999), The pragmatics of caregiver-child pretending at home: Understanding culturally specific socialization practices. In A. Goncu (Ed.), *Children's engagement in the world: Sociocultural perspectives*. Cambridge, UK: Cambridge University Press.

differences. Thorne, in her book *Gender Play*, described how American school-age boys and girls mark boundaries on the playground (and elsewhere) between girls and boys—boundaries that are then crossed, undermined, and challenged through playful behavior. Play reflects not only individual differences and gender-related social behavior but also the distinctive cultural context.

#### 4. THERAPEUTIC PLAY

Play has been regarded, both theoretically and practically, as a healing or coping activity to be engaged in when children are under stress or facing emotional challenges. A widely known application of therapeutic play takes place within the clinical modality of play therapy employed by social workers, psychologists, child life therapists, psychiatrists, and others working in counseling children. Less prominent in the scholarly literature is the fact that children also use play on their own, without explicit professional intervention, as a means of coping with adverse or stressful circumstances.

One example of child-initiated therapeutic play is the self-initiated use of stories or pretend play as a means of expression and resolution. Paley, in her well-known series of reports on young children's playful enactments, reported how children use role-playing to serve emotional ends by reworking details, for example, by eliminating the frightening big bad wolf when enacting the "Three Little Pigs," by resurrecting a dead grandfather during role-play, or by continually improvising threats to a baby (e.g., a baby-eating lion) in reflecting on the vulnerable infantile state.

Eisen, in his book *Children and Play in the Holocaust*, chronicled how children assimilated the horrors and traumas of Holocaust death camps through their own play. While waiting for deportation for internment, children pretended to comfort their dolls. Living in camps surrounded by death and cruelty, children constructed playful pastimes such as corpse tickling and pretend games of gravedigging, funerals, and gas chamber. Experiencing the most intense horrors, children found ways of playing that addressed the very traumas of their environment. Their play was a reaction to atrocity, yet it was also a buffered context in which children could creatively make meaning of a chaotic gruesome circumstance.

In Clark's book *In Sickness and in Play*, she described how American children with chronic illness used self-initiated play as a vehicle for coping with illness. Clark found that chronically ill children spontaneously, and without formal therapy, often employed play as a means of coping with the concerns and hardships of diabetes or asthma. Playfulness, such as pretending a superhero would protect the children from peril or making a game out of an onerous treatment, made space for children to examine and reframe illness-related dilemmas.

In short, play need not be guided by an adult (therapist, teacher, or child life specialist) to provide

meaningful transformative value to children confronting everyday stresses. Even without clinical intervention, play seems to be a child-manageable domain for raising and addressing emotional troubles. Play allows children to approach and express feelings that, if not masked in the "disguise" of play, might be inexpressible, whether through independent play or in therapy.

The development of formalized play therapy began during the 1920s, when Anna Freud advocated the use of toys and games as a way in which to build relationships with child patients. She claimed that this method was superior to the verbal adult method of free association. Her influential idea was that whereas adults "talk out" conflicted emotions, children are better suited to "play them out." Her techniques were further developed by Klein, Axline, Erikson, and others. Klein proposed that play was children's natural means of expression and could serve as a direct substitute for verbalizations, providing sustained symbolic expression to be interpreted by analysts. Axline, inspired by the client-centered therapy movement, proffered that in therapeutic play children should be given free rein to express feelings without judgment or direction by therapists. Erikson regarded play as a natural self-healing measure, that is, a given ability of childhood by which children could be helped to help themselves. Play therapy has developed in multiple directions during recent years, serving as an adjunct to other modes of treatment such as behavior modification and proliferating in forms such as art therapy, sand tray therapy, storytelling, music and dance therapy, puppet therapy, costume therapy, bibliotherapy, and numerous forms of symbolic and game play.

Play therapy has been used in individual therapy one child at a time, but it has also found wide application with groups of children who are facing or have experienced a shared trauma such as war. As an example of the latter approach, Simo-Algado and colleagues used therapeutic play to strengthen personal resources among children in Kosovo who were exposed to trauma during Serbian occupation and the ensuing 1999 NATO military campaign. The goal of the intervention was to preventively buffer against further effects of trauma among a child population that was impoverished, subject to injury from land mines, and dealing with difficult issues such as missing persons, the discovery of mass graves, and the memories of massacres, rapes, burnings, and shootings. Under circumstances of war, the play of Kosovo children had been narrowed to the play theme of reenacted war. Occupational therapists and aids worked with the children, engaging them in other activities that reflected their shared

cultural identity, for example, songs and poems in Albanian and games incorporating native cultural themes. Drawing, clay, narratives, and puppets were introduced as media through which children could vent, process, and discuss their feelings, including fear, anxiety, hatred, and despair. Play, which indigenous teachers were trained to facilitate, provided support, insight, external expression, and a sense of mastery. Following the therapeutic sessions, children took part in sports such as soccer, providing an opportunity for adults to observe the children at play (as well as those children who seem inhibited in free play) and to identify particular children who might benefit from additional individual attention. Through the play-based interventions, children gave expression to their war experiences and then gradually reflected new meanings about past trauma, in the words of Simo-Algado, "moving from victims to survivors." The United Nations Children's Fund (UNICEF) also used art and play in its psychosocial assistance program for children in the former Yugoslavia, through which children were encouraged to recall events from the past and express hope for the future. In crisis circumstances of mass trauma, play interventions can be taught to local professionals and volunteers, who in turn can help many children in need of healing.

Play therapy has also become a mainstream means of clinical intervention with children in individual therapy. For example, play is used as a means of helping pediatric patients by child life specialists who use play to provide child-relevant support in hospitals. Play is implemented as a means of instructing and preparing children for medical treatment procedures as well as a means of expressing or relieving emotional distress among ill youngsters.

Play serves multiple therapeutic purposes. First, it is an appropriate vehicle of communication between a professional and a child. Second, play can serve to relieve emotional tension by providing a means of expressing and reworking issues. Finally, play therapy can introduce children to new forms of beneficial play behavior, potentially to be extended into children's lives outside of therapy.

Although play need not be guided by an adult (therapist, teacher, or child life specialist) to provide

meaningful transformative value to children confronting everyday stresses, adult support of therapeutic play is helpful when children face extraordinary circumstances. With or without clinical intervention, however, play seems to be a child-manageable domain for raising and addressing emotional troubles.

Nevertheless, the underlying process by which emotional resolution takes place in play is not completely understood. The therapeutic utility of play and the role of sociocultural contexts in shaping the trajectories of play are areas in need of additional investigation. Although play has come to be appreciated as a powerful human activity, its ambiguity and transformative dynamism continue to pose challenges for further scholarly understanding.

### See Also the Following Articles

Child Development and Culture ■ Coping

### Further Reading

- Bergen, D. (Ed.) (1987). *Play as a medium for learning and development*. Portsmouth, NH: Heinemann.
- Bruner, J., Jolly, A., & Sylva, K. (Eds.) (1976). *Play: Its role in development and evolution*. New York: Basic Books.
- Clark, C. D. (2003). *In sickness and in play: Children coping with chronic illness*. New Brunswick, NJ: Rutgers University Press.
- Goncu, A. (Ed.) (1999). *Children's engagement in the world: Sociocultural perspectives*. Cambridge, UK: Cambridge University Press.
- Oremland, E. (2000). *Protecting the emotional development of the ill child*. Madison, CT: Psychosocial Press.
- Schwartzman, H. (1978). *Transformations: The anthropology of children's play*. New York: Plenum.
- Simo-Algado, S. (2002). Occupational therapy intervention with children survivors of war. *Canadian Journal of Occupational Therapy*, 69, 205–217.
- Sutton-Smith, B. (1997). *The ambiguity of play*. Cambridge, MA: Harvard University Press.
- Thorne, B. (1994). *Gender play: Girls and boys in school*. New Brunswick, NJ: Rutgers University Press.
- Webb, N. B. (1999). *Play therapy with children in crisis*, 2nd ed.: *Individual, group, and family treatment*. New York: Guilford.



# Political Psychology, Overview

Patrizia Catellani

Catholic University of Milan, Milan, Italy

1. Scope of Political Psychology
  2. Political Knowledge and Political Attitudes
  3. Political Identity and Conflict
  4. Political Communication
  5. Political Participation
- Further Reading

## GLOSSARY

**accessibility** The ease with which information in long-term memory can be retrieved.

**bias** A systematic distortion in the processing of information.

**entitativity** The extent to which a social category is perceived as having real existence.

**heuristic** A simplified procedure or rule of thumb for making a decision or solving a problem without systematic processing of information.

**ideology-based voting** Voting choice based on ideology or values.

**issue-based voting** Voting choice based on parties' position on given political issues.

**matching effect** The different impact of persuasive messages as a function of authoritarianism and other characteristics of the target.

**media agenda setting** Decision process regarding the relative space and time to be devoted to different news by the media.

**performance-based voting** Voting choice based on evaluation of past performance of a party or a political leader.

**political identity** The part of the self-concept that derives from belonging to a political group.

**salience** Prominence, distinctiveness; any aspect of a stimulus that causes it to stand out and attract attention.

**third person effect** A tendency to overestimate media influence on others and underestimate media influence on oneself.

**trait-based voting** Voting choice based on personal characteristics of a political candidate that may have either a positive or a negative connotation.

**voter's paradox** A definition of voting as an apparently contradictory action, since costs implied by the act of voting are higher than the benefits the single citizen may expect from that act.

Political psychology investigates representations and actions of political actors, be they citizens, politicians, or members of groups characterized by collective and public goals. This overview begins with a definition of the area of interest of political psychology and of the main stages that have characterized the development of the discipline. We will then examine how citizens know and evaluate political reality, with a special focus on political candidates. Attention will also be devoted to the differences between experts and novices in the political domain, and superordinate principles (such as values and ideologies) around which political attitudes may be organized. In the third section of the article, political identity will be dealt with. Different groups one may identify with in the political context will be highlighted, with a particular stress on regional, national, and supranational identities, and on the roles they play in international conflicts. Media and various

forms of political communication, such as political interviews or discourses, influence most of our interpretations of political reality. These issues will be analyzed in the fourth section, with special emphasis on motivations and goals that underlie political messages, as well as their persuasive effects. Finally, we will consider psychological factors involved in voting behavior and in other forms of political participation.

## 1. SCOPE OF POLITICAL PSYCHOLOGY

### 1.1. Area of Interest

Let us consider millions of people in different countries all over the world gathering on the same day in main squares and streets of their towns and demonstrating in favor of peace, shouting slogans against terrorism and war. Let us consider other millions of people watching the event on television. If we zoom in on any of the demonstrators, politicians, bystanders, or the people simply watching this political event, we may investigate how they represent people struggling or not struggling for peace, what expectancies they have regarding possible success of peace movements, what plans they have regarding their future engagement in peace movements, and so on. We may also move beyond the specific event and investigate their political views, their political identities, or their voting intentions. All these issues fall within the area of interest of political psychology, which studies the representations and actions of all the people involved in political reality, from simple citizens to political leaders.

Most research in political psychology originates from the extension of “basic” psychological research to the political context, an extension that implies consideration of two main categories of context-specific factors: (1) rules and constraints that are present in the political context, and (2) motivations and goals of individuals and groups who act in the political context. For example, research on how citizens form an impression of political candidates takes into account the fact that in the political context, unlike other contexts, knowledge of a target person is often not direct, but is filtered by media. It also considers that motivations underlying the citizen’s perception of a political candidate may be different from those underlying perception of a target person in other contexts, and these differences may have an influence on perception accuracy or

on traits perceived as most relevant. The influence of the citizen’s ideological orientation on candidate perception is also taken into account—a variable that may not be as relevant in other contexts (see Section 2). These are examples of how research in political psychology may identify a series of factors that moderate or mediate effects already observed by basic psychological research, in some cases also highlighting previously neglected effects.

### 1.2. Historical Development

A short reference to the historical development of political psychology may offer an idea of the main issues dealt with by the discipline so far, as well as an idea of the psychological theories that have been most frequently assumed as frames of reference. Four main stages may be identified.

1. *1940s to 1950s: Politicians’ personality.* Studies of political psychology initially focused on the personality of political leaders, assuming psychoanalysis as their main frame of reference. Biographies of famous leaders (e.g., Alexander the Great, Hitler) were carefully reconstructed through historical and archival data, with a special stress on the educational environment that characterized the leaders’ development. Features of the adult personality, for example, a strong tendency to dominate over others or to be aggressive toward external enemies, were traced back to a series of psychological processes originating in childhood. Other studies, not focusing on political elites but on the wider category of political activists, adopted Maslow’s motivational model as a frame of reference. According to this model, people would perceive the need for engaging in political activity only when other, more basic needs have already been satisfied (e.g., health or social needs). As a consequence, politically active people would be characterized by high self-esteem and strong personal efficacy, that is, they would be deeply convinced of being able to engage successfully in political action.

2. *1960s to 1970s: Public opinion and voting behavior.* In this period, research focus changed dramatically from political elites to normal citizens and from personality to behavior. Surveys aimed at tapping political attitudes and voting intentions were being carried out on a regular basis, before or after national political elections. In 1960, Campbell et al. published a volume, *The American Voter*, in which they suggested that political attitudes would develop quite early, influenced by the individual’s family environment and, by adult age,

would become rather stable and resistant to change. Party choice would therefore be the consequence of a largely irrational process, defined as party identification, instead of being based on a thorough consideration of information regarding candidates, parties, and political issues. Although this approach evokes the image of an irrational voter, a very different, substantially opposite approach to voting behavior also developed in the same years, this one supporting the image of a rational voter. In *The Economic Theory of Democracy* (1957), Downs made reference to rational choice theory (RCT), according to which individuals, in this case voters, may effectively use all the information they have at their disposal so as to choose, from a series of possible alternatives, the one characterized by the maximum subjective expected utility. Decision-making would consist in an accurate evaluation of the pros and cons of each alternative in order to choose the one that is perceived as the potentially most useful for the individual, where utility is intended as functional to the personal, mainly economic, interest of the individual.

3. *1980s: Political cognition.* As compared to the previous stage of research in political psychology, in this stage attention shifted from political attitudes and behaviors per se to the cognitive processes that underlie and influence attitudes and behaviors. The 1980s were characterized by the extension of the “social cognition” approach to the political domain. Political psychology borrowed from social cognition a well-articulated corpus of theory and data regarding what happens in the different stages of human information processing, from initial coding of new information, to its organization in the individual’s mind, to its retrieval from memory. The citizen was seen as actively selecting and processing political information, attributing meaning to it through comparison with information already stored in memory, thus arriving at a personal interpretation of political reality. Information processing is constrained by the limits of human memory, mainly the capacity for dealing with only a limited amount of information at the same time. Because of these limits, citizens often do not thoroughly examine all the information they are exposed to. To save mental energy, they recur to simplified and rather superficial reasoning strategies, so-called heuristics. These strategies are functional to the everyday goal of making decisions and planning behaviors in a reasonably short time, but, being based on partial and approximate processing of available information, they may also lead to systematically distorted evaluations (biases). Thus, the political

cognition perspective offers an alternative to the above-mentioned opposition between the irrational and rational voter by suggesting that voters are characterized by a limited rationality.

4. *1990s to today: Social psychology of politics.* Although the synthesis of research developments that are still unfolding should be made with some caution, two main tendencies in contemporary political psychology may be singled out. The first one involves an enhanced attention to context-related factors (citizens’ motivations and goals, power relationships, normative and political constraints, etc.) and how these factors may influence both quality and quantity of political information processing. For example, cognitive theories of decision-making may be applied to voting decisions, taking into account the fact that context-related factors, such as the voting system of a given country or the dimensions and numbers of the opposing parties, may condition the way in which citizens make their decisions. The second tendency identifiable in contemporary political psychology is a gradual abandoning of a substantially individualistic perspective and the assumption of a more marked social perspective. Considering that one of the functions of politics is to regulate the relationships between individuals and groups, especially resources and power distribution, it is difficult to conceive of a psychology of politics that is completely detached from the social dimension. Social identity, social influence, intergroup relations: these are some of the factors that have been widely investigated by social psychology and are now also being taken into account by political psychology.

## 2. POLITICAL KNOWLEDGE AND POLITICAL ATTITUDES

In this section, we consider how political reality is perceived, felt, and reflected upon by the individual. As we have seen, studies of political psychology in the 1960s focused on public opinion: the substance of people’s political attitudes and how their attitudes on various political questions are interrelated. While these questions remain relevant, and indeed central to political psychology, in the following years there was greater recognition of the interdependence of attitudes and cognition, namely, how being in favor of or against a political candidate or issue is related to what we know (or believe we know) about them.

As compared to the past, contemporary society has developed an amazing number of ways through which information, including political information, can be transmitted and disseminated (TV, radio, Internet, mobile phones, etc.), and the potentiality of these media is still far from having been entirely explored. This means that, in principle, everybody has access to an incredibly large amount of information. However, such an access is filtered and regulated by a number of selection processes. Whereas some of these processes are independent from the individual, others depend (at least partly) on them.

1. *Media agenda setting.* This information selection process is operated by the media, that is, by those who decide the relative space and time to be devoted to each piece of information, as well as the language and images that convey information. These choices inevitably reflect a given way of reconstructing and interpreting reality, and are not under the control of the individual who receives information via the media.

2. *Non-voluntary media exposure.* This second source of selection is linked to the individual's life but at the same time is largely not dependent on the individual. Being born in a given place, living in a given family environment, attending a given school, these are all factors that inevitably influence the kind of media, and therefore the political information, to which the individual is exposed.

3. *Voluntary media exposure.* Although influenced by the individual's social and family sphere, media exposure is also voluntary, because individuals, once they have reached a certain developmental stage, start to choose the kind of media to which they want to be exposed.

4. *Selective attention.* Being exposed to a given piece of information does not necessarily mean paying attention to it. Social cognition studies have clearly shown that people engage in a process of selective attention, more or less consciously, that leads them to process only the information that, for various reasons, they perceive as salient or relevant in the given context.

In this section, we shall examine the latter type of selection process. We shall see what information is more likely to catch the individual's attention, with a special focus on information regarding political candidates. We will then consider how a well-established attitude regarding a political candidate or issue may influence or bias the processing of new information. Differences between experts and novices in political information processing will also be highlighted. Finally, we shall consider various theoretical proposals regarding

organizing principles (ideologies, values, etc.) of political knowledge and attitudes.

## 2.1. Candidate Perception and Evaluation

Citizens process information regarding political events, issues, programs, parties, candidates, or leaders. These various categories of information are obviously related to each other. However, the relative salience of each category may vary according to the goal pursued in processing that information and other context-related factors. In particular, when elections are upcoming, information regarding political candidates is especially likely to become salient. Actually, perception and evaluation of political candidates has been the most investigated area of political information processing so far, borrowing from the consistent body of results gathered by social cognition research regarding how people form descriptive and evaluative judgments about target persons.

### 2.1.1. Traits

When we form a judgment about a political candidate, we are likely to devote our attention to three main information categories: (1) the candidate's party, (2) the candidate's position on political issues, and (3) the candidate's traits, that is, personal characteristics that may have either a positive or a negative connotation. An analysis of the answers to an open question in a pre-electoral survey carried out on a regular basis in the United States (National Election Study) has shown that when people are asked to speak about candidates, they focus on their personal characteristics more than on their party or their position on political issues. This tendency is not significantly influenced by the individual's education level, suggesting that the stress on personal traits is not necessarily due to a lack of political competence, but may instead be due to the prevailing need for the individual to choose the right person to represent his or her own political needs and positions. Although some variation in the relative stress on personal traits has been found according to the political system and context in which individuals find themselves, recent years have witnessed an increasing personalization of politics, which is likely to make perception of the candidates' traits more and more important in citizens' political choices.

There is substantial consensus on what traits an individual focuses on when evaluating a political

candidate. These traits may be divided into five main categories: (1) competence (intelligence, ability in dealing with political issues); (2) integrity (being honest or, on the contrary, being prone to corruption); (3) reliability (being consistent, being able to keep one's word); (4) charisma (having capacity for leadership); and (5) image (being good-looking, attractive, photogenic). While the importance of the first two categories is generally more heavily stressed than the others, the relative weight attributed to each category may vary according to the context in which candidate evaluation takes place. For example, the salience of traits related to the candidate's integrity may be enhanced in a context characterized by a series of scandals involving corrupt politicians. These events usually get wide media coverage and may induce citizens to construct their evaluation of candidates around the dimension of integrity versus corruption. In these cases, a "figure-ground phenomenon" is likely to happen, in that the honesty of a given politician becomes especially salient because it is perceived as peculiar in a context of diffused corruption.

### 2.1.2. Emotions

The evaluation of candidates is influenced not only by their traits but also by the emotions (happiness, anger, etc.) they trigger. We may in fact wonder whether our judgment of a given candidate is based more on the traits we attribute to him or her (cognitive component of the attitude) or on the emotions we associate with him or her (affective component of the attitude). Research suggests that emotions, especially positive ones, are very good predictors of candidates' evaluation and voting intention. However, the relative weight of traits and emotions can vary according to the context. It has been suggested that emotions might become especially relevant in a changing context, when new candidates or new parties emerge in the political scene. The relative weight of positive versus negative emotions is also likely to depend on the context. Consistent with the figure-ground phenomenon described previously, higher relevance of positive emotions might be due to the fact that citizens are especially inclined to notice something positive when the context is mainly negative.

### 2.1.3. Influence of Ideological Orientation

So far, we have focused on traits and emotions that are generally perceived as relevant when evaluating a political candidate. However, some differences in perceived

relevance of traits and emotions have also been highlighted. In particular, citizens with opposite ideological orientations have been shown to differ in the positive traits they perceive as most relevant in political candidates. In a study by Caprara *et al.*, left-wing citizens asked to rate personality traits of candidates of different parties stressed the dimension of sincerity/reliability (sincere, loyal, reliable, accountable, authentic, etc.) in candidates of their own political side and the absence of the same traits in candidates of the opposite side. Right-wing citizens performing the same task have instead focused on the dimension of energy/innovativeness (dynamic, resourceful, resolute, energetic, creative, etc.), rating these traits as high in candidates of their own side and as low in candidates of the opposite side.

Interestingly, when evaluating their own personality, citizens tend to rate themselves as high in the same traits they perceive as typical of their favorite candidates. Can we therefore infer that citizens tend to choose candidates that are similar to them? Not necessarily. It may well be that an assimilation effect is taking place here, according to which we are inclined to perceive people we like as more similar to ourselves than they actually are.

## 2.2. Assimilation and Contrast Effects

As social cognition research has widely demonstrated, once a positive (or negative) attitude toward a given target has developed, processing of new information regarding that target may be biased, in the sense that we look for a confirmation of our own attitudes. This may happen in various ways, through discarding inconsistent information, devoting selective attention to consistent information, or reinterpreting neutral information in a way that is consistent with our attitudes. Attitude-consistent processing of new information is especially likely when the relevant attitude is strong and when information presents itself as somehow contradictory or ambiguous. Both conditions may be often found in the political domain. People typically develop strong attitudes in favor of or against political candidates, parties, or issues. At the same time, these attitude objects are not always stable over time, but may instead change, sometimes rather substantially. Let us consider a politician who argues about the opportunity for a country to restrict or enlarge its immigration policy. The politician's position on such an issue is not necessarily the same in different places and at

different times. With regard to place, the politician may present his or her stand on the issue with different tones and nuances according to the audience. Politicians know very well how to tailor their message, using all the possibilities offered by the richness of language and rhetorical strategies (see Section 4). Regarding time, politicians may change their position on a given issue due to several reasons, such as shifts in political alliances or changes in the national or international political situation. The more mutable and ambiguous a political position on an issue, the more strongly the citizen may be inclined to select, interpret, or even distort information in a way that is consistent with his or her consolidated attitude.

If we have a positive attitude toward a candidate, we are especially inclined to perceive the candidate's position on relevant political issues as being very similar to our own (assimilation effect). For example, we may ask people to rate their position on a restricted immigration policy on a 7-point scale, ranging from 1 (fully against) to 7 (fully in favor). Then we may ask them to use the same scale to rate the supposed position of political candidates on the same issue. Research results show the presence of a high and positive correlation between one's position and the one attributed to the preferred candidate: if one's position is rated with a 5, the candidate's position will also be rated with a 5, whereas if one's position is more extreme (i.e., rating 7), the same extreme rating will be attributed to the candidate. An opposite tendency to differentiate our position on a given issue from that of a candidate we do not like has also been observed (contrast effect). The contrast effect is weaker, however, than the corresponding assimilation effect. This asymmetry has been attributed to a positivity effect, according to which people prefer to focus their attention more on what they like than on what they dislike.

### 2.3. Expertise

So far, we have focused on how people know and evaluate political candidates and issues irrespective of the degree of competence these people have in the political domain. However, political information processing appears to vary according to people's level of expertise. Political expertise has been defined as the synthesis of different dimensions, such as political interest, political knowledge, and media use. Experts and novices in the political domain seem to process information very differently, and these differences affect all stages of information processing.

1. *Coding.* Experts understand new political information more quickly than novices do, and are quicker in expressing evaluations on political issues and candidates. Such evaluations are also more stable over time. These differences are due to the fact that frequent reasoning about politics makes a number of political concepts immediately accessible to the experts' minds and therefore easy to employ in the interpretation and evaluation of new information.

2. *Organization.* Experts' political attitudes are more consistent than those of novices, and experts experience a larger number of thoughts related to a political candidate or issue. This difference has been attributed to how concepts and their positive or negative connotations are organized in people's minds. Experts' conceptual networks are not only characterized by a larger number of concepts, but also by a larger number of links between concepts. Moreover, a hierarchical structure is often well developed in experts' mind, so that a series of concepts is organized around a limited number of more abstract concepts. This means that the activation of one concept easily leads to the activation of other related concepts at the same or a higher level of abstraction. Such an interrelated network increases the experts' consistency. Different political issues, such as unemployment subsidies, norms toward immigration, and the welfare state, may be linked to each other in the experts' mental network, and all of them may in turn be linked to a more abstract concept such as solidarity. If the experts' attitude toward applying the solidarity principle in politics is positive, related attitudes will also be positive. This may not be the case for novices, who may be more prone to express inconsistent attitudes. At the same time, novices are likely to be less sensitive to inconsistency in politicians' attitude.

3. *Retrieval.* Better organization of political information in experts' minds facilitates the process of retrieving information; this explains why the experts' recall of political information is better than that of novices.

### 2.4. Ideological Orientation and Values

As we have seen, specific political concepts and attitudes may be cogently linked to more abstract, unifying principles. But what are these principles? In his article in this encyclopedia entitled "Ideological Orientation and Values," Hans De Witte identifies them with values and ideologies, showing how research on this issue has developed over time. Research on values has focused

on single relevant value dimensions, such as conformity versus self-direction or materialism versus post-materialism, but it has also developed more encompassing models, including several value types (e.g., power, security, conformity, or universalism) and their compatibility relations. Prominence attributed by individuals to one or the other of these values has been shown to be a good predictor of political attitudes and voting. Research on ideology has especially focused on the notion of conservatism (as opposed to progressiveness) and has highlighted two different dimensions of it, economic conservatism and sociocultural conservatism. Like values, ideology dimensions are predictive of political attitudes and voting.

In his article “Human Rights,” Willem Doise also makes wide reference to values and ideology. The general relevance people attribute to human rights is related to values, and this reaches its highest level in people for whom the values of universalism and social harmony are very important. People may also differ in their representation of human rights according to their values and ideology. For example, for some people, the right to equality means that everybody has to be treated in the same way by the law or by public institutions, whereas for others, equality also implies that all should have equal access to a decent standard of living, education, and health care. Finally, people with different values and ideologies seem to differ in their opinions of the relative role individuals and institutions (e.g., governments) should play in safeguarding human rights.

### 3. POLITICAL IDENTITY AND CONFLICT

Campbell's early reference to the notion of party identification as a way of explaining people's voting behavior (see Section 1) undoubtedly had the merit of highlighting how through their vote people may express a feeling of belonging to a given party—a feeling made up of a mixture of cognitive and emotional components. However, the notion of identification as developed by Campbell had a limited heuristic validity. He saw strong party identification as a cause of a stable vote for a party, but envisaged no specific measure of identification. On the contrary, it was precisely the observed stability of voting that induced Campbell to infer the presence of identification. As such, his explanation of stable voting through identification amounts to little less than a tautology. In addition, Campbell's

view of identification as being deeply rooted in developmental age failed to explain possible shifts in voting preferences.

Research on identification processes in the political domain has gathered new strength in recent years. Thanks to the substantial progress made by psychosocial research on identity, we know a lot about how people define themselves as individuals or as members of meaningful groups, how they perceive members of other groups, and what processes characterize intergroup relations, leading to overt competition and conflict or, on the contrary, to peaceful coexistence and cooperation. While extension of this approach from the social to the political domain has proved to be fruitful, specific research in this field has also started to highlight some peculiarities of identification processes in the political domain.

#### 3.1. Political Identity

In some circumstances of our lives, we tend to define ourselves as unique and different from anyone else (personal identity). In other circumstances, we define ourselves as members of a group: we perceive ourselves as similar to the other members of the group, sharing the same values and goals (social identity). One may identify with several groups (multiple identities); these may include groups that one cannot choose to belong to (e.g., gender or race) or groups chosen by the individual (e.g., a sports team, an association, a circle of friends). Moreover, some of these groups may be nested, with a more inclusive or superordinate group (e.g., the company one works for) including other less inclusive or subordinate groups (e.g., a specific sector of the company or a specific workgroup). According to the context, a given identity may be perceived as more salient than others. For example, party identity may become especially salient during the electoral period, when people discuss a politically sensitive issue, or when we confront ourselves with people who belong to different parties. Apart from the context, a given identity may also be more “chronically” salient to the individual; this means that a given identity may be especially relevant to the individual's self-definition. As compared to other kinds of identity, political identity seems to be especially characterized by the need for expressing the group's values in public and, more generally, for acting together in order to gain further consensus around these values. As Bert Klandermans highlights in his article “Collective Action,” a shared identity is one of the main antecedents of people's

involvement in collective action. At the same time, being involved in collective action leads to enhanced political identity.

An analysis of autobiographical accounts of political activists carried out by Catellani *et al.* has shown the presence of three main levels of political identity, characterized by a different degree of abstractedness and embedded within each other: ideological identity, party identity, and sub-party identity. These identity levels are likely to satisfy different needs, and, accordingly, the relative salience each of them assumes for the individual is likely to vary depending on circumstances. Ideological identity is characterized by the highest degree of abstractedness. When people identify themselves with left-wing or right-wing people, they state their inclusion within a social category that is difficult to visualize, with wide and ill-defined boundaries, including a large number of heterogeneous individuals. At the same time, such an abstract category of identification is likely to be perceived as more stable over time and less subject to transformations as compared to a more concrete category, such as a political party. In fact, ideological identity has been shown to be especially salient after a political schism, when a previous existing party splits into two or more new parties. In these changed circumstances, stressing one's ideological identity has the function of safeguarding a certain degree of continuity with the past, to keep a balance in one's own identity.

### **3.2. Regional, National, and Supranational Identity**

Investigation of how multiple or nested identities may exist and coexist in the political domain may be usefully extended to the analysis of regional, national, and supranational identities. The issue has become of increasing relevance in a global world that is witnessing an ever-increasing close cohabitation of citizens of different nationalities as well as attempts to create solid supranational realities such as the European Union.

As already mentioned, a superordinate social category may be more difficult to visualize, because it is formed by a large number of people, possibly rather heterogeneous. Developing identification with a category of this type first requires developing a perception of its existence. Some studies by Castano *et al.* have investigated how a political institution such as the European Union may be perceived by people not so

much as an abstract and far-away entity but as a real community of citizens. These studies have referred to the notion of entitativity, the extent to which a social category is perceived as having a real existence. It is assumed that four factors especially contribute to enhance a social category's entitativity: common fate, similarity, proximity, and boundedness. If we ask people to read a journal article or to see a video stressing either common trends or differences in the economic development of European countries, we are manipulating the common fate factor. If we show people a map in which the boundaries of single European nations are thin lines but the external perimeter of the European Union is a thick line (or vice versa), we are manipulating the boundedness factor. Similarity and proximity may also be manipulated in various ways. Such manipulations have been shown to be effective in increasing people's degree of identification with the European Union. Their effect, however, seems to be felt mainly by people holding moderate attitudes toward the European Union and is not perceived by the so-called Euro-skeptics or Euro-enthusiasts.

Investigation of a higher order identity, such as supranational identity, usually focuses not only on its conditions of existence, but also on its conditions of coexistence with lower order identities. For example, can people strongly identify with Europe and at the same time strongly identify with their nation or region? Studies carried out in different countries have shown that citizens may show high identification with both the superordinate and the subordinate categories. These results suggest that regional, national, and supranational identities may indeed coexist in the citizens minds. This is more likely to happen when the frame of reference is one of integration between different identities, in particular when stressing the superordinate identity does not mean absorbing lower level identities and forcing homogeneity but, on the contrary, accepting a certain degree of internal heterogeneity. Ideally, diversity will become one of the core values of the superordinate identity. This means that belonging to a social category made up of different components that integrate with each other may become a core feature of, for example, North American or European identity. When, however, the frame of reference is one of competition between different identities, in particular when stressing the superordinate identity means absorbing lower order identities or threatening them, competition easily turns into open conflict, and fear of losing a given regional or national identity may hinder development of a supranational identity.

### 3.3. International Conflict

Analysis of what conditions may favor development of a new supranational identity starting from well-established national identities may be useful in highlighting strategies of international conflict resolution. In his article “International Conflict,” Herbert Kelman describes psychosocial processes that underlie conflict insurgence and its escalation, as well as psychosocial processes that may instead favor a break in the spiral of conflict in the direction of conflict resolution. An essential step of an effective strategy of conflict resolution is the acknowledgement of each party’s needs. An effective, long-term solution is unlikely to arise if either party has the feeling that its needs or its very existence is being denied or simply ignored by the other party. A similar premise may be found in the dual-identity model of conflict resolution, according to which a first essential step in negotiation between groups consists of creating conditions for the already existing group identities to be expressed and safeguarded. Although open expression of each group’s identity and related needs may initially enhance intergroup conflict, it also has the advantage of enhancing each group member’s perception of having a well-established identity. This is an ideal psychological premise for an effective interaction with the other party, in order to find solutions that address not only a single party’s interests but also superordinate interests. If, however, the first step—expressing one’s group identity—is lacking, members of each group may feel that their group’s identity is threatened. The resulting uncertainty prevents the group from taking the next crucial step in the negotiation process, finding a solution that satisfies superordinate interests.

## 4. POLITICAL COMMUNICATION

When we think of politics, we are likely to think of a phenomenon encountered largely through the mass media. The link between politicians, events, and issues on the one hand and most citizens on the other is mediated through the complex web of information generated by newspapers, magazines, television, radio, and the Internet. While the mass media has significantly increased the opportunity for politicians to reach citizens, communicate, inform, and put forward persuasive messages, it has also offered citizens the means to scrutinize politics and politicians more closely while offering information on a previously

unthinkable scale. However, if the media multiply the opportunities for communication, it must also be recognized that the process of mediation is not neutral. Through the people who control them, the media select, classify, and shape political information, thus contributing to the creation of political reality.

The influence exerted by the media on political information processing and the degree to which citizens are aware of this influence will be the first issues dealt with in this section. Next, we will consider how political actors communicate and employ language and other communicative strategies in order to fulfill their goals. Our attention will be especially focused on two communicative contexts, political interviews and political discourses.

### 4.1. Mass Media and Politics

Studies on information processing have shown that a concept or an attitude that has been frequently retrieved in the past becomes more accessible to the individual’s mind, that is, more likely to be retrieved again and to be perceived as relevant. Therefore, when media coverage of a given political issue is high, individuals are repeatedly exposed to it, and its accessibility is likely to become higher. Studying the cause-effect relationship between media coverage and accessibility or perceived relevance of a political issue is not an easy task. Is a political issue perceived as relevant because it is widely covered by media, or is it the other way round? Adopting a temporal series technique, some studies have measured perceived relevance of a number of political issues at different moments in time and have monitored media coverage of the same issues in between. Results suggest that media coverage significantly increases the relevance of political issues to which people already attribute a certain degree of relevance. In other words, media have the effect of amplifying the importance of a given issue and possibly of altering the hierarchy of the relevance of political issues. This may have not only generic consequences on the evaluation of political reality, but also more specific effects on the evaluation of a given politician. For example, before the Gulf War, U.S. citizens’ evaluation of three main policy areas of President George Bush (foreign affairs, domestic affairs, and economic affairs) equally contributed to form the global judgment of Bush, whereas after the Gulf War, evaluation of Bush was based significantly more on his foreign affairs policy. Thus, media coverage of a political issue leads people to perceive that

issue as more important for the country than they thought before; consequently, that issue weighs more in their evaluation of political leaders.

Are we aware of media influence? Yes and no. We are inclined to think that the mass media influence “people” or “others,” but not ourselves. This is the so-called third person effect, according to which we overestimate media influence on others and underestimate media influence on ourselves. This tendency has been defined as a self-serving bias, in the sense that the perception of having full control on the surrounding reality and on one’s own choices contributes to raise the individual’s self-esteem. Thus, one’s capacity of resistance to media influence is enhanced if compared with others’ vulnerability to the same influence. The third person effect is more evident in those who perceive the media as untrustworthy and perceive target people as distant from themselves. In a post-electoral survey carried out in Australia by Duck *et al.*, people asked to evaluate the influence of a number of television programs (political debates, news, electoral spots, etc.) on voting choice declared themselves to be the least vulnerable to that influence, followed by voters of their party, voters of the opposing party, and voters in general.

#### 4.2. Political Language

We may undoubtedly say that politics is made up of words. Politics deals not only with concrete, tangible objects, but also with abstract concepts or notions to which people attribute a shared meaning (e.g., democracy, liberalism, welfare state). Moreover, the planning of future events is an essential component of politics, which therefore often refers to something that does not necessarily exist now but that might become real in the future. Such referential elusiveness explains why political language has often been described as polysemous, many-faceted and, in the end, ambiguous. At the same time, such elusiveness makes it clear that political reality is mainly construed through language. Politicians propose a certain interpretation of political issues and events, which may be defined as successful when it becomes shared by a large number of people. Such interpretation often implies stressing similarities or differences between political groups, thus favoring or hindering citizens’ identification with these groups.

In studying political language, psychology focuses on the analysis of cognitive, motivational, and psychosocial factors that guide speakers and listeners in the political domain (politicians, activists, journalists, or

citizens). Special attention is devoted to the study of the various kinds of communicative contexts in which politicians find themselves. Although each context has peculiar characteristics, politicians’ communication is generally guided by three main goals that may be found across contexts: (1) presenting a positive image of themselves and their party, (2) presenting a negative image of opposing parties, and (3) trying to gain consensus. This means that on the one hand, politicians try to define themselves, to make clear what their political views are and how they differ from other politicians’ views. On the other hand, they aim to extend the boundaries of their party to include as many people as possible. Language helps politicians in the difficult and contradictory task of defining their group and at the same time trying to widen the boundaries of this group. In the communicative contexts we discuss below, we will consider how politicians defend their image when interviewed and how they use political discourse to define and depict social categories in order to gain consensus.

#### 4.3. Interviewing Politicians

Thanks to the growing media-centered and spectacular character of politics, interviews have become one of the main channels through which politicians present themselves and their political positions. Psychosocial research on question–answer exchanges has shown many, often subtle, strategies employed by interviewer and interviewee to pursue their communicative goals. Extension of this approach to the analysis of political interviews started from two considerations that easily arise when listening to political interviews: (1) journalists’ questions are often not neutral and (2) politicians often do not reply to journalists’ questions. A systematic examination of interviews given by politicians has led to identifying an interesting typology of questions and answers as well as conditions that make non-replies more likely. As already mentioned, one of the politician’s main goals is to present a positive image of him- or herself. According to the so-called face model of political interviews, developed by Bull *et al.*, during interviews politicians tend to defend (1) their personal-political face; (2) their party face; and (3) the face of significant others. Journalists’ questions that are perceived as face-threatening are therefore more likely to get a non-reply. Questions may threaten a politician’s personal-political face in various ways, for example, by creating a negative impression of the politician (e.g., “Isn’t all this emphasis on personality a cover for the

fact that you haven't got a big idea?"), hinting at possible future difficulties, highlighting contradictions with past statements, or stressing difficulties in clarifying personal beliefs (e.g., "I wonder whether wavering voters aren't influenced by not quite knowing where you, Prime Minister, stand"). Similar threats may be found in questions regarding the politician's party (e.g., "Things aren't looking that good for your party, are they?"). Finally, questions may focus on other people or groups but still imply a threat to the politician's face. This is the case for questions implying that the politician does not care for the electorate or support a friendly country, and for questions luring the politician into awkward admissions regarding opposing parties (e.g., "Is there really a shift of opinion toward the Liberal Democrats or is it because they have run a better campaign than you have?", a situation in which confirming either alternative offered by the question would lead the interviewee to support an opposing party).

Although the interviewer has various ways to threaten the politician's face, the politician has various ways to avoid replying. Not replying does not mean being silent, but rather adopting communicative strategies aimed at avoiding a direct reply without necessarily appearing rude or not collaborative. The outcome will be a reply that may be defined as equivocal, because it is not consistent with the interviewer's communicative intention; it is, however, consistent with the politician's intention of presenting a positive face to the electorate. Equivocal replies may include ignoring the question, replying with another question, criticizing the question, or refusing to answer.

#### 4.4. Political Discourse

In their discourses, the politicians' ultimate goal is gaining the widest possible consensus from their audience. They aim to create in their public a political identity that is consistent with their own and that of their party, so as to favor mobilization in their favor. To do so, politicians address not only single individuals but also significant groups to which their audiences belong and are likely to identify with. For example, they can explicitly address blue-collar workers, entrepreneurs, or housewives. They may also address wider social categories, such as the citizens of their country, or some vague social categories, such as all honest people: making these categories salient means addressing millions of potential voters. Whatever social category they are trying to make salient to their audience, politicians will use arguments that highlight the

fact that the values and goals of that category are consistent with the ones their party is ready to fight for (e.g., "Citizens want to be able to walk safely in their neighborhoods, and that's exactly what we're aiming for with our new bill on criminal offenses"). Very likely, they will also argue that the same values and goals are inconsistent with those of the opposing party. Whenever this is possible, politicians also try to present themselves as prototypical members of the salient category (e.g., "I am a father/mother myself" when speaking about school reform), thus implicitly stating that they are trustworthy and entitled to represent the category.

A confirmation of the relevance social categories have in politicians' discourses has been offered by some studies that compared how politicians of opposing parties interpret the same political event. Differences were observed regarding not only the description of facts, but also the definition of social categories involved in the event, as well as of the values and goals presented as typical of those categories.

An example may be found in Reicher and Hopkins' 1996 study on how a famous miners' strike in England was reconstructed from opposite perspectives by political leaders Margaret Thatcher (Conservative Party) and Neil Kinnock (Labour Party). In her interpretation of the events, Mrs. Thatcher created a picture in which miners that did not participate in the strike, as well as their relatives, were attributed personality traits such as courage and determination that were presented as typical of authentic English people. Mrs. Thatcher's definition of the category of strikers, on the other hand, identified them with violent people and even with terrorist groups. Thus, through her discourse Thatcher evoked a large ingroup of "no-strikers/authentic English people" on the one hand and a small outgroup of "strikers/terrorists" on the other. She consistently presented the Conservatives' reaction to the strike as a way of defending the nation at large against a minority of destabilizing forces. A similar though opposite argumentative strategy may be found in Mr. Kinnock's interpretation of the same event. He described the strikers as people who wanted to defend their rights against the repressive policy of Mrs. Thatcher. In doing so, Mr. Kinnock compared strikers to all Labour voters and more generally to all the people who were against Mrs. Thatcher, so that the initially small category of strikers ended up being redefined as the large ingroup of "strikers/liberals."

To conclude, the way in which both a political event and the people involved in it are defined is of crucial

relevance to increasing the public's feeling that the politician, the politician's party, and the public belong to the same wide ingroup and share the same goals and values. Politicians try to define the boundaries and the characteristics of social categories in order to mobilize people in a direction that is favorable to them and unfavorable to the political adversaries. Social and political categories are not stable and unchangeable; on the contrary, they are constantly being redefined, and political leaders contribute to their redefinitions through political discourse.

#### 4.5. Political Persuasion

So far, we have focused our attention on how politicians present their political identity and try to shape the public's political identity. Research on political persuasion, however, has been more generally aimed at investigating how politicians' messages may have an impact on the target's (i.e., the public's) political attitudes. The typical research design of this kind of study implies a comparison of the target's attitude toward a given political issue (e.g., the tax system) before and after exposure to a persuasive message regarding that issue (e.g., a political discourse in favor of a reduction of the tax load on wages). Recent studies in the area have focused on the so-called matching effect, according to which the persuasive impact of different types of messages varies as a function of given characteristics of the target. For example, Lavine *et al.* investigated the different persuasive impact that threat-related and reward-related political messages may have on individuals characterized by different degrees of authoritarianism. The messages were intended to persuade young people to vote in presidential elections. While threat-related messages highlighted the possible negative consequences of non-voting (e.g., "Not voting allows others to take away your right to express your values"), reward-related messages highlighted the possible positive consequences of voting (e.g., "Voting enables one to bring about the kinds of public policies he or she believes in"). The change in attitudes toward voting after exposure to either type of message was then measured, comparing participants with high versus low degrees of authoritarianism. Threat-related messages had a high persuasive impact on authoritarian people and a low impact on non-authoritarian people, and the opposite was true of reward-related messages. Thus, the matching effect suggests that tailoring political messages with regard to the target's characteristics (e.g., authoritarianism) may help to increase the persuasive effect of those messages.

## 5. POLITICAL PARTICIPATION

For most people, voting is the only practical way to take part in political life. At the same time, in a democratic regime, it is also the most important one. In this section, we will focus on how voting behavior has been explained, by referring to motivational, cognitive, and social factors that may underlie political choice. Finally, we will briefly hint at more involved forms of political participation.

### 5.1. Voting Behavior

Since the pioneer work by Campbell *et al.* on party identification, that is, on voting as the stable result of an identification process that dates back to developmental age and to family influences, several other explanations have been proposed. As we have seen in Section 1, the extension of rational choice theory from mathematics and economics to politics has led to the view of voting behavior as the outcome of a rational process. In other words, people consider all the information they have regarding the various possible alternatives (parties or candidates), and choose the one that is characterized by the highest possible utility for them. However, ideas of human rationality and utilitarian goals implied by this approach have been seriously challenged by subsequent psychological and empirical research on voting behavior. The main acquisitions of this research are summarized here.

#### 5.1.1. Cognitive Factors

When deciding how they will vote, people do not necessarily consider all the information they have acquired over time about the various parties or candidates. Given that the quantity of information they can pay attention to at a given moment is limited, they tend to focus only on information they perceive as salient and relevant at that very moment. Their decision-making process does not consider all pros and cons of each alternative. Instead, a series of heuristics or simplification strategies is adopted, which allow making a decision through a limited consumption of mental energy. Thus, people base their decision on a limited amount of information they perceive to be relevant at a given moment. Such information may be categorized into the following four main dimensions:

1. *Trait-based voting.* As we have already seen (see Section 2), candidate evaluation is a very good predictor

of voting behavior. The relevance of this dimension is likely to have increased in recent years, due to the growing personalization of politics favored by the media.

2. *Issue-based voting.* Voting choice may be based on parties' political programs or their positions on given political issues, such as social welfare, the tax system, or immigration rules, and people will choose the party or candidate whose positions on those issues are more consistent with their own. According to the so-called consumer model, recent years have been characterized by an increasing tendency for people to base their voting choice on their agreement with a party position on one or more relevant issues. This would lead to a party choice that is related to specific issues and consequently that is not stable over time.

3. *Ideology-based voting.* In contrast with the consumer model, the ideological model assumes that people choose parties or candidates with whom they share not so much a position on a given issue but general beliefs based on shared ideology or values. As the organization of political attitudes around abstract principles is typical of politically sophisticated people, ideology-based voting is likely to be more concentrated among this kind of people.

4. *Performance-based voting.* Voting choice may also be influenced by the evaluation of past performance of a party or a political leader. Actually, representation of the past is often the main source of information people use to make predictions about the future. Not only reference to the actual past, but also mental simulations of how things might have developed differently, significantly contribute to shape people's interpretation of political reality. For example, one may think that the armed intervention of the United States and their allies in Iraq in 2003 was appropriate if the most readily available alternative to the individual's mind was a course of events implying Saddam Hussein's alliance with Al Qaeda and a rise of terrorist attacks in Western countries. On the other hand, one may evaluate the same intervention as inappropriate if to the individual's mind the available alternative was a successful mediation process handled by the United Nations. A negative evaluation of past action of politicians may sometimes lead to strong disenchantment with politics and increase the possibility of success for new emerging or extremist political parties.

While each of the preceding factors has been shown to play a significant role in people's choices, recent research has been developing more complex predictive models of voting behavior aimed at assessing the

relative weight of each factor on voting. Through application of these models, one may be able to assess how much the electoral success of a given party may be attributable to the party leader, to party ideology, to past performance of either the party or the leader, and so on.

### 5.1.2. *Personal Interest*

According to rational choice theory, utility pursued in voting decision corresponds to self-interest, intended as material, economic interest. Recent empirical research in political cognition has reconsidered this issue in a different perspective, focusing attention on political attitudes that are perceived by people as personally important. Issues of domestic policy, such as welfare or taxation, are usually perceived as more personally important than other issues. When asked to express their attitudes regarding issues considered personally important, people reply more readily than regarding other issues. This suggests that attitudes toward domestic policy issues are more immediately accessible to the individual's mind and therefore are more likely to influence voting behavior. The same does not hold for attitudes toward issues of foreign policy, such as defense expenses or non-intervention in war. Though these issues are perceived as more important than domestic issues at a national level, related attitudes are less easily accessible to the individual's mind and therefore are less predictive of voting behavior. Meta-analysis of a number of survey data sets in the United States has confirmed that attitudes on policy issues that are personally important are more stable over time and are more predictive of candidate preferences and voting.

Analysis of what factors may be at the origin of a political attitude's personal importance has shown that self-interest, as defined by rational choice theory, is just one of these factors. Another factor that significantly contributes to raise the personal importance of a political attitude is its relation with one or more core values, values that hold a high position in the individual's hierarchy. This means that, for example, some people may perceive having an efficient public health system as important not only because of the personal advantage implied by such a system but also because solidarity is one of their core values.

### 5.1.3. *Social Factors*

In addition to self-interest and core values, a third factor that has been shown to underlie the personal

importance of political attitudes is social identification with reference groups. As we have seen in Section 3, social identity is that part of an individual's self-concept that derives from awareness of belonging to one or more groups. If voting for a given party is a shared and relevant characteristic of a group one identifies with, voting for that party may become a way of expressing and reinforcing one's group identity. A confirmation of this can be found in research that has investigated psychological antecedents of the so-called regional vote, that is, the prevalence of a given party in a given region or area of a country. For example, Scottish people are traditionally more in favor of the Labour Party: research results by Abrams and Emler show that Scottish people who vote for the Labour Party have a higher Scottish identity than Scottish people who vote for the conservatives. Although people with higher Scottish identity feel they have fewer economic opportunities as compared to people living in England, these people are also more willing to remain in Scotland. These data suggest that voting may be a way of expressing one's social identity, in this case, regional identity. Even more, they suggest that in voting behavior, expression of group identity may challenge, if not overcome, the pursuit of economic self-interest.

In conclusion, empirical research on voting behavior has shown that vote is not the outcome of fully rational thinking aimed at maximizing personal benefits and minimizing personal costs. It is a more complex process that implies selective information processing, reference to values or other abstract principles, and the expression of one's own identification with one or more social groups. All these factors allows one to overcome the so-called voter's paradox described by rational choice theorists, which portrays voting as an apparently contradictory action: The costs implied by the act of voting (e.g., acquiring knowledge of the issue, renouncing a day's leisure) are higher than the benefits the single citizen may expect from that act, since one single vote cannot decide the outcome of the election. Thus, if voting were based on a mere cost–benefit ratio, no single citizen would vote. As we have seen, however, voting decision is not the outcome of a simple cost–benefit ratio but rather of a range of cognitive and psychosocial factors that must be taken into account if satisfactory predictive models of voting behavior are to be developed.

## 5.2. Collective Action

When we shift our attention from voting as the basic level of political participation to more involved forms

of participation, such as demonstrations, sit-ins, or strikes, finding satisfactory explanations of why people get involved in such actions becomes even more challenging. On the one hand, the cost of similar actions may be very high for the individual, in terms of time, money, and energy spent in participation, but also, in extreme cases, in terms of stigmatization or physical damage suffered because of participation. On the other hand, possible benefits deriving from the success of collective action may be enjoyed just as much by people who did not actively participate in that action, which may be one further reason to not undertake the risk of participation. In his article “Collective Action,” Bert Klandermans deals with this issue in detail, focusing on three main motivations that lead people to get involved in collective action: instrumentality, identity, and ideology. Klandermans also offers an articulate definition of collective action, distinguishing it from other types of group actions. Finally, he analyzes a number of factors that play a role in transforming potential participation into actual participation.

### See Also the Following Articles

Collective Action ■ Ideological Orientations and Values  
■ International Conflict

### Further Reading

- Abrams, D., & Emler, N. (1992). Self-denial as a paradox of political and regional social identity: Findings from a study of 16- and 18-year olds. *European Journal of Social Psychology*, 22, 279–295.
- Boninger, D. S., Krosnick, J. A., & Berent, M. K. (1995). Origins of attitude importance: Self-interest, social identification, and value relevance. *Journal of Personality & Social Psychology*, 68, 61–80.
- Bull, P., Elliot, J., Palmer, D., & Walker, L. (1996). Why politicians are three-faced: The face model of political interviews. *British Journal of Social Psychology*, 35, 267–284.
- Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American voter*. New York: Wiley.
- Caprara, G. V., Barbaranelli, C., & Zimbardo, P. G. (1999). Personality profiles and political parties. *Political Psychology*, 20, 175–197.
- Castano, E., Yzerbyt, V., & Bourguignon, D. (2003). We are one and I like it: The impact of ingroup entity on ingroup identification. *European Journal of Social Psychology*, 33, 735–754.

- Catellani, P., & Quadrio, A. (1991). Ideal and real in the representation of politics. *Revue Internationale de Psychologie Sociale*, 4, 231–256.
- Catellani, P., Milesi, P., & Crescentini, A. (in press). One root, different branches: Identity, injustice, and schism. In B. Klandermans, & N. Mayer (Eds.), *The psychology of right-wing extremism in Europe*. London: Routledge.
- Downs, A. (1957). *An economic theory of democracy*. New York: Harper & Row.
- Duck, J. M., Hogg, M. A., & Terry, D. J. (1995). Me, us, and them: Political identification and the third-person effect in the 1993 Australian federal election. *European Journal of Social Psychology*, 25, 195–215.
- Gaertner, S. L., Rust, M. C., Dovidio, J. F., Bachman, B. A., & Anastasio, P. A. (1994). The contact hypothesis: The role of a common ingroup identity on reducing intergroup bias. *Small Group Research*, 25, 224–249.
- Granberg, D. (1993). Political perception. In W. J. McGuire, & S. Iyengar (Eds.), *Explorations in political psychology* (pp. 70–112). Durham, NC: Duke University Press.
- Haslam, S. A. (2001). *Psychology in organizations. The social identity approach*. London: Sage.
- Krosnick, J. A., & Brannon, L. A. (1993). The media and the foundations of presidential support: George Bush and the Persian Gulf conflict. *Journal of Social Issues*, 49, 167–182.
- Lavine, H., Burgess, D., Snyder, M., Transue, J., Sullivan, J. L., Haney, B., & Wagner, S. H. (1999). Threat, authoritarianism, and voting: An investigation of personality and persuasion. *Personality and Social Psychology Bulletin*, 25, 337–347.
- McGraw, K. M. (2000). Contributions of the cognitive approach to political psychology. *Political Psychology*, 21, 805–832.
- Miller, A. H., Wattenberg, M., & Malanchuk, O. (1986). Schematic assessment of presidential candidates. *American Political Science Review*, 80, 521–540.
- Reicher, S., & Hopkins, N. (1996). Self-category constructions in political rhetoric; an analysis of Thatcher's and Kinnock's speeches concerning the British miners' strike (1984–5). *European Journal of Social Psychology*, 26, 353–371.



# Postmodernism in Psychology

Teresa Cabruja Ubach

University of Girona, Girona, Spain

1. Introduction
2. Contextualization: Issues of Postmodern Thought as a Critique of Modern Rationality
3. Postmodernism Comes to Psychology: The Crisis of Modern Psychology and Its Different Theoretical and Practical Proposals
4. Implications for Traditional Psychological Topics Such as Emotion, Attitudes, and Memory
5. Implications for Psychological Fields and the New Role of the Psychologist
6. An Example: Implications of the Postmodern Turn on the Study and Conception of Identity and Gender  
Further Reading

**enlightenment** Characteristics of psychological knowledge, producing changes on studied objects that have created this knowledge.

**postmodernism** Movement or set of ideas that implies an epistemological and social change with respect to the basic principles of the modern rationality.

**social construction** Idea that human and nonhuman objects do not exist independently of human actions and their linguistic practices.

Postmodernism can be understood as the critique of modern scientific rationality. It may be thought as a loss of belief in truth, in universalistic ideas of historical progress, in unitary subjects, and in objective scientific method. This also means incredulity regarding the great narratives of legitimacy. Postmodern thought contains diverse interpretations and proposals that replace a positivist method of knowing the “real world,” by local and engaged accounts of reality with focus on social and linguistic construction. The influence of these ideas can be found in psychology, as in any other discipline. Here, it reflects, among others things, shifting conceptions in the main epistemological and methodological topics that comprise the field of psychology. Postmodernism can be found overall in the social constructionist approach (Gergen, Ibáñez, etc.), but also in others’ approaches. This article introduces the general themes in postmodern thought, and its bases in French poststructuralist philosophy, feminist studies, and postcolonial studies. All these have had their effects on psychology and we shall see

## GLOSSARY

**critical psychology** Psychological theories that try to substitute the study of psychological processes used in positivist psychology for a study based on language and social practices.

**cyberpsychology** An approach that allows thinking about technoscience, cultural theory, and political effects.

**deconstruction** Action to disassemble structures that are supposed to be well known in order to know the non-explicit way in which they have been organized and the premises that construct them.

**discourse analysis** Variety of forms developed by linguistics, philosophy, and psychology to analyze both social linguistic practices that produce social meaning and the effects of the discourse.

how its principal topics of study such as emotion, attitudes, memory, self, and gender have all been challenged.

## 1. INTRODUCTION

During the 1970s, a “crisis” in psychology reflected a more general crisis in the social sciences: the crisis of the paradigm of positivist modern rationality as a privileged form of knowledge of the human being. This debate was inspired by French poststructuralist philosophy (Foucault, Derrida, etc.), feminist studies (Kristeva, De Lauretis, etc.), and postcolonial studies (Gayatri-Spivak, Bhabha, etc.). The postmodern thought was represented by Lyotard, Baudrillard, Haraway, Bauman, and Butler, among other thinkers.

This crisis emerged from the recognition that a general openness to change had, by the 1970s, precipitated transformations in both the academic and the personal spheres under the influence of social movements such as feminism, ethnic movements, and the political, social, and cultural transformations that drove the new technologies.

Events of the 20th century have prompted several questions about truth, science, reason, and the concept of reality. Within the social sciences, a critique has been developed, based on the questioning of the basic dichotomies of Western modern thought such as reality–fiction, nature–culture, emotion–reason, objective–subjective, and mind–world.

Mainstream psychology, until the end of the 1970s and the beginning of the 1980s, had followed the dominant paradigm of the natural sciences. It put its faith in the hypothetical-deductive method, a methodology for human research that separates science, on the one hand, from “the world” on the other. The fundamental ideas that constitute this form of knowledge are objectivity and neutrality. These are supposed to produce knowledge that is both universal and predictive. Nevertheless, in different fields of psychology, doubts have arisen as to whether objectivity and neutrality can handle complex reality. These reframings suggest that we see psychological knowledge as impregnated of values and “experimental emptiness.” In psychology, this is signaled, for example, in the emphasis on the Western cultural, gender, and sexual values as individualism, ethnocentrism, logocentrism, and androcentrism that are inherent to psychological practice and research. Psychology is challenged in its claims to objectivity and neutrality and in its belief that the world and the subject are independent from each other.

The aim of change can be seen clearly in the works of Gergen, in his treatment of psychology as history, in the socioconstructionist paradigm, and in the analytic critiques of experimental psychology. Challenges to traditional psychology also exist in Latin America with the development of a more reflexive and engaged communitarian psychology. It is important to emphasize that of the several ways aimed at changing the traditional psychological form of thinking and doing research, only some are directly acknowledged as postmodern. Other are recognized as “critical psychology” (among other terminologies).

General topics that are often taken for granted in psychology have begun nowadays to be deconstructed. In other words, this postmodern turn shows how the system of representation of the world is not free of values and can operate by promoting or oppressing certain forms of life or personhood, within the hierarchy of power relationships.

## 2. CONTEXTUALIZATION: ISSUES OF POSTMODERN THOUGHT AS A CRITIQUE OF MODERN RATIONALITY

### 2.1. Postmodernity and the “Postmodern Condition”

Postmodernity is a controversial term. The French philosopher Lyotard claimed that it neither defines nor identifies a period or a historical time. Postmodern rather serves to indicate that something is in decay in modernity, in the universality of the philosophical discourse of the enlightenment. Terms and their opposites begin to lose their sense and enter the play of auto-reflexivity, namely, the differences between fact and fiction, language and its objects, art and “kitsch,” as well as the limits between styles, disciplines, genres, etc.

How can postmodernity be understood? Postmodernity can be found in the goal, common to literary theory, philosophy, and the social sciences, to reveal how it is that positivist dominant discourses are able to create “illusions” of theoretical coherence and empirical credibility from certain institutional conditions and with consequent effects of power.

The post-structuralist, feminist, and post-colonial contributions constitute a source of inspiration for the “postmodern turn” as well as the new sociology of science and neopragmatism. The philosophical base of the modernity–postmodernity debate is found in

French poststructuralist thought in its various forms. In particular, the work of Foucault on matters such as power, madness, sexuality, prisons, and subjectivity develops the thesis of the relationship between power-knowledge-truth and the political, historical, and rhetorical character of psychological sciences.

One of the key methods of postmodernity is the deconstruction of Derrida and one of his main ideas: that language does not have any sense or truth, directly or literal, and its relation with social institutions and hierarchical and power relations. A continuous self-reflexivity is proposed. One consequence is to consider scientific knowledge as a form of knowing the world that is not beyond any other human activity.

What is the postmodern condition? According to Lyotard, it is understood as the generalized acknowledgement of the exhaustion of reason and is based in the crisis of “representation.” That is, it includes the crisis of the social and political movements and discourses, as they are constructed from modern rationality, as well as the transformations in social relations, changes of production, and the global culture of new technologies.

## **2.2. Three Main Ideas from Postmodern Thought**

To be precise, the postmodern critique of modern scientific rationality involves the following three ideas.

1. The critique of great narratives that legitimize a multitude of social actions, erected as discourses of “truth,” and of the linear and evolutionary ideas of progress and history inherent to the modern project. Instead, a local and committed knowledge is proposed.

2. The critique of the concept of representation that entails the dissolution of the border between the language and its object and of the idea of mind as the location of knowledge. A socially constructed reality is proposed.

3. The critique of the subject of Western knowledge and the vindication of “differences,” and the death of the subject as the center of representation and history, and the recognition of its universalism and androcentrism. The consequences for emancipatory social movements appear controversial. The proposal is to avoid essentialism and universalism and to consider using identity categories of a more temporary and local dimension.

With these three critiques, statements of “objective truths” become constructions of communities with

particular values; interest moves towards the forms in which the world “is represented” and what consequences these representations carry. This will bring about a change in the traditional assumption of the way in which language operates. It will no longer be seen as an instrument for the expression of internal representation of external reality, but rather something that obtains its meaning from its use in social life. What is in play, then, is the notion of objectivity as a legitimacy of the truth about how the world and what people are like.

## **3. POSTMODERNISM COMES TO PSYCHOLOGY: THE CRISIS OF MODERN PSYCHOLOGY AND ITS DIFFERENT THEORETICAL AND PRACTICAL PROPOSALS**

### **3.1. Modern Psychology and the Analysis of Its Own Discourse**

With the postmodern turn, psychology, from the reflexivity and the deconstruction point of view, analyzes its own discourse. It tries to understand the way in which it opts to produce certain knowledge and what relations of power are produced. Social constructionism in psychology requires that we put aside the basic presupposition of modern rationality that, as objectivity, had led psychology to constitute itself, as Tomás Ibáñez indicates, as an extraordinarily authoritarian system. Psychological entities were naturalized as corollaries of the human condition, pressuring it to adapt itself to them and freeing the psychologist from the responsibility of the effects of power over the individual. Thus, the productive character of psychology has made a vocabulary and new knowledge of the individuals possible. But simultaneously, its commitment to a rhetoric of truth has entailed effects of social regulation. An individual’s motivations or attitudes, satisfaction, self-esteem, etc., are the elements that are used in deciding who gets a job, who may look after their children, etc.

To consider, then, that psychology is not an objective science like the natural sciences entails a series of changes. Is the conduct of human beings really predictable and not altered by people’s knowledge of themselves, or by the continuous transformations of the world around them? Kenneth Gergen calls this effect enlightenment.

### 3.2. The New Conception of Psychology: The Postmodern Psychology

Due to its symbolic character, situated in a concrete society, culture, and history, psychology's theories, from a postmodern point of view, are to be understood as a practice, and its concepts are to be understood in a dialogical form. For example, if a complex, a frustration, a motivation, an attitude, or a phobia exists, it is because psychology segments, differentiates, and compartmentalizes certain phenomena and gives them a name. It divides populations and their actions into normal and pathological, or normal and abnormal. The possibility of decoding an interview or an observation of a child in a daycare center depends on training that has separated out different behaviors, compared them, named them, and classified them. In other words, it has made available to the people trained in this discipline the authority to carry out the actions that correspond to this domain.

An important change is to stop conceiving the concepts developed by psychology as individualized and as the result of cognitive or cerebral operations. Rather, they should be thought of as relational and evaluative, or in other words, as ways to understand the world and to agree or disagree with it, or to act on it or cause reactions in it. In addition, this way of thinking about psychology totally changes the form of acting and producing personal and social transformations. In the field of psychotherapy, as in educational or social psychology, it leads to specification of from where and why the psychologist wants to intervene, not from an aseptic and neutral form, but rather from an engaged and participative position.

Postmodern psychology replaces a concern for the subject of investigation and the adequacy of theories with an interest in the procedures and instruments used for the study of the socially constructed subject and the way in which we relate to it. And, above all, it is interested in the social conditions that give rise to such a subject and its relations. The concepts inherited from positivist psychology are reanalyzed, and the "great narrative" of psychology is replaced with local knowledge. The attempt to find universal laws is dropped in favor of circumscribing knowledge to certain socio-historical conditions of production, susceptible to change. It has been suggested that the professional practice of psychology may get along well with postmodern thought, since it focuses on "situated" knowledge of human activity, the everyday world, and validation through practice; dynamic, dependent on context, and fragmentary.

### 4. IMPLICATIONS FOR TRADITIONAL PSYCHOLOGICAL TOPICS SUCH AS EMOTION, ATTITUDES, AND MEMORY

From postmodern psychology's point of view, our conception of self, science, representation, and language influences the way that the human being has been studied. This implies a deconstruction of the typical themes of the discipline. Identity, emotions, gender, attitudes, and memory are some of the subjects that have been reconsidered along the lines previously mentioned.

#### 4.1. Bases for Reconsidering the Main Topics of Psychology

1. Proposing a less individualistic and more collective or relational conception;
2. introducing a dialogical perspective, a more contextual and a non-universalistic or essentialist analysis;
3. being conscious of one's own prejudices of interpretation;
4. recovering the role of language not as a simple transmitter but as constructivist;
5. introducing the creative, resistant, and transforming aspects of subjects; and
6. introducing a dimension that defines the moral, evaluative, political, and functional character that must lie in any object of analysis within a certain social order and intersubjective relation.

#### 4.2. Examples: The Study of Emotions, Attitudes, and Memory

In the study of emotions, they are not seen as responses or reactions, but as a part of a contextual representation, not in the intentional sense, but rather as significant in each different situation. In this sense, the emotions are considered as having performativity and pragmatic meaning, showing how the individual understands their context and how they express moral valuations.

Another example is the study of attitudes, almost always considered to be prior to performance, and locatable. But they become understandable as dilemmatic and challengeable constructions, dialogical in character, and as possible *a posteriori* elaborations of our behavior. This way, when we try to understand an event, we relate it to others or to ourselves, resorting to

this cause–effect relation in which we have been socialized. Thus, perhaps we looked for motives and reasons that were not always there before, and which are perhaps reconstructed in some cases *a posteriori*, and that are more applicable to current circumstances than an expression of a previous reason or idea.

In the study of memory, individual memory is not considered as an individual cognitive property, but rather in relation to its function and its use in a certain context. In the production of a particular memory, what is more important than its activation or configuration is the moment and the other people involved. This way, the details, the precision, or the changes in narration are not established by the ability of memory to reproduce reliable truth, but in its relative performativity at the time of production. That is to say, it serves to say something about the individual or to obtain something of the others, and obeys a system of expression of values, beliefs, and functions. Something similar happens in the case of collective memory that is considered in relation to its historic and political role in a concrete socio-economical and political context, with a sense only within a political, economic, institutional and communicative relationship. The implications, for example, in individual or family therapy's ability to solve conflict or trauma are clear, as are the implications in situations of intergroup conflicts and in institutions.

## **5. IMPLICATIONS FOR PSYCHOLOGICAL FIELDS AND THE NEW ROLE OF THE PSYCHOLOGIST**

### **5.1. Implications of the Postmodern Turn for Psychological Fields**

What does this postmodern movement imply for treatment and therapy, for educational, environmental, psycho-social, and organizational intervention, or for the resolution of problems, mediation, and so on? That is to say, what does it imply for applications other than psychology? Since, as we have commented, from the postmodern psychology point of view, any type of individual or group psychological intervention is not considered merely an “application” of the theory, the psychologist cannot be placed in an external relation—as could happen with other sciences—and the values that guide an application and the directions of the transformation need to be specified.

As examples of methods, even if they do not have their origins in postmodern thought but are developed

or used by postmodern psychology, we have discourse analysis, with a variety of schools and developments, which is centered on identifying the discursive practices in relation to their functions and pragmatic effects, and conversational analysis, with its emphasis on the interpretation and meaning of the participants in a concrete context and in its institutional background. Both constitute bases for the study of daily interactions and institutional contexts in research, therapy, education, racism, etc., to transform concepts and methods of work into the day-to-day practice of the psychologist.

An example is the communitarian psychology in Latin America, based on the Action-Research of Lewin and ethnographies that have developed the theology of liberation or the socioconstructionist movement. Its basic characteristic is to obtain a more dialogical relation of a horizontal nature. Another example is in psychotherapy, with the reflective teams, inspired by Andersen and White, in which interdisciplinary teams of psychologists and relatives use these developments to jointly find formulas of dissolution or changes of problematic aspects or in work on the psychologist-client dialogue developed by socioconstructionist therapy. In family therapy, counseling, and individual psychotherapy, the emphasis moves away from the individualized problem and is related to cultural and institutional practices, opening therapeutic work to social justice and to larger sociopolitical and relational contexts. Works is also done on what is defined as “problematic,” to reconstruct it from alternative forms or, even, deconstructing diagnoses to make possible different solutions.

### **5.2. Implications for the New Role of the Psychologist**

To summarize, some of the proposals of postmodern psychology to equip psychology with a less authoritarian and more democratic, plural, and participative position are related to the new role of the psychologist.

1. The role of the psychologist is understood to be that of an agent who facilitates personal transformation and social change, while being aware of the values and interests of professional performance.

2. The activity of the psychologist depends on the participation of the people involved in the investigation or the intervention, co-producing the obtained knowledge and not holding a privileged position as expert. The procedure should include, for example, the people involved in the decision making and the

evaluations before, during, and after treatments and interventions.

3. The psychologist should be sensitive to diversity and differences as well as the plurality of opinions, situations, etc.

4. A role is developed to commit with its ideas by its professional action, which makes explicit and does not hide under neutrality or aloofness.

5. The psychologist develops a less authoritarian intervention in respect to the status that psychology has in our society, beyond the pursuit of the deontological codes of the professional practice.

6. The psychologist incorporates the personal or communitarian explanation of what happens.

7. The psychologist should incorporate a constant reflexivity in all actions and professional practices in order to question and to query the methods of investigation or intervention.

8. The psychologist should develop and apply methods and techniques that incorporate parts of this program.

## **6. AN EXAMPLE: IMPLICATIONS OF THE POSTMODERN TURN ON THE STUDY AND CONCEPTION OF IDENTITY AND GENDER**

### **6.1. The Deconstruction of the Subject in the Postmodern Perspective and the Changes with Regard to Identity**

In fact, this critique of psychology has been made by a great number of psychologists influenced by symbolic interactionism, ethnomethodology, poststructuralism, and deconstruction, who have questioned the individual/society dualism, the idea of the mature, autonomous, and independent person who is seen as an agent planning his or her life, and who has values in virtue of which he or she makes choices (the self-sufficient individual). Consequently, the idea of evolution or development of modern rationality that impregnates a great part of evolutionary theories of personality or child development is replaced by more relational options and different understandings. The works of Foucault on the technologies of the self, developed by Rose, Walkerdine, Sampson, etc., show that the knowledge of our individuality (for example the search for “our own identity” or the wished-for social objectives, such as well-being,

happiness, efficiency, etc.) has been constructed in relation with the use of the capacities of citizens and also for its regulation. The idea of a “saturated self,” as Gergen proposed, as the result of a social saturation process, with the new technologies and the variety of human relationships, leads to a condition in which doubts are instilled about who and what we are. This is not the result of our personal essence: real feelings, deep beliefs, etc., but of how we are constructed.

In particular, these developments imply that we should consider a more relational and dialogical “self” and the replacement of individualized theories of identity with more relational theories. A consequence is a work with a symbolic plurality of meanings with its respective positions in networks of power, shared or constructed between different people, present or not. When constituting intersubjectivity in the interaction, “self” can be understood discursively as a “text,” constructed by a multiplicity of narratives, about how we relate ourselves, full of other stories that we receive explicitly and implicitly throughout our life, among which it is possible to register not only those of human actors, but also those of nonhuman (nature, for example), as proposed by “actor-network theory.” The metaphor of the text is used to talk about the identities in a continuous process, in an intimate relationship with the social institutions and practices. One works with the change from applying different techniques to narrate in another form the events, and thus to transform them.

### **6.2. Cyberpsychology and the Deconstruction of Gender**

The idea of biotechnological self, grounded in the cyborg, an organism that the anthropologist Donna Haraway developed in her study on primates, is an organism made up of technology, biology, culture, and politics, and is used to destabilize the essential differences between body and mind, body and machine, and masculine and feminine. Also, it brings closer another form of new identities that start from, for example, the use of prostheses. Additionally, we can find changes for the gender relations as in cyberpsychology. Cyberpsychology allows us to work with technology and subjectivity, and in politics, feminism, and technoscience with transgressive identities. In fact, a postmodern feminism, as psychologists Raquel Hare-Mustin and Jeanne Marecek suggest, is brought about through a deconstruction of the polarity of sex and

gender, to go beyond the different previous totalitarian theories while maintaining a feminist critique on the effects of how psychology has constructed sexual and gender difference.

### See Also the Following Articles

Attitude ■ Measurement ■ Attitudes ■ Cyberpsychology ■ Emotion ■ Gender and Culture ■ History of Applied Psychology, Overview

### Further Reading

Gergen, K. J. (2000). *An invitation to social construction*. London: Sage.

Gergen, K. J. (1994). *Realities and relationships: Soundings in social construction*. Cambridge: Harvard University Press.

Gergen, K. J. (1991). *The saturated self: Dilemmas of identity in contemporary life*. New York, Harper Collins.

Gordo-López, A., & Linaza, J. (1996). *Psicologías, discursos y poder*. Madrid: Visor.

Gordo-López, A., & Parker, I. (1999). *Cyberpsychology*. London: Sage.

Hare-Mustin, R., & Marecek, J. (1990). *Making a difference. Psychology and the construction of gender*. London: Yale University Press.

Henriques, J., Hollway, W., Walkerdine, V., & Venn, C. (1984). *Changing the subject: Psychology, social regulation and subjectivity*. London: Methuen.

Ibáñez, T., & Iñiguez, L. (1997). *Critical social psychology*. London: Sage.

Ibáñez, T. (1994). *Psicología social construccionista*. Guadalajara: Universidad de Guadalajara.

Kvale, S. (1992). *Psychology and postmodernism*. London: Sage.

McNamee, S., & Gergen, K. J. (Eds.) (1992). *Therapy as social construction*. London: Sage.

Middleton, D., & Edwards, D. (1990). *Collective remembering*. London: Sage.

Montero, M. (1994). *Construcción y crítica de la psicología social*. Barcelona: Anthropos.

Nicholson, L. J. (Ed.) (1990). *Feminism/Postmodernism*. New York: Routledge.

Parker, I. (1999). *Deconstructing psychotherapy*. London: Sage.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology*. Sage: London.

Rose, N. (1989). *Governing the soul. The shaping of private self*. London: Routledge.

Shotter, J., & Gergen, K. J. (Eds.) (1989). *Texts of identity*. London: Sage.

Shotter, J., & Parker, I. (1990). *Deconstructing social psychology*. London: Routledge.

Simons, H. W., & Billig, M. (Eds.) (1994). *After postmodernism*. London: Sage.

Walkerdine, V. (Ed.) (2002). *Challenging subjects: Critical psychology for a new millennium*. New York: Palgrave.

Wetherell, M., Taylor, S., & Yates, S. J. (2001). *Discourse theory and practice: A reader*. London: Sage.



# Postsecondary Education Students, Counseling of

Kathleen L. Davis and P. Gary Klukken

University of Tennessee, Knoxville, Tennessee, USA

1. Overview: The Need for Counseling
  2. Problems That Interfere with College Goals
  3. Counseling Resources
- Further Reading

## GLOSSARY

**adjustment** Psychological processes and behaviors that people use to manage or cope with the demands and stressors of everyday life.

**anxiety** Feelings of excessive apprehension and worry often accompanied by physical disturbances such as racing heart and sweating.

**attention deficit/hyperactivity disorder (AD/HD)** A mental disorder characterized by difficulties in maintaining attention with or without hyperactive behaviors.

**bipolar disorder** A mental and physical disorder characterized by very low mood with intermittent high mood swings; it was previously called manic-depressive disorder.

**depression** A mental and physical disorder characterized by mood disturbances such as persistent feelings of sadness and despair with a lack of energy.

**Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)** The classification manual used by health professionals to identify people's mental disorders.

**stress** Environmental demands made on individuals that may exceed or tax their ability to adapt, cope, or adjust.

Students engaged in postsecondary education often are confronted with stressful situations. Although many students adjust well to the college situation and have the coping mechanisms in their repertoires to successfully negotiate/navigate the stressful events on their own, some students do not. Postsecondary institutions recognize the potential for students' problems that may interfere with institutions' and students' educational endeavors by offering assistance, primarily in the form of counseling. This article reviews normal development and stressors usually encountered by college-age or older adults. It also recognizes many problems with which students present when seeking counseling.

## 1. OVERVIEW: THE NEED FOR COUNSELING

The majority of high school students planning to continue their education tend to enter a university, community college, junior college, or technical school during the fall immediately following their high school graduation. However, there is also a growing population of high school graduates who enter a postsecondary institution after being employed for a number of years. This article focuses on students seeking postsecondary education at a university or 4-year college, the

problems that interfere with students reaching their goals, and counseling services offered on campus.

### **1.1. The Role of Stress and Adjustment**

Students who begin immediately after high school graduation and those who resume their studies after a hiatus from formal education naturally will encounter various sources of stress. Stress may be experienced in relation to external events (positive and/or negative) or to internal psychological and developmental issues. During the late 1960s, Holmes and Rahe developed a list of traumatic life events that tend to cause stress in most people. The list of events included things such as changes in schools, living conditions, social groups, and work hours or conditions. Many of these events occur when students begin or are engaged in their postsecondary education. Students who move away from home or lose their support groups in other ways must adjust daily to new situations, including different people (e.g., roommates, various campus groups). Some must split their time between work and school. All experience pressures to succeed in the academic environment. These pressures may be self-imposed or a result of parental expectations. Additional pressure to maintain high grades is placed on scholarship recipients. Many students must reevaluate their view of self in regard to their strengths and skills. Individuals who were identified as extremely successful, academically, politically, and/or athletically, might now be among hundreds of students with similar skills, achievements, and abilities. For someone who has been successful in the world of work and “the one in charge,” there now is a need to make conceptual adjustments to being in the student role that often requires more passive behaviors. These changes in environment often induce stress in many individuals. Therefore, in response, students must make adjustments and learn new coping behaviors.

Developmental theorists, such as Erikson and Marcia, suggest that identity concerns are very prominent and normal during late adolescence (i.e., college age). This period is one of questioning and exploring personal values. Values may be examined within the context of areas such as religion, gender, sexual orientation, and potential careers. In addition, as college students enter early adulthood, the quest for intimate relationships becomes paramount. The search for identity and intimate relationships is potentially stressful and anxiety producing, but it is part of the normal developmental process.

Although these stressors are to be expected, individuals often need assistance in learning to adjust or cope with the external and psychological stressors associated with college. There are many ways in which people respond to stressors. How individuals adjust or fail to cope with stressors is dependent on how they appraise and what meaning they attach to the demands and pressures associated with the potential stressors. The perceived degree of threat to one’s well-being and coping abilities is related to each individual’s previous successful or unsuccessful coping behaviors and experiences. For example, two people may assess attendance at a social event differently. One may be excited and look forward to a wonderful evening, whereas the other may view an evening interacting with others with trepidation. In general, individuals tend to respond to stressful events by developing preferred coping methods or various degrees of maladaptive reactions. It is important that students recognize when they can benefit from assistance in dealing with academic, relationship, and/or personal stressors. Students may seek counseling as a means of learning new ways of dealing with new situations or in recognition of their own inability to adjust effectively to this developmental period and the college experience. Counseling can be preventive, developmental, or remedial.

Prevention or outreach activities of counseling psychologists may entail services such as making a presentation to a dormitory, fraternity, or sorority on the identification and avoidance of date rape or eating disorders. Examples of developmental activities include offering support groups on improving social skills and on learning methods for time management. Training others (e.g., resident hall personnel) on communication skills and peer counseling may be preventive as well as developmental. When a student or others close to the student believe that a problem exists, remediation is usually the typical intervention.

### **1.2. Influence of Diversity**

In addition to the stress and adjustments with which college students must deal, students also bring a diversity of experiences and backgrounds to this new setting. Increasingly large numbers of students are not in the traditional 18- to 22-year-old age range; are from diverse ethnic, racial, and cultural backgrounds; and bring backgrounds of widely varying experiences to the college classroom. Their family systems may range from close and supportive to demanding and

abusive. Their cultural backgrounds may or may not be supportive of higher education. They may have lived on their own or never left home. They may have worked in significant jobs or had careers before going back to school. These students bring new expectations, new goals, and new issues. As a result, campuses are no longer islands of simple intellectual pursuit and career development (if they ever really were); instead, they are dynamic and demanding social structures that must be negotiated on a daily basis.

## **2. PROBLEMS THAT INTERFERE WITH COLLEGE GOALS**

The issues of stress and adaptation appear to have been increasing in intensity over the past decade or so and have been the subject of considerable discussion among higher education mental health professionals. This section briefly examines some of the major problem areas resulting from this increase.

### **2.1. Academic Problems**

Academic problems are common among students. This is the area involving the job of students, that is, to learn new information, new concepts, and new skills. This is also the area for which there is constant evaluation of success or failure. Most students go on to postsecondary education because they did well in prior academic settings. However, being successful in secondary education can happen for a number of reasons, not all of which are useful at the next level. Being academically successful due to easy classes, special privileges, extracurricular renown, or any other basis that resulted in background deficits is a major detriment to college success. Following are some of the most common academic problems and the application of psychology to their resolution.

#### **2.1.1. Study Skills**

Many students have not learned how to study. They are usually very intelligent and have used their ability to memorize quickly and/or to write convincingly to succeed in high school. When faced with increasingly difficult material at a college level and competing against equally intelligent classmates, they discover that the tasks of taking good notes, writing college-level papers, and studying for difficult exams is very

different from what they were used to doing in high school. Classes and workshops in study skills are offered at most schools using information and techniques developed as a result of psychological studies in human learning. Some of these include overlearning, multichannel information acquisition, and note-taking organization.

#### **2.1.2. Time Management**

Many people are very poor time managers, and students are no exception. Choosing what to do and when to do it is a constant decision-making process and is a skill that many students do not have. Students begin to feel overwhelmed and out of control of their lives as academic demands increase in the context of everything else in their lives. Many schools offer training in time management using psychological techniques for values clarification, task organization, and assertiveness training.

#### **2.1.3. Student Role Conflicts**

As mentioned previously, the role of students is to learn. This role can be especially difficult for returning students who may have spent years working in jobs, directing other people, and/or being experts in their fields. Although the return to school may be the fulfillment of their dreams, the return to the role of students may be much more difficult. They quickly discover that students are not in charge of the curriculum, do not get to choose their professors' teaching styles, are not experts in every field, and do not get to decide the merits of their own writing styles. In addition, they learn quickly that the way in which they learned to do something on their jobs might not be acceptable in their academic programs. This readjustment of roles may require individual or group counseling focusing on issues of self-identity, motivation, and role flexibility.

### **2.2. Career Decision Problems**

For much of human history, there was not a lot of choice in careers. Usually, an individual's occupation was determined by what his or her parents did as an occupation. Although this process did not provide much freedom of choice, it also eliminated the stress of decision making. When a different occupation was chosen, it was usually based on community need. That is no longer the case in our current society. Each year, there are new

possibilities for careers and old careers may disappear. The result is that decision making about one's career has become difficult and stressful. Following are two types of career decision-making problems.

### **2.2.1. Undecided Students**

There are students who come to school knowing what they want to do for careers and who then continue on into those careers. These are a minority of all students. Many students come to school with no idea of what they eventually want as careers. Others come with ideas or even clear goals and then change their minds. Most students change their intended majors several times over their college careers. Some of these decisions are a result of developmental changes that redirect their interest, whereas others are pragmatic decisions based on classes that they discover they really enjoy, do not like, or cannot pass. There are a number of career development theories about how people make these important life decisions, and counselors will often spend time with students using psychological tests to assess their strengths and interests, provide them with information about employment opportunities and forecasts, and help them to evaluate their past successes as indicators of future possibilities.

### **2.2.2. Indecisive Students**

There are many students whose problems with career choices stem from difficulties with decision making of any kind. Because their career choices are such significant decisions, their indecisiveness may completely stop their progress, eliminate their motivation, increase stress, and cause them to change their majors whenever something new attracts their attention. Counselors are able to make use of knowledge about the fear of failure, the fear of disappointing loved ones, and other unreasonable expectations that may stop students from making decisions. The principles of cognitive-behavioral psychology have very direct application to this problem.

## **2.3. Personal Problems**

All individuals bring their own histories, their own traits, and their own problems with them wherever they go. College is a place where stress can amplify these issues to such a degree that they can interfere with life, sometimes in a serious way. Following are

some of the typical problem areas and the application of psychology in their amelioration.

### **2.3.1. Relationship Problems**

Problems with relationships are perennially one of the major concerns for college students. Most students are involved in, or at least concerned about, their search for life partners. Consequently, social relationships are a significant part of their world. They may be dealing with roommates for the first time in their lives and working out the day-to-day adjustments of living with other people who are not family members. They may be in intimate relationships where they must make decisions, set boundaries, and take risks all on their own for the first time. They often hold memberships in various groups such as classroom teams, social organizations, and athletic teams. Here, they must negotiate the more complex social politics of group dynamics. There are also relationships with authority figures that are unmediated by parents or other hometown supporters. Professors have a wide range of personal styles, pedagogical practices, and worldviews. Residence hall personnel bring their own customs and communication styles. Offices of deans and other college administrators have their own agendas and rules. Through this often confusing maze of people, students must successfully negotiate a variety of relationships. It is not surprising that they run into difficulties. This is an area where it is critical to have counselors who are well versed in the application of group dynamics and interpersonal relationships.

### **2.3.2. Family Problems**

Most students are still a part of family units. These special relationships are very powerful, and students' advancement to postsecondary education does not guarantee that these relationships will be trouble free. Some of these family relationships are problematic from the beginning. Many students come to college with histories of abuse. These childhood experiences usually result in long-term damage to self-esteem, interpersonal relationships, and general psychological health. Students with abuse histories may discover their learned wariness of others creates relationship difficulties in many areas of their lives. These problems often require extensive individual or group therapy to explore and work out new patterns of thinking, feeling, and behaving.

Family relationships that have not been problematic might still create considerable stress for students.

Parents do not receive training in how to let go of the more directive role that they have played in their children's lives for many years; however, abandonment of all control is not a healthy solution either. Thus, a new version of old relationships must be developed, often without the mutual agreement of both parties. This is very stressful for both parents and students. It is a developmental issue that everyone must face. Students want more freedom than they had previously (especially those students who are now used to living on their own), whereas parents are often concerned about, or even frightened by, the lack of experience or appropriate behavior exhibited by their sons and daughters. Issues of independence, control, acceptance in a new role, and responsibility for their own lives are very important to resolve as students mature and take their place in adult society. Counseling must be readily available for short-term interventions, with attentive listening mixed with appropriate support, confrontation, and reality testing. Most students survive this stage without the help of professionals, but some outside help at this point can often make the passage smoother and quicker.

### **2.3.3. Alcohol Use and Abuse**

The use of alcohol among students in postsecondary education settings is both a time-honored tradition and a serious mental health problem. Many students use this freedom from parental oversight to increase their drinking behavior. Usually, students do not seek help for problems in this area of their own volition but rather are forced into counseling by concerned others. This may be problematic behavior for only a small percentage of people, but it is a problem that may cost them their academic careers and considerable legal and relationship problems. Although the application of behavioral techniques may be useful here, often more intensive inpatient rehabilitation is needed.

### **2.3.4. Depression**

Perhaps the most prevalent issue in any higher education counseling service is depression. During recent years, this has even exceeded the traditional number one issue of relationship problems. Depression is identified by the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) as having a number of the following symptoms: depressed mood much of the time, diminished interest or pleasure in nearly all activities most of the time, significant change in appetite, significant change in weight, insomnia or hypersomnia,

psychomotor agitation or retardation, fatigue, feelings of worthlessness or excessive guilt, diminished ability to think or concentrate, and recurrent thoughts of death or suicide.

Although there is considerable discussion in the professional community about the prevalence of biochemical depression, it is clear that depression may have either or both a psychological basis and a biochemical basis. The evaluation of a student with symptoms of depression will often include an exploration of current and past stressors, a personal and family history of depression or other psychological issues such as alcoholism, and the current or past use of medication. Depression may have a range of impacts on college students, from simple lack of energy to complete withdrawal from interaction with anyone and serious contemplation of suicide. This can result in academic dysfunction, with concomitant course failures, academic dismissal, and financial repercussions. In addition, interpersonal relationships may deteriorate, social and organizational obligations may be ignored, and physical health may be compromised. Suicide is the most dangerous outcome of depressive behavior; therefore, considerable attention must be devoted to students with depressive symptomatology. Although college students typically commit suicide at a lesser rate than do non-college students of a similar age, it is still a major concern for college mental health professionals. It is important to differentially diagnose depression from bipolar disorder as well as from a fairly large number of potential physical problems. During recent years, there has been an increased use of consultations with medical colleagues as this vital physical/psychological link has become more understood. Many mental health professionals now see that the prescription of psychotropic medication is an effective way in which to quickly handle the crisis experienced by severely depressed students. Usually, the medicating of depression is done in conjunction with psychological interventions, a strategy that appears to have the greatest ameliorative effect. The goal is to keep these students alive, to return them to an acceptable level of functioning, to help them learn more effective ways of handling stress (if it is more than simply biochemical), and to enable them to continue their educational pursuits.

### **2.3.5. Anxiety**

An area of concern that often coexists with depression is anxiety. This may range from minor apprehensions about performance, appearance, and social acceptability

to panic attacks, phobias, and obsessive–compulsive disorders. Although not as life-threatening as depression, anxiety can still become life-disabling. Unfortunately, students with anxiety problems are often overlooked. Because everyone is anxious about something—especially in school—its very commonality means that it can be an invisible plague, damaging performance and interfering with students' quality of life. Evaluation of anxious students involves an exploration of current and past stressors, a personal and family history of anxiety or other psychological issues such as depression, and the current or past use of medication. Anxious students may present themselves with a variety of symptoms such as inability to concentrate while studying or sitting in class, difficulty in sleeping, avoidance of social situations, and compulsive performance of repetitive behaviors. Again, as with depression, coordination of treatment with medical professionals can be the most effective intervention.

### **2.3.6. Attention Deficit/Hyperactivity Disorder**

Attention deficit disorders are relatively new areas of concern for college counselors. Although these issues have been around for as long as there have been learners, only recently have there been concerted efforts to respond appropriately to these problems. These are clearly dysfunctions in the neurophysiology and biochemistry of the individuals concerned. Unfortunately, these problems have been very difficult to measure and document, leaving sufferers with significant deficits and very little remediation. Adult attention deficit/hyperactivity disorder (ADHD) has only recently been clearly recognized, diagnosed, and medicated. Some of the signs include making careless mistakes, being unable to concentrate, not seeming to listen, often losing things, being easily distracted by external stimuli, and being forgetful. Many students who are evaluated as having ADHD have at least one parent who had the same problems and so they thought it was normal. Others have had ADHD identified all the way through school but have parents who have been reluctant to use medication due to side effects, stigma, or disbelief that their children had anything wrong with them. For many of those who have ADHD, the use of medication can literally change their lives, enabling them to concentrate and to actually enjoy learning for the first time.

## **2.4. Environmental Problems**

External events may affect and interfere with students' academic progress. The accidental or intentional deaths

of fellow students can cause distress among a large number of students, even if they are not acquainted with those who have died. Special counseling services dealing with bereavement issues may be necessary for a large number of students at one time. Fear resulting from a series of physical assaults and/or robberies on campuses is a realistic response, yet it is one that affects students' mental functioning.

Rejection of individuals due to their lifestyles (e.g., gay, lesbian, bisexual, transgender), religion, race, or native countries on the part of a large number of students, faculty members, and/or administrators often causes disruptions in students' efficiency. Students facing the decision to “come out”—and to whom—are often conflicted about the best course of action. Recently, the invasion of Iraq caused many Muslim students to fear for their safety on campus. Racial tension on campus often is the result of one group being perceived as receiving more support from the university. Individual or group counseling can be an important resource during these times.

## **3. COUNSELING RESOURCES**

Recognizing the large number of adjustment demands on students, a variety of counseling services are supported by postsecondary institutions. Sources of counseling include individuals, agencies, and organizations. Colleges often employ academic advisers, psychologists, and psychiatrists. In addition, campus ministries have people trained in pastoral counseling, and they are a wonderful source for personal and religious counseling. Resident hall assistants who, like the residents, are also students are a source of peer counseling. Many counseling centers hire counseling psychologists to work with college-age students in the development of more appropriate adjustment strategies and in the alleviation of stress-related symptoms. Agencies such as career service centers, ombudsperson's offices, and student health centers offer specialized counseling that may be preventive and/or remedial in nature. Postsecondary institutions recognize the need for, and their responsibilities in offering, counseling services to their students.

### **See Also the Following Articles**

Alcohol Dependence ■ Attention Deficit/Hyperactivity Disorders (ADHD) ■ Career Counseling ■ Diagnostic and Statistical Manual of Mental Disorders ■ School-to-Work Transition ■ Stress ■ Vocational Psychology, Overview

### Further Reading

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Benton, S. A., Robertson, J. M., Tseng, W., Newton, F. B., & Benton, S. L. (2003). Changes in counseling center client problems across 13 years. *Professional Psychology: Research and Practice*, 34, 66–72.
- Erikson, E. H. (1968). *Identity: Youth and crisis*. New York: Norton.
- Gallagher, R. P. (2003). *National Survey of Counseling Center Directors*. Alexandria, VA: International Association of Counseling Services.
- Garfield, S., & Bergin, A. (Eds.) (1986). *Handbook of psychotherapy and behavior change* (3rd ed.). New York: John Wiley.
- Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, 11, 213–218.
- Marcia, J. E. (1980). Identity in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology* (pp. 159–187). New York: John Wiley.
- Maslow, A. H. (1970). *Toward a psychology of being* (2nd ed.). New York: Harper & Row.
- Mental health: Does therapy help? (1995, November). *Consumer Reports*, pp. 734–739.
- Rathus, S. A., & Nevid, J. S. (1998). *Adjustment and growth: The challenges of life* (7th ed.). Fort Worth, TX: Harcourt Brace.
- Weiten, W., & Lloyd, M. A. (2000). *Psychology applied to modern life: Adjustment at the turn of the century* (6th ed.). Belmont, CA: Wadsworth/Thomson Learning.



# Posttraumatic Disorders

David S. Riggs and Edna B. Foa

University of Pennsylvania, Philadelphia, Pennsylvania, USA

1. Definition of Trauma
  2. Common Reactions to Trauma
  3. Course of Recovery from Trauma
  4. Psychological Disorders Following Trauma
  5. Assessment in the Wake of Trauma
  6. Psychological Treatment of Posttraumatic Disorders
- Further Reading

## GLOSSARY

**acute stress disorder (ASD)** A psychiatric disorder that develops in some people following a trauma; symptoms include dissociation, reexperiencing of the trauma, avoidance of trauma-related cues, and increased arousal.

**arousal** The tendency for individuals who have experienced a trauma to have elevated levels of physiological and emotional tension; this may be manifest as increased agitation, irritability, and/or startle as well as difficulty in concentrating or sleeping and in a sense of hypervigilance.

**avoidance** The tendency for individuals who have experienced a trauma to stay away from reminders of the trauma.

**numbing** A sense of diminished emotional responding commonly seen in trauma survivors; it is sometimes described as an empty or blank feeling.

**posttraumatic stress disorder (PTSD)** A psychiatric disorder that develops in some people following a trauma; symptoms must persist for at least 1 month and include reexperiencing of the trauma, avoidance of trauma-related cues, and increased arousal.

**psychological trauma** The emotional impact of a traumatic event on the person who experienced it.

**reexperiencing** The tendency for memories, thoughts, feelings, and/or images related to a past trauma to intrude into the current day; manifestations of reexperiencing symptoms include intrusive thoughts or images, nightmares, and extreme emotional and physiological distress when reminded of the trauma.

**traumatic event** An event that involves actual or threatened death or injury or a threat to the physical integrity of oneself or others.

**traumatized people** People who have experienced a traumatic event that resulted in psychological trauma.

The psychological aftereffects of traumatic events can be dramatic, and in some cases they can be prolonged. There are a group of common psychological reactions to traumatic events that reflect the overwhelming nature of the trauma. These reactions, when severe or prolonged, may develop into the psychological disorders of posttraumatic stress disorder (PTSD) and acute stress disorder (ASD). This article describes common reactions to trauma, trauma-related psychiatric disorders, common issues in the assessment of trauma and posttrauma symptoms, and treatments for posttraumatic stress disorders.

Psychological reactions following traumatic events have been recognized for hundreds (or perhaps thousands) of years in literary descriptions and for more than 100 years in the psychiatric field. Terms such as war neurosis, shell shock, and battle fatigue reflect the fact that these reactions were largely associated with war. More recently, researchers and

clinicians have begun to examine the psychological impact of a wide range of traumatic events, including rape, physical assault, natural and manmade disasters, motor vehicle accidents, and chronic abuse in various populations such as child abuse, domestic violence, and political imprisonment. This research clearly demonstrates that traumatic events can have devastating and long-lasting effects on the psychological well-being of survivors. At the same time, this research also indicates that humans are remarkably resilient and that, with time, most trauma survivors recover on their own.

## 1. DEFINITION OF TRAUMA

Individuals will label many different events as traumatic. The formal definition of a traumatic event adopted by the American Psychiatric Association in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) consists of two distinct aspects. The first describes the characteristics of the event itself, and the second describes the subjective emotional responses of the individual who experiences the event:

1. The person has experienced, witnessed, or been confronted by an event that involved actual or threatened death or injury or a threat to the physical integrity of oneself or others.
2. The person responds to this event with intense feelings of fear, horror, or helplessness.

The definition allows different types of events to be identified as traumatic and recognizes that some people will experience psychological trauma from a particular event, whereas others will not. Although experiencing traumatic events was once thought to be rare, it is now known that most people are exposed to at least one traumatic event over the course of their lifetimes.

## 2. COMMON REACTIONS TO TRAUMA

The psychological impact of a traumatic event differs across types of events and individuals. Some events (e.g., rape, combat) are more likely to result in long-lasting psychological problems than are others (e.g., motor vehicle accidents, natural disasters). Also, individuals who experience the same event may differ

greatly in their reactions to it. Despite this variation, there is a cluster of reactions that commonly occur following a trauma. Table I presents a group of reactions that many trauma survivors experience. Research on recently traumatized people suggests that nearly all trauma survivors will experience at least some of these reactions. These include symptoms that reflect elevated levels of emotional distress, such as increased feelings of fear and anger and general increased arousal, as well as feelings of guilt, shame, and grief. In addition, shortly after a traumatic event, people often find themselves replaying or reliving the event in their minds. These intrusive thoughts are usually accompanied by feelings similar to those experienced at the time of the traumatic event. The distress associated with the traumatic memories often leads trauma survivors to avoid people or situations that remind them of the traumatic event.

## 3. COURSE OF RECOVERY FROM TRAUMA

Shortly after a traumatic event, most traumatized people experience significant emotional distress that, in the majority of cases, diminishes over time. Although there are individual differences in patterns of emotional recovery from a traumatic event, most people recover within the first few weeks or months after the event. However, a substantial minority experience persistent chronic psychological problems related to the trauma. In some cases, individuals appear to recover relatively well from the trauma, only to have trauma-related problems surface months or years later.

Several factors have been found to distinguish people at most risk for persistent trauma-related problems, particularly those people who develop chronic posttraumatic stress disorder (PTSD). The factors that appear to be related to persistent posttraumatic problems can be loosely divided into (a) characteristics of the traumatic event, (b) characteristics of the individual exposed to the trauma, and (c) characteristics of the environment in which the individual is attempting to recover.

Among the characteristics of the traumatic event that are associated with persistent posttrauma problems is the type of trauma, with sexual assaults and combat being more likely to result in prolonged problems than nonsexual assaults and motor vehicle accidents. Exposure to a more severe traumatic event

**TABLE I**  
**Common Psychological Reactions to Traumatic Events**

<i>Reaction</i>	<i>Examples</i>
Fear and anxiety	Fear of a repetition of the trauma Fear in response to environmental cues reminiscent of the trauma such as people, places, times of day, smells, and noises Anxiety attacks that feel as though they come out of the blue General sense of dread or anxious arousal
Reexperiencing of the trauma	Unwanted thoughts of the trauma Unwanted mental images of the trauma Flashbacks in which the person feels the trauma is occurring again Bad dreams or nightmares
Increased arousal	Feels jumpy, jittery, or shaky Feels irritable or angry Has difficulty in concentrating Has difficulty in sleeping
Avoidance	Attempts to suppress trauma-related thoughts and feelings Avoids trauma-related conversations Avoids situations or activities that are reminiscent of the trauma
Numbing	Sense of emotional numbness or emptiness Detachment from surroundings and other people Depersonalization or derealization
Guilt and shame	Blames self for things he or she did or did not do Is ashamed of what happened during the trauma or by the nature of the traumatic event Is ashamed or guilty about the emotional or behavioral reactions the person experienced during and after the trauma
Grief and depression	Feels down, sad, hopeless, or despairing Has loss of interest in, and decreased enjoyment from, normal activities
Altered image of self and the world	Sees the world as generally or uniformly dangerous Views the self as incompetent or bad in some way Views others more negatively and less trustworthy Is concerned that the person's reactions to the trauma mean that he or she is losing control or "losing my mind"
Disrupted relationships	Has difficulties with intimate relationships Feels distanced or cut off from friends and family
Alcohol and substance use	Increased use of alcohol and drugs to "self-medicate" reactions to the trauma

is also more likely to produce persistent problems. Markers for a severe trauma include physical injuries and the duration of the event. Individual characteristics that are associated with persistent problems include prior exposure to traumatic events and the presence of diagnosable depression or anxiety prior to the event. Among posttrauma factors, the absence of a supportive social environment appears to be associated with persistent problems following a trauma.

Additional markers that predict persistent emotional distress reflect the interaction of trauma and

individual characteristics. For example, the perception that one is likely to die during the trauma, which reflects both the severity of the trauma and the individual's interpretation of the events of the trauma, is associated with persistent posttrauma problems. Similarly, intense emotional reactions (e.g., fear, helplessness, shame, anger) during and shortly after the trauma, as well as severe initial posttrauma reactions (e.g., intrusive memories, increased arousal, dissociation), are associated with more persistent problems.

## 4. PSYCHOLOGICAL DISORDERS FOLLOWING TRAUMA

### 4.1. Posttraumatic Stress Disorder

PTSD was first introduced into the diagnostic nomenclature in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) in 1980. PTSD is classified as an anxiety disorder, but it incorporates three distinct groups or clusters of symptoms: reexperiencing, avoidance, and increased arousal. The reexperiencing cluster consists of five symptoms that reflect distressing spontaneous or cued recollections of the traumatic event such as intrusive memories, nightmares, and emotional distress in response to trauma-related cues. The seven avoidance symptoms reflect attempts at avoiding memories or reminders of the event as well as disruptions in emotional/interpersonal experience such as a numbing of emotional reactivity and a sense of detachment from other people. The increased arousal cluster consists of five symptoms that reflect an elevated level of alertness (e.g., hypervigilance, increased startle) and a high level of general arousal (e.g., sleep disturbance, irritability). To be diagnosed with PTSD, an individual exposed to a traumatic event (as defined by the DSM-IV) must experience symptoms in each of these clusters. Specifically, the person must manifest at least one reexperiencing, three avoidance, and two increased arousal symptoms for at least 1 month after the traumatic event. If the symptoms persist for more than 3 months, the condition is considered chronic.

### 4.2. Acute Stress Disorder

Acute stress disorder (ASD) was introduced as a distinct disorder in the DSM-IV. ASD is characterized as an anxiety disorder and shares many symptoms in common with PTSD. One important difference between the two disorders is the time frame of the symptoms. The diagnosis of ASD requires that the symptoms occur within the first 4 weeks after a traumatic event, whereas PTSD requires that the symptoms persist for at least 1 month after the traumatic event. The other feature that distinguishes the two disorders is a greater emphasis on dissociative symptoms in ASD than in PTSD.

To be diagnosed with ASD, a person must be exposed to a traumatic event and develop symptoms in each of four categories: dissociation, reexperiencing,

avoidance, and arousal. In contrast to PTSD, the avoidance cluster of ASD symptoms consists of only effortful avoidance of trauma reminders. Symptoms of emotional numbing and amnesia are included in the dissociation cluster, and other symptoms (e.g., anhedonia, foreshortened future) are not included. In contrast to the diagnosis of PTSD, ASD requires only one symptom in each of the reexperiencing, avoidance, and arousal clusters.

Initially, it was thought that the presence of ASD would identify people who would go on to develop chronic PTSD. However, recent research suggests that the relationship between acute reactions to trauma and more chronic problems is not so straightforward. ASD does appear to serve as a risk marker for PTSD with many people who develop chronic PTSD meeting criteria for ASD shortly after a trauma. However, some survivors with ASD do not go on to develop PTSD, and some traumatized people who do not have ASD may develop PTSD later. The imperfect relationship between the diagnoses of ASD and PTSD probably reflects the fact that the initial severity (not simply the presence or absence) of symptoms following a traumatic event is related to their persistence. Also, relating the presence of ASD directly to a PTSD diagnosis ignores factors that augment or impair recovery during the first month following a traumatic event. Regardless of its utility as a precursor to PTSD, the ASD diagnosis does capture the significant distress that some people experience shortly after trauma—distress that cannot be diagnosed as PTSD because the symptoms have not yet persisted for more than 1 month (Table II).

## 5. ASSESSMENT IN THE WAKE OF TRAUMA

Assessments of posttrauma pathology focus on three areas: (a) exposure to traumatic events, (b) trauma-related symptoms, and (c) establishment of the link between the traumatic event and the symptoms. Establishing the link between the event and the symptoms is particularly difficult in cases of chronic PTSD where the traumatic event may have occurred years or even decades in the past and in cases where the patient has experienced multiple traumas. Establishing this link is typically less complicated when assessing ASD where the traumatic event must have occurred recently.

**TABLE II**  
**Symptoms of Posttraumatic Stress Disorder and Acute Stress Disorder**

<i>Posttraumatic Stress Disorder</i>	<i>Acute Stress Disorder</i>
<i>Reexperiencing symptoms (need at least one)</i>	<i>Dissociation (need at least three)</i>
1. Intrusive thoughts or memories of the trauma	1. Subjective sense of emotion numbing or detachment
2. Recurrent dreams related to the trauma	2. Reduced awareness of surroundings
3. Flashbacks of the trauma	3. Derealization
4. Emotional distress to trauma reminders	4. Depersonalization
5. Physiological reactivity to trauma reminders	5. Unable to recall important aspect of the trauma
<i>Avoidance (need at least three)</i>	<i>Reexperiencing (need at least one)</i>
6. Efforts to avoid thoughts or feelings of trauma	6. Recurrent trauma-related images, thoughts, dreams, or flashbacks
7. Efforts to avoid trauma reminders	7. Emotional distress to trauma reminders
8. Unable to recall important aspect of the trauma	<i>Avoidance (need at least one)</i>
9. Diminished interest in activities	8. Marked avoidance of trauma-related thoughts, feelings, or reminders
10. Detachment from others	<i>Arousal (need at least one)</i>
11. Restricted range of affect	9. Marked symptoms of arousal such as sleep problems, irritability, difficulty in concentrating, hypervigilance, exaggerated startle, restlessness
12. Foreshortened sense of the future	
<i>Arousal (need at least two)</i>	
13. Sleep disturbance	
14. Irritability or angry outbursts	
15. Difficulty in concentrating	
16. Hypervigilance	
17. Increased startle response	

**5.1. Assessing Exposure to Trauma**

The specific guidelines for assessing exposure to traumatic events depend on the goals of the assessment. This being said, five aspects of traumatic exposure are typically assessed during the evaluation: (a) establishing the presence of trauma in the life of the client, (b) identifying the nature of the traumatic exposure, (c) estimating the severity of the trauma, and (d) establishing the timing of the traumatic exposure.

**5.1.1. Establishing the Presence of Trauma**

Typically, the assessment of traumatic exposure includes an evaluation of traumatic events across the life span. In some cases, the primary focus may be on a single event, but understanding the prior trauma history of the person who experienced the trauma is important.

The presence of traumatic exposure is typically determined by presenting the person with a list of traumatic events and having the person indicate those events to which he or she has been exposed. The specific events included on the list will vary across settings, situations, and clinicians. The most

commonly evaluated traumas include motor vehicle accidents, physical and sexual assaults, physical and sexual abuse (childhood or adult), fires, natural disasters, and combat. It is better to use structured questions that describe each event in detail rather than a single question such as “Have you experienced a trauma?” This helps to ensure that the person’s response does not rely on his or her understanding of a single word such as “trauma.” Similarly, the wording of each item should be designed to reduce the likelihood of misunderstanding or misinterpretation. For example, it is better to ask “Has anyone ever used physical force or threats of force to make you do something sexual that you didn’t want to do?” instead of “Have you ever been raped?” There are many times when a person reports having experienced forced sexual contact but will not view it as rape.

In addition to determining what types of traumatic experiences a person has experienced, assessments should inquire about the number of times various types of events have occurred. For example, the assessor will want to know how many earthquakes the person has experienced or how many times the person has been assaulted.

### **5.1.2. Identifying the Nature of the Traumatic Exposure**

Although traumatic events are thought to share certain characteristics, it is clear that they are not the same in their emotional impact. A more complete understanding of the traumatic event and its potential impact on a person requires a thorough understanding of the specific characteristics of the event and the nature of the person's exposure to that event. The definition of trauma included in the diagnostic criteria for PTSD and ASD allows for three types of exposure to the event: (a) directly experiencing the event, (b) witnessing the event, and (c) being confronted with the event. In general, but not always, directly experiencing a trauma is more likely to produce lasting effects on the person than is witnessing the event or being confronted by it.

Often, a distinction is made between interpersonal traumas (e.g., criminal assaults, child abuse) and other types of traumas (e.g., fires, natural disasters). Research suggests that interpersonal traumas are more likely to produce persistent reactions than are other types of traumatic events. However, this might not be true in all cases.

Traumas have also been differentiated by whether the event happened to a particular individual or to a group of people. Events such as criminal assault, child abuse, and life-threatening illnesses typically happen to only one person. In contrast, events such as natural disasters, wars, and the terrorist attacks of September 11, 2001, affect entire communities simultaneously. Although individual reactions may be quite varied, there is some suggestion that individual traumas are more likely to result in chronic problems than are community traumas. However, the distinction between individual traumas and community traumas is far from perfect. For example, a violent crime may typically be perpetrated against an individual, but the cumulative effect of crime can affect an entire community. Similarly, although a hurricane or another natural disaster may affect an entire community, the specific experiences of individuals within that community may be quite varied.

### **5.1.3. Assessing the Severity of the Traumatic Event**

Two aspects of trauma severity are typically assessed: objective severity and subjective severity. These map loosely onto the two aspects of the DSM-IV definition of trauma in the diagnostic criteria for PTSD and ASD. Measures of objective dangerousness will depend on the specific type of traumatic event, but they reflect

characteristics of the event such as whether someone was killed, whether and how badly the person was injured, and whether an assailant used a weapon during the assault. Other objective measures of trauma severity might include the duration of the event and the amount of monetary or property loss.

Subjective severity of the trauma refers to the person's evaluation of the event and overlaps considerably with the second part of the trauma definition mentioned previously. Therefore, these estimates of severity tend to focus largely on feelings of fear, horror, and helplessness. Typically, survivors are asked to what degree they felt each of these emotions during the traumatic event. For diagnostic purposes, all that the assessment needs to establish is that the person experienced the emotion, and simple questions such as "Did you fear that you would be killed during the trauma?" are sufficient. In other cases, one may wish to establish the subjective severity relative to other events. In these cases, it is often helpful to have the person rate the intensity of the feeling on an arbitrary scale.

The objective and subjective ratings of trauma severity may be quite different. Seemingly minor (based on objective ratings) events are rated as subjectively very severe and vice versa. Therefore, it is important and informative to assess both the objective severity and subjective severity of traumatic events.

### **5.1.4. Assessing the Timing of the Traumatic Event**

Timing of the trauma refers to both the developmental level (i.e., age at which the event occurred) and the duration over which the event occurred. In this case, duration usually refers to whether the event occurred a single time (e.g., assault by a stranger on the street) or was repeated over time (e.g., physical abuse by a parent over years). There is some indication that exposure to a traumatic event at an early age or repeated exposure to a traumatic event over a period of time is more likely to lead to psychological problems than is a single traumatic event during adulthood. Some have argued that these problems extend beyond those captured by the PTSD diagnosis and include disrupted sense of self and relationships.

## **5.2. Assessing Early Emotional Reactions to the Traumatic Event**

As described earlier, it is common for people who have experienced a traumatic event to report reactions such

as intrusive memories, emotional distress, and increased arousal. In most situations, there will be little formal assessment of these reactions unless the person is being seen by a mental health professional. Informal assessment of these reactions might consist of a review of the common reactions listed in Table I. Such informal assessments respect the survivor's wish to limit what he or she discusses, and the presence of the common reactions discussed previously should be normalized and validated. Because these reactions will diminish over time naturally, it is recommended that a recently traumatized person be encouraged to use his or her naturally occurring support system and to follow his or her natural inclination with regard to how much to say and to whom to talk. However, an individual who experiences very severe emotional reactions or who shows high levels of dissociative symptoms should be referred for further evaluation and possible treatment.

### **5.3. Assessment of Symptoms**

Two areas are typically evaluated in a comprehensive assessment of trauma-related symptoms. The first is the specific symptoms included in the PTSD or ASD diagnostic criteria. The second involves the assessment of associated symptoms and problems. As with the assessment of trauma exposure, the specific questions or instruments included in the assessment battery will depend largely on the purpose of the assessment and the specific trauma history of the person being evaluated.

#### **5.3.1. Assessing the Symptoms of PTSD or ASD**

Both self-report and interview-based instruments have been developed to assess the symptoms of PTSD and ASD. In some cases, only the presence or absence of each symptom is assessed, but most instruments also call for rating the severity of the symptoms. The specific metric on which the severity of the symptoms is measured varies across instruments, but usually the measurement incorporates an evaluation of the frequency with which each symptom has occurred and the intensity or severity of each symptom when it is experienced. Typically, PTSD symptoms are assessed over a 1-month time frame, reflecting the criteria that symptoms must be experienced for at least 1 month to be diagnosed with the disorder, but symptoms have been assessed over periods as short as 1 week when necessary. Because ASD can occur only within the month immediately following a trauma, these

symptoms are generally evaluated relative to a time frame of 1 week or even a few days.

When assessing symptoms of PTSD or ASD, it is important to relate the symptoms to a specific traumatic event and to determine whether the current presentation reflects a shift from prior functioning. In the case of some of the symptoms, this is relatively straightforward because the symptoms are directly tied to the event (e.g., intrusive thoughts or memories of the event, efforts to avoid reminders of the event). In the case of other symptoms, however, the association with the specific traumatic event in question is not necessarily clear (e.g., sleep disturbance, sense of detachment from other people). In the latter case, the evaluator will generally find it necessary to use the "change from prior functioning" criteria to link the symptoms to a particular trauma. As noted earlier, this task is easiest when the traumatic event in question is relatively recent and discrete. When the trauma occurred a long time ago (particularly when it occurred during childhood) or over a long period of time (as in the case of domestic violence), it is much more difficult to establish the link to a particular event or a shift from prior functioning.

#### **5.3.2. Assessing Broader Difficulties Associated with Trauma**

Beyond the symptoms of PTSD and ASD, trauma survivors can experience a wide range of problems, including (a) mental health problems such as depression, anxiety, and substance abuse/dependence; (b) problems in functioning such as disrupted relationships, difficulty in working, and physical health problems; and (c) cognitive and emotional difficulties such as intense feelings of guilt or shame, anger control problems, and problem-solving deficits. Some instruments have been developed to assess these broader difficulties associated with trauma, but for the most part, evaluators have used established instruments for assessing each of these areas of interest.

## **6. PSYCHOLOGICAL TREATMENT OF POSTTRAUMATIC DISORDERS**

Psychological treatments for trauma survivors can be divided into those designed to be implemented in the immediate aftermath of the traumatic event (often termed "debriefing"), those aimed at treating ASD,

and those aimed at treating chronic PTSD. Treatments aimed at alleviating PTSD have been studied much more extensively than have those used for ASD or debriefing.

### 6.1. Treatment of Chronic PTSD

The most well-studied and empirically validated treatments for PTSD are based on behavioral and cognitive-behavioral approaches to therapy. Three major categories of treatments have been tested and found to be somewhat effective for patients with PTSD: (a) exposure therapy in which patients are encouraged to systematically confront reminders of the traumatic event that elicit anxiety, (b) cognitive techniques in which patients are taught to challenge cognitive distortions thought to underlie the symptoms of PTSD, and (c) anxiety management techniques in which patients are taught methods for managing their trauma-related anxiety. In addition, various combinations of these treatments have been examined and found to be useful in treating chronic PTSD.

Of the treatments used for chronic PTSD, exposure therapy has been the most widely validated. Two forms of exposure are typically included in these programs: imaginal and in vivo. Imaginal exposure refers to patients intentionally confronting specific memories of the traumatic event in their imaginations either by verbally describing the traumatic event or by writing a description of the trauma. Typically, this is done for a prolonged period of time (i.e., 30–60 minutes) and is repeated over the course of treatment. The second type of exposure, in vivo exposure, refers to real-life confrontation with situations that are objectively safe but elicit intense anxiety or with situations that elicit memories of the trauma itself. The most effective therapies for chronic PTSD include these two exposure techniques or combine one or both of these techniques with cognitive therapy techniques aimed at altering PTSD-related dysfunctional cognitions.

### 6.2. Treatments for ASD

Treatments for ASD have been studied much less extensively than have treatments for chronic PTSD. The available studies suggest that the cognitive-behavioral techniques found to be useful with cases of chronic PTSD are also helpful in the more immediate aftermath of trauma. These interventions are typically brief (four or five sessions, with each session lasting 90–120 minutes) and are conducted during the first

2 to 4 weeks after the trauma. Despite the early promise of these interventions, it is premature to conclude that treatment targeting ASD symptoms during the first month after a traumatic event are effective in reducing long-term difficulties associated with trauma.

### 6.3. Debriefing: Treatment in the Wake of Trauma

Although debriefing interventions (typically administered in a single session during the first several days after a trauma) have become widely used, they have been poorly studied. These interventions aim to reduce long-term mental health problems associated with the trauma, but there is little evidence to support their efficacy in achieving this goal. Indeed, there is growing evidence that debriefing does not substantially reduce the rates of PTSD among trauma survivors. Of more concern are recent findings suggesting that debriefing might actually be detrimental to some trauma survivors, leading to more long-term emotional disturbance than would have occurred with no intervention at all.

### See Also the Following Articles

Stress

### Further Reading

- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Press.
- Bryant, R. A., & Harvey, A. G. (2000). *Acute stress disorder: A handbook of theory, assessment, and treatment*. Washington, DC: American Psychological Association.
- Davidson, J. R. T., & Foa, E. B. (1993). *Posttraumatic stress disorder: DSM-IV and beyond*. Washington, DC: American Psychiatric Press.
- Foa, E. B., Keane, T. M., & Friedman, M. J. (2000). *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies*. New York: Guilford.
- Follette, V. M., Ruzek, J. I., & Abueg, F. R. (1998). *Cognitive-behavioral therapies for trauma*. New York: Guilford.
- Saigh, P. A., & Bremner, J. D. (1999). *Posttraumatic stress disorder: A comprehensive text*. Boston: Allyn & Bacon.
- Wilson, J. P., Friedman, M. J., & Lindy, J. D. (2001). *Treating psychological trauma and PTSD*. New York: Guilford.
- Wilson, J. P., & Keane, T. M. (1997). *Assessing psychological trauma and PTSD*. New York: Guilford.
- Yehuda, R. (1998). *Psychological trauma*. Washington, DC: American Psychiatric Press.



# Power, Authority, and Leadership

Lourdes Munduate and Francisco J. Medina

*University of Seville, Seville, Spain*

1. Power and Influence
  2. Formal Authority and Leadership
  3. Models of Leadership
- Further Reading

## GLOSSARY

**authority** The institutionalized power between a superior and a subordinate that ensures compliance with the superior's wishes because he or she is the boss.

**empowerment** A motivational process involving feelings of self-efficacy among organizational members; to enhance the perception of self-efficacy, members use a set of strategies to influence and improve their behavior, taking responsibility for their own lives rather than depending on a leader to direct and motivate them.

**influence** A force that one person (the leader) exerts on others (followers) to induce a change in the latter, including changes in behaviors, attitudes, and values.

**leadership** Power relationships and influence processes between a leader and followers in which the leader exercises greater influence over the followers to accomplish group or organizational objectives.

**power** The potential ability of a person to influence others; thus, influence is "power in action," just as power is "potential influence."

Leadership involves power relationships and influence processes between a leader and followers in which the leader exercises greater influence over the followers to

accomplish group or organizational objectives. Thus, power is essential to leadership and processes of influencing members of a group or an organization. However, although power is necessary for leadership, it is insufficient by itself for leadership. The same person could be described as having influence, being persuasive, and/or being a leader. As a result, social scientists have attempted to discover how leaders use power to influence followers, looking at the difference between exercising formal authority and leadership and followers' responses to these influence attempts. It should be noted that there is a clear difference between yielding to social pressure from an individual and being genuinely persuaded. This article first examines the various bases and sources of power that are used to influence people and the responses they elicit in followers as well as the difference between management and leadership. Then, it focuses on three fundamental features that are critical to leadership: traits or personal characteristics, behaviors, and situations. Finally, it describes some integrative approaches represented mainly by charismatic and transformational leadership.

## 1. POWER AND INFLUENCE

In trying to understand how people influence each other when they interact, social psychologists have used the concepts of social power and influence tactics. The two terms have sometimes been interchanged, but a clear distinction needs to be made between them.

Social power reflects the repertoire of tools available to a person to influence the environment or the other party. Influence tactics refer to the actual use of a certain tool in a specific situation. The definition of the two concepts helps to clarify the differences between them. Following Lewin's initial conceptualization in 1941, French and Raven in 1959 defined influence as a force that one person (the leader) exerts on others (followers) to induce a change in the latter, including changes in behaviors, attitudes, and values. Social power was subsequently defined as the potential ability of a person to influence others. Thus, influence is "power in action," just as power is "potential influence."

The conceptualization of power highlights two important characteristics of power. The first reflects that power is potential and, as a result, that a person can have power without using it. The availability of tools does not mean that the tools must be used. The second characteristic refers to the relational aspect of power in that it does not reside in the person but rather resides in the relationship between people in a given situation. In general, power is derived from the situation in which the person operates, and in the event of the situation changing, the person's power also changes.

### **1.1. Social Influence Dynamic**

In studies examining people's responses to leaders' influence attempts, there is a clear difference between yielding to direct or indirect social pressure from a leader and being genuinely persuaded. For example, the leader's influence may be strong enough to wield control over the followers' behavior, ensuring public agreement, regardless of whether the followers are privately convinced. On other occasions, the influence process may change followers' private attitudes or opinions and make them committed to the leader's request. Therefore, social psychologists have proposed various types of influence processes, the most important of which was Kelman's traditional distinction, made in 1958, among compliance, internalization, and identification. Compliance refers to a surface change in behavior and expressed attitudes, often as a consequence of coercion or the followers' desire to obtain a reward. Because compliance does not reflect internal change, it usually persists only while behavior is under surveillance. In contrast to compliance, internalization means subjective acceptance that produces true internal change that persists in the absence of surveillance. The norm becomes an internalized

standard for behavior. There is genuine support for the leader's proposals because they appear to be intrinsically desirable and are suited to the followers' values and beliefs. Identification is based on the leader's attractiveness to the followers, who imitate the leader's behavior and attitudes to gain his or her approval. The maintenance of close relationships with the leader involves the followers' need for acceptance and esteem.

The outcomes of these various forms of social influence in terms of compliance, internalization, and identification may occur at the same time. For example, a person who works in a nongovernmental organization (NGO) with exploited children in the Third World may accept a proposal from a superior to design a new program because he or she will obtain tangible rewards for carrying it out (compliance) and believes in the necessity and effectiveness of the program (internalization). At the same time, the person is proud of working with his or her superior on this project, and this will increase the level of satisfaction in the relationship with the superior (identification).

### **1.2. Bases and Sources of Power**

To understand the different ways in which power can be exercised, it is best to look at the various bases and sources of power. In 1959, French and Raven proposed a well-known taxonomy of power that considered five bases of power: reward, coercion, legitimacy, reference, and expertise (a sixth base of power, information, was added a little later). A leader possesses these resources and can use them to change the beliefs, attitudes, and/or behaviors of followers. Reward and coercive power rely on others believing that the leader can provide them with the desired rewards or can punish them, respectively. Using either of these bases will induce only a superficial change in the followers. In other words, none of the followers' privately held beliefs, attitudes, or values is changed. Only compliance is obtained, and the continuation of this compliance depends on successful surveillance of the followers by the leader. A person possesses legitimate power when others believe that he or she has a legitimate right to exert influence over them and they are obliged to accept this influence. It cuts across all three types of influence processes and, as a result, may share elements of each of them. Referent power refers to followers' identification with the leader. It leads to private acceptance by followers by enabling them to maintain a satisfactory relationship with the leader and to see themselves as similar to the leader in certain

important areas. Expert power depends on followers' perception of having expertise or knowledge in a specific domain. If the followers perceive the leader as an expert, this will result in private acceptance on the part of the followers. Finally, informational power leads to internalized and lasting changes in the followers' beliefs, attitudes, and/or values. Compared with other social power bases, the changed behavior resulting from information is maintained without continued social dependence on the leader and is instead based on the perceived importance and validity of the information. Only informational power leads to cognitive change in followers because it immediately becomes independent of the leader.

Another related dimension of power refers to sources of power. When dealing with the bases of power, the focus was the repertoire of tools available to a person to influence the other party or the environment. In terms of sources of power, the focus points toward the way in which a person comes to control these tools. Two different sources of power have been identified: position and personal power. Position power arises from the formal position held in a group or an organization, whereas personal power arises from personal attributes and the kind of relationship established with the other party. In position power, potential control is derived from legitimate authority. In personal power, it stems from task expertise or the opportunity to access certain information that is important for other members of the group or organization.

So far, this article has identified two sources of power, six bases of power, and three influence processes. Figure 1 illustrates the relationships among these aspects of power.

### 1.3. Influence Tactics

As was explained earlier, literature on social power takes power to be the capacity to exert influence. However, power can be used in different ways, and the specific ways in which power is enacted involve influence behavior. With the publication of *The Powerholders* in 1976, Kipnis approached influence behaviors from the perspective of the agent of influence. He stated that the choice of a means of influence depends mainly on the resources at the agent's disposal (i.e., on the power bases that the target possesses), on the target's inhibition to invoke a power base (as determined by the target's estimate of the costs of using a power base, the target's subjective values and attitudes, individual differences, and social norms),

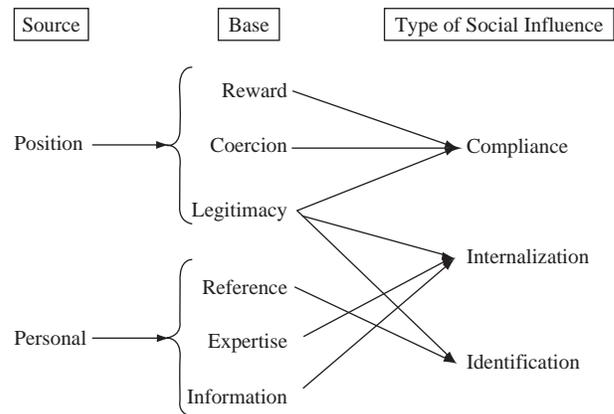


FIGURE 1 Relationships of sources, bases, and types of social influence elicited in followers.

and on the resistance that one expects from the target toward the influence attempt. Subsequent developments by Kipnis and colleagues in 1980 identified the specific behaviors that people have at their disposal for influencing others in organizations. The various categories of proactive influence behaviors that were proposed were referred to as "influence tactics."

Building on and extending the previous work, Yukl and colleagues developed a classification with nine proactive influence tactics: rational persuasion, inspirational appeals, consultation, ingratiation, personal appeals, exchange, coalition tactics, legitimating tactics, and pressure. In general, effective leaders use a combination of influence tactics that they select depending on their judgment of what is most appropriate for a given situation. As Yukl pointed out in his recent book titled *Leadership in Organizations*, a leader prefers to use tactics that are socially acceptable, feasible in terms of the leader's power in relation to the followers, likely to be effective for a particular objective, and unlikely to require much time, effort, or cost to the leader.

## 2. FORMAL AUTHORITY AND LEADERSHIP

As noted in the previous section, power that stems from a person's formal position in a hierarchical structure is sometimes called formal authority or legitimate power. Authority represents the institutionalized power between a superior and subordinates that ensures compliance with the superior's wishes because he or she is the boss. Thus, authority involves a

person's right to exercise control over other people or the environment. For example, a manager usually has the legitimate right to make a job-related request to subordinates, who in turn have the legitimate right to request job-related information and assistance from the manager. However, a manager may have a great deal of legitimate power due to his or her position but may have little, if any, expert power. As a result, followers may think that the manager does not understand the problem they must deal with but that they must do the work because they have been told to do so by the boss.

### 2.1. Management and Leadership

Although nobody has argued that managing and leading are equivalent, it has sometimes been assumed that managers' control over some important resources involves leadership processes. In fact, the exercise of power that stems from a person's hierarchical position may facilitate the emergence of leadership. But hierarchical position is simply a context from which leadership may or may not emerge. Leadership does not arise by virtue of occupying hierarchical roles. There are differences between formal role designations (which deal with formal aspects of the relationship) and the emergence of leadership (which deals with informal aspects of the relationship). Thus, the emergence of leadership involves more than subordination and is related to the kind, or depth, of social influence elicited from followers. People who achieve acceptable influence, such as internalization or identification, are the ones who are defined as leaders. Therefore, leaders can be identified because they demonstrate leadership. Countless presidents, kings, generals, and military figures who were chosen or appointed (bottom-up vs top-down), and who have power and authority, give few examples of behaviors that could be associated with actual leadership. If a person with formal authority (i.e., a manager) is able to induce internalization or identification, he or she would have managed to exercise leadership. If this person is not able to do so, he or she would have exercised authority. In turn, a person who develops a relation that induces internalization or identification despite not holding a hierarchical role (i.e., a nonmanager), is exercising leadership. In other words, a person can be a leader without managing, and a person can be a manager without leading. From this perspective, leading and managing are viewed as different processes, although leaders and managers are not considered to be different types of people.

TABLE I

#### Authority and Leadership as Results of Social Influence

<i>Social influence</i>	<i>Type of power</i>
Compliance	Formal authority (management)
Internalization	Leadership (transactional)
Identification	Leadership (transformational)

Nevertheless, it must be emphasized that some overlap does exist among the various types of social influence and that more than one influence process may occur at the same time, as stated previously. It may also be the case that the same type of influence attempts may produce different individual responses. As a result, not all theoretical perspectives share the same conceptualization on the exercise of power, authority, leadership, and management. In fact, the relationship among these concepts has been the subject of extensive debate. Table I presents a summary of the conceptualization developed in this article.

### 2.2. Empowerment

Leaders and managers often share power with followers or subordinates for a period of time to accomplish a variety of organizational tasks and objectives. This process is known as power delegation. In 1988, Conger and Kanungo proposed that empowerment differs from delegation in that the former is a motivational process involving feelings of self-efficacy among organizational members and is not temporally limited by the leader. To enhance the perception of self-efficacy, members use a set of strategies to influence and improve their behavior, taking responsibility for their own lives rather than depending on a leader to direct and motivate them. In conventional terms of social influence, empowerment can be understood as a process of identification focused toward the development of a shared vision among organizational members. The aim of the empowerment process, by strengthening individual and/or collective efficacy, is to enable organizational members to solve organizational problems arising from the turbulent environment and diversification of the workforce.

## 3. MODELS OF LEADERSHIP

As stated previously, a person cannot lead unless there are power relationships between the leader and the

followers. However, power by itself is not a sufficient condition for leadership. Various perspectives and models have approached the question of the variables that need to be considered to achieve the exercise of leadership. The conventional perspective emphasized the personal attributes and acquired behaviors that leaders bring to the situation as well as the situational forces that shape and determine effective leadership. Some new perspectives emphasize that leadership is not only a transactional relationship between the leader and the followers but also produces important effects in followers' value systems.

### **3.1. Traits, Behaviors, and Situational Models**

Some of the earliest research to approach the subject of leadership consisted of trait or dispositional studies. Trait models focus on what a person brings to leadership activities. This approach emphasizes the personal attributes of a leader that are treated as innate and unique. Accordingly, attention has focused on certain combinations of personality traits (e.g., tolerance of stress, emotional maturity, self-efficacy), motives (e.g., power orientation, relatively strong motivation to achieve, need for affiliation), and skills (e.g., technical, conceptual, interpersonal) because these might be reflected in the achievement of leader status and as possible predictors of a leader's effectiveness in promoting group or organizational performance. However, the studies aimed at uncovering the characteristics constantly linked to leadership have been inconclusive, and no definitive predictors of effective leadership have been established.

Researchers became discouraged with the trait approach, and around 1950 they began to look at what leaders or managers do in their jobs. The behavioral models of leadership analyze what leaders or managers do with whatever personal traits, skills, or motivational capacities. The primary concern of behavioral models focused on developing methods to determine what leaders do and measure relevant dimensions of their behavior that might be linked to group effectiveness and subordinate satisfaction. Two main dimensions that reflected how subordinates perceived a manager's behaviors were identified. Consideration (or relationship-oriented) behaviors reflect the extent to which a leader shows concern for subordinates' satisfaction and well-being. Initiating structure (or task-oriented) behaviors reflect the degree to which a leader explains and

defines the roles of a task, assigns subordinates to various task roles, controls subordinates' performance, and provides feedback to subordinates. Because these are two independent dimensions, the behavior of a leader may be described as a combination of both. According to this perspective, the ideal leader would be someone with high scores on both dimensions.

Both the trait and behavioral models focus on personal attributes and acquired behaviors that leaders bring to the situation and provide a framework for identifying and cultivating leaders in organizations. However, an important difference between the models lies in the fact that traits are considered innate in a person—a person either does or does not have the gift—whereas specific behaviors may be learned. A difference in the practical implications of the two models is that if the trait proposals were correct, leaders could be selected to hold formal posts in groups and organizations. However, if the behavioral models were correct, it would be possible to teach leaders, such that programs could be designed to train people to become effective leaders. The interesting aspect of this last perspective is that it could increase the number of available leaders.

During the 1960s, situational models supplanted universalistic approaches. Situational models focus on situational variables that moderate the relationship between the leader's attributes and behaviors and his or her effectiveness. Researchers became convinced that predicting the success of leadership was more than just a case of signaling certain preferable traits or behaviors. The impossibility of obtaining congruent empirical results gave rise to a new emphasis on situational factors. The relation between leadership style and efficacy indicated that Style X would be appropriate for Condition A, whereas Style Y would be appropriate for Condition B. However, it was one thing to assert that the efficacy of leadership depended on the situation, but managing to identify those situational conditions was something completely different. Attempts to isolate those conditions gave rise to a wide variety of situational models of leadership, all of which focused on explaining how aspects of the situation moderate the relationship between leaders' traits, behaviors, or both and the outcomes arising from a particular situation. A representative sample of these models is next described briefly.

The least preferred coworker (LPC) contingency model introduced by Fiedler in 1967 considered two types of leadership styles that were regarded as fixed or immutable dispositional traits: directive or task-oriented

style and nondirective or relationship-oriented style. The effectiveness of both styles depended on three situational variables that were identified as the leader-member relationship (e.g., the degree of confidence and commitment between the leader and the followers), task structure (e.g., the extent to which procedures for performing the task are established), and position power (e.g., control of resources such as money or information). Thus, when selecting leaders, this model suggests matching a person's leadership style with the situations. For example, if there is a group or situation in which the leader is trusted, the task is highly structured, and position power is good, it would be more appropriate to assign a task-oriented leader than to assign a relationship-oriented leader.

The popular life cycle model proposed by Hersey and Blanchard in 1977 identified four leadership styles—telling, selling, participating, and delegating—each of which is basically a mixture of task behavior and relationship behavior styles. The model viewed the “maturity” of followers—defined in terms of ability, motivation, and self-confidence—as the chief contingency variable to which one of these leadership styles had to be matched. Thus, followers with low maturity, as is often the case with new employees, would be most responsive to a telling leadership style (characterized as relatively high on task behaviors and low on relationship behaviors), whereas a delegating style (low in both task and relationship behaviors) would be more appropriate for followers with high maturity, as is usually the case with experienced employees. Therefore, the model states that the leader should adapt his or her leadership style to the maturity of the followers. The life cycle model was widely marketed and used in many leader training and development programs in organizations during the 1980s despite having received only partial support from empirical research.

According to the path-goal model developed by House in 1971, the leader's main role is to help his or her followers achieve goals and to provide sufficient leadership and support to guarantee that their goals are compatible with the overall objectives of the group or organization. To achieve these objectives, the leader must identify two critical situational variables (follower needs and task demands) and then choose an appropriate leadership style (consideration style or initiating structure style). The behavior of the leader is motivating for his or her followers in that, first, the satisfaction of the requirements of these followers depends on a good performance and that, second,

these followers need to receive the leadership, support, and rewards that are vital for achieving a good performance.

The normative decision model built by Vroom and Yetton in 1973, and developed by Vroom and Jago in 1988, considers that decision making is the critical leadership activity with groups. The model went further than did previous proposals on decision procedures by specifying which procedure will be the most effective in each of several specific situations. Two intervening variables for measuring decision effectiveness were suggested: decision quality and decision implementation or acceptance by the parties involved in the decision. In turn, these variables are affected by the decision-making procedure used by the leader: two varieties of autocratic decision (AI and AII), two varieties of consultation (CI and CII), and one variety of joint decision making as a group (GII). Depending on seven key situational variables, such as the amount of relevant information possessed by the leader and followers and the likelihood that followers will accept an autocratic decision, the model provides a set of rules for identifying the best decision procedure in terms of decision quality, parties' acceptance, and costs of decision making. Vroom and Yetton developed a decision process flow chart, known as a decision tree, to help leaders identify the optimal decision procedure for each situation.

A more recent situational model is the cognitive resource model, introduced by Fiedler and Garcia in 1987, which focuses on the cognitive resources of the leader (intelligence and experience), leader behavior (directive leadership), and two aspects of the leadership situation (interpersonal stress and the nature of the group's task). Basically, the theory points out that leader intelligence contributes to group performance only when the leader is directive and the followers require guidance to perform the task effectively. However, perceived stress moderates the relation between intelligence and decision quality. Finally, perceived stress moderates the relationship between leader experience and performance. The general idea is that leader effectiveness is predicted by intelligence in low-stress conditions and by experience in high-stress conditions.

The final situational model of leadership considered here is the tridimensional leadership theory developed in 2003 by Yukl, who considered that leadership behavior can be described in terms of three broad categories: task oriented (primary effect on efficiency and reliability), relations oriented (primary effect on

innovation and human resources), and change oriented (primary effect on innovation and adaptation to the environment). The theory states that effective leaders integrate behaviors in a way that is consistent with the situation and that the primary situational variables are the type of organization or unit and the amount of environmental volatility and uncertainty.

In general, situational models of leadership have received only moderate support from empirical studies. However, most of them have the advantage of providing detailed prescriptions for practitioners that are theoretically rooted in the emphasis of leaders' needs to interpret situations and to bring specific traits and acquired behaviors to these situations so as to exercise effective leadership.

### 3.2. Charismatic and Transformational Leadership

During the 1980s, the previous leadership models began to come under criticism due in large part to their dependence on the methodology used that gave rise to what is known as the "performance-cue" effect. This effect refers to the fact that groups with knowledge of their performance levels reported "objective" leadership behaviors that were consistent with performance levels (e.g., good leadership behavior was reported when performance was good, whereas poor performance led to less flattering ratings), regardless of the actual behavior of the leader. During these years, conventional leadership wisdom came under heavy attack, and this led to a "leadership gap" or "leadership crisis." Nevertheless, this gap was soon filled with a new kind of leadership whose characteristics were connected to one of the oldest traditions in leadership studies, that is, charisma. This Greek word meaning "special or divine gift" was already prominent in Weber's analysis of organizational authority in 1921. The charismatic leadership theory developed by Conger and Kanungo in 1988 relies on its effect on followers and society in general and stemmed from their radical vision of a new world, the exhibition of their own behavior for followers to imitate, and the confidence expressed in the followers' capacity to achieve challenging goals.

Therefore, this new perspective of charismatic leadership differs from the conventional line on leadership studies. To this point, this article has analyzed various aspects of leadership that have been grouped under the term "transactional leadership" because a type of

transaction is established between a leader and followers in all of them. The members of the group or organization recognize the leader as such and accept his or her authority, but in turn the leader must provide the followers with valuable resources. As we have seen, these resources may occasionally refer to the cognitive skills and knowledge that enable objectives to be obtained, whereas other times they may be rewards such as pay increases and recognition of jobs well done. However, if one examines social reality and history itself, one finds that there is another type of leadership that produces important effects in followers in particular but also in society in general. This type of leadership has been labeled "transformational leadership" and refers basically to charismatic leadership.

The differences between transactional leadership and transformational leadership, as defended by Bass in 1985, are significant in terms of followers' commitment to challenging group or organizational goals. In transformational leadership, these goals are presented as a compelling vision and are expressed in ideological and inspirational language (e.g., "what we are able to achieve by challenging ourselves continuously"). In fact, transformational leadership is most likely to emerge when there is a group or organizational crisis or when followers are disenchanted with the current state of affairs. In other words, transformational leadership affects, first of all, followers' value systems and ambition levels. Its ultimate effect is that followers will appreciate projects and options that they would not appreciate unreservedly on their own. It transforms the value systems of followers in such a way that they start pursuing new ideals. Thus, the role of the transformational leader is to attempt to bring about followers' acceptance of a bold vision by inciting positive feelings toward desired behaviors. Change, then, comes about through an induced reevaluation of behavioral options by followers.

In contrast to transformational leadership, transactional leadership does not intervene in followers' value systems. Leaving these systems untouched, it tries to perceptually or materially model the contents of influence attempts to fit the existing values of followers. Followers become committed to supporting and implementing the proposals espoused by the leader because these proposals are intrinsically desirable and match their values and beliefs. The leader achieves this by clarifying or guaranteeing the profitability of requested efforts.

By the 1990s, issues of transformational leadership and charisma had become the dominant themes in leadership studies, and there was a proliferation of

models and perspectives on charismatic and transformational leadership. They emphasized, for instance, the connection between charisma and culture as well as between charisma and affective communication.

### **3.3. Individual Differences and Cognitive Processing Models**

As theories on transformational leadership developed, new models in the domain of transactional leadership appeared, featuring a renewed interest in long since discredited trait-based models. Although trait approach research has been out of favor for a long time, recent work has once again suggested that stable individual differences in leader behavior do exist. Studies have become increasingly interested in the special endowments of leaders, and new models have focused on the intellectual and creative abilities of leaders. One good example of this development is the role motivation theory of organizational leadership that was introduced by Miner at the beginning of the 1970s and has been developing constantly ever since. The theory states that different leader attributes fit different types or roles in the organization and that leadership is a function of the larger organizational form within which it operates. Using a projective test, Miner described the type of motivational traits required for success in most leadership (management) positions in hierarchical organizations. Leader role requirements and the matching motive patterns are derived from the form of the organization. This person-organization fit model involves individual differences, strategy, organizational structure, and environment.

The individual differences research domain was strongly influenced by the information-processing concepts applied to the study of leadership perceptions. Cognitive research generally emphasizes the role of followers as well as leaders in leadership perception. Extensive research using an information-processing approach shows that individuals can be recognized as leaders based on their fit with leadership prototypes or that leadership can be inferred from favorable outcomes. Lord and colleagues' categorization theory and Mitchell and colleagues' attributional model represent that trend. The general idea is that leadership resides (at least in part) in the minds of the followers and that individual differences are important perceiver frameworks for assessing emergent leadership. This approach emphasized the role of individual differences in defining implicit theories or leader prototypes. The categorization theory assumes that relatively fixed

prototypes, which are learned through extensive experience by followers, are the basis for leadership categorization. Followers appear to share (within a context) a set of general beliefs about the characteristics related to leadership (e.g., decisive, determined, intelligent) in diverse situations and tend to use their implicit theories and leader prototypes to decide whether or not an individual is judged to be an emergent leader. Specifically, if an individual is perceived to match followers' leader prototype, the person is likely to be viewed as an emergent leader and rated accordingly. From Lord and Maher's theoretical perspective, although leader qualities are clearly important for understanding leadership, followers' social-cognitive process is also important. Attributional notions are also introduced into the leadership cognitive approach, and the mutual influence of a superior's behavior and subordinates' behavior on one another is considered, inserting each person's causal attributions into the loop as mediational variables. As for future developments using the cognitive perspective, Lord and Emrich suggested in 2000 that one promising area covers the effect of self-concepts and self-regulatory processes such as self-efficacy. Although much of the research in the cognitive domain has focused on the relationship between a leader's behavior or traits and subordinates' satisfaction, behavior, and performance, increasing attention is being devoted to understanding the intervening mechanisms, such as values and self-concept, with which leaders influence followers.

As for conclusions on leadership research, Sternberg and Vroom provided a clear summary about theoretical leadership approaches in 2002: "The persons, the situation, and the interaction of the person with the situation . . . all matter for leadership effectiveness." Other recent reviews of leadership literature have called for increased attention on leadership context (e.g., Lowe and Gardner in 2000) as well as a greater focus on personal attributes, especially in the area of executive leadership (e.g., Hooijberg and Schneider in 2001). Similarly, Lord and Emrich suggested, in their review of cognitive leadership research in 2000, that "information processing research has established . . . that leadership factors reside not only in the minds of followers, but also in the minds of leaders, in leaders' behavior and attitudes, and in the social context in which leaders and their followers interact."

### **See Also the Following Articles**

Authoritarianism ■ Leadership and Culture ■ Traits

### Further Reading

- Andriessen, E. J. H., & Drenth, P. J. D. (1998). Leadership: Theories and models. In P. J. D. Drenth, H. Thierry, & C. J. de Wolff (Eds.), *Handbook of work and organizational psychology* (2nd ed., Vol. 4, pp. 321–355). Hove, UK: Psychology Press.
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.
- Bass, B. M. (1990). *Bass & Stogdill's handbook of leadership* (3rd ed.). New York: Harper & Row.
- Conger, J. A., & Kanungo, R. N. (1988). *Charismatic leadership: The elusive factor in organizational effectiveness*. San Francisco: Jossey-Bass.
- Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.
- Fiedler, F. E., & Garcia, J. E. (1987). *New approaches to leadership: Cognitive resources and organizational performance*. New York: John Wiley.
- French, J., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies of social power* (pp. 150–167). Ann Arbor, MI: Institute for Social Research.
- Hardy, C., & Clegg, S. R. (1996). Some dare call it power. In S. R. Clegg, C. Hardy, & W. R. Nord (Eds.), *Handbook of organizational studies* (pp. 622–641). London: Sage.
- Hersey, P., & Blanchard, K. H. (1977). *Management of organizational behavior: Utilizing human resources*. Englewood Cliffs, NJ: Prentice Hall.
- Hollander, E. P. (1985). Leadership and power. In G. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., pp. 485–537). New York: Random House.
- Hooijberg, R., & Schneider, M. (2001). Behavioral complexity and social intelligence: How executive leaders use stakeholders to form a systems perspective. In S. Zaccaro, & R. Klimoski (Eds.), *The nature of organizational leadership* (pp. 104–131). San Francisco: Jossey-Bass.
- House, R. J. (1971). A path-goal theory of leader effectiveness. *Administrative Science Quarterly*, 16, 321–339.
- House, R. J. (1988). Leadership research: Some forgotten, ignored, or overlooked findings. In J. G. Hunt, B. R. Baglia, H. P. Dachler, & C. A. Schriesheim (Eds.), *Emerging leadership vistas* (pp. 179–199). Lexington, MA: D. C. Heath.
- Kelman, H. C. (1958). Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution*, 2, 51–60.
- Kipnis, D. (1976). *The powerholders*. Chicago: University of Chicago Press.
- Kipnis, D., Schmidt, S. M., & Wilkinson, I. (1980). Intra-organizational influence tactics: Explorations in getting one's way. *Journal of Applied Psychology*, 65, 440–452.
- Lewin, K. (1941). Analysis of the concepts whole, differentiation, and unity. *University of Iowa Studies in Child Welfare*, 18, 226–261.
- Lord, R. G., & Emrich, C. G. (2000). Thinking outside the box by looking inside the box: Extending the cognitive revolution in leadership research. *The Leadership Quarterly*, 11, 551–580.
- Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, 34, 343–378.
- Lord, R. G., & Maher, K. J. (1991). *Leadership and information processing: Linking perceptions and performance*. Boston: Unwin Hyman.
- Lowe, K. B., & Gardner, W. L. (2000). Ten years of The Leadership Quarterly: Contributions and challenges for the future. *The Leadership Quarterly*, 11, 459–514.
- Meindl, J. R. (1995). The romance of leadership as a follower-centric theory: A social constructionist approach. *The Leadership Quarterly*, 6, 329–342.
- Miner, J. B. (1993). *Role motivation theories*. New York: Routledge.
- Miner, J. B. (2002). The role motivation theories of organizational leadership. In F. J. Yammarino, & B. J. Avolio (Eds.), *Transformational and charismatic leadership: The road ahead* (pp. 309–338). New York: Elsevier.
- Mitchell, T. R., Green, S. G., & Wood, R. E. (1981). An attributional model of leadership and the poor performing subordinate: Development and validation. In L. L. Cummings, & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 3, pp. 197–234). Greenwich, CT: JAI.
- Sternberg, R. J., & Vroom, V. (2002). Theoretical letters: The person versus the situation in leadership. *The Leadership Quarterly*, 13, 301–323.
- Vroom, V. H., & Jago, A. G. (1988). *The new leadership: Managing participation in organizations*. Englewood Cliffs, NJ: Prentice Hall.
- Vroom, V. H., & Yetton, P. W. (1973). *Leadership and decision making*. Pittsburgh, PA: University of Pittsburgh Press.
- Weber, M. (1921). The sociology of charismatic authority. In H. H. Gerth, & C. W. Mills (Trans. & Eds.), *From Max Weber: Essays in sociology* (pp. 245–252). New York: Oxford University Press.
- Yukl, G. A. (2002). *Leadership in organizations* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Yukl, G., (2003). Tridimensional leadership theory: A roadmap for flexible, adaptive leaders. In R. J. Burke, & C. Cooper (Eds.), *Leading in turbulent times*. London: Blackwell.
- Yukl, G., & Falbe, C. M. (1990). Influence tactics in upward, downward, and lateral influence attempts. *Journal of Applied Psychology*, 75, 132–140.
- Yukl, G., & Tracey, B. (1992). Consequences of influence tactics used with subordinates, peers, and the boss. *Journal of Applied Psychology*, 77, 525–535.
- Zaccaro, S., & Klimoski, R. (Eds.). (2001). *The nature of organizational leadership*. San Francisco: Jossey-Bass.



# Prejudice and Discrimination

Victoria M. Esses, Antoinette H. Semanya, and Monika Stelzl

University of Western Ontario, London, Ontario, Canada

1. Stereotypes, Prejudice, and Discrimination
  2. Causes of Prejudice and Discrimination
  3. The Experience of Being a Target of Prejudice and Discrimination
  4. Psychological Strategies for Reducing Prejudice and Discrimination
  5. Programs for Reducing Prejudice and Discrimination in Educational and Employment Settings
- Further Reading

Prejudice toward and discrimination against members of other groups can take a variety of forms and are caused by a number of factors. Their consequences for target group members are extensive and can potentially pervade all aspects of life. Several strategies for reducing prejudice and discrimination have been suggested, based on an understanding of the processes that serve to maintain negative intergroup relations. These strategies have led to the development of programs for use in educational and employment settings.

## GLOSSARY

**discrimination** Negative behavior toward members of a social group based on group membership.

**group identification** The degree to which a person associates himself or herself with a group in terms of identity, shared interests, and outcomes.

**group identity** A person's view of, or sense of, himself or herself as a member of a social group.

**illusory correlation** Falsely perceiving a relationship between two variables where none exists.

**in-group** A group with which a person identifies or feels a sense of belonging.

**minimal group paradigm** Arbitrary assignment of individuals to different groups, with minimal information other than group membership provided.

**out-group** A group other than those to which a person belongs.

**prejudice** A negative prejudgment about or attitude toward members of a social group.

**stereotypes** Beliefs about the specific characteristics possessed by members of a social group.

## 1. STEREOTYPES, PREJUDICE, AND DISCRIMINATION

People are judged and treated not only based on who they are as individuals, but based on the various groups to which they belong. For example, almost universally, members of the national group within a country are viewed more positively and receive better treatment than immigrant groups. Traditionally, men and women have been seen as possessing different characteristics, and women have frequently been viewed as inferior to men, with consequences for their treatment in society. Thus, relations among individual members of social groups can be strongly influenced by stereotypes, prejudice, and discrimination.

Although stereotypes, prejudice, and discrimination are conceptually related, they are not necessarily highly correlated. People may hold similar attitudes toward a variety of out-groups, whereas their stereotypes of these groups

may differ considerably. For example, although people may hold similarly negative overall evaluations of two groups, they may perceive one group as lazy and unintelligent and the other as conniving and cutthroat. In such cases, the stereotypes differ but the overall attitude toward the two groups is the same. Of importance, prejudice is not always related to discrimination. One major reason for this lack of association is that people do not necessarily act on their attitudes, particularly when such behavior may be prohibited by society. For example, even though an individual may hold a negative attitude toward members of a particular ethnic group, that individual might not be able to overtly express this attitude or blatantly discriminate against a member of the group due to societal constraints.

Some people claim that prejudice and discrimination are things of the past and that they no longer influence individuals in society. Although it is the case that overt, blatant prejudice has decreased considerably in recent years, it seems to have been replaced by a more subtle, covert prejudice. This subtle prejudice may be the result of unconscious negative emotions toward a group (of which prejudiced individuals are not even aware). It may be expressed in situations in which appropriate behavior is not clear-cut and where a variety of explanations for the negative behavior are available. For example, an individual might not hire an immigrant job candidate based on claims about the candidate's lack of local experience, rather than expressing negative attitudes toward the candidate's group.

Prejudice and discrimination may operate at different levels of society. At an individual level, a single person may hold prejudicial attitudes and discriminate against members of other groups. At an institutional level, prejudice and discrimination may operate through the practices and policies built into the institutional framework, which may prevent the full participation of members of certain groups. At a societal or cultural level, prejudice and discrimination may operate through the beliefs and values of the society that encourage and justify discrimination against particular groups. In all cases, the consequence is that certain groups receive negative outcomes and are not able to enjoy the benefits of full participation in society.

## **2. CAUSES OF PREJUDICE AND DISCRIMINATION**

Prejudice and discrimination may be caused by a variety of factors. Chief among these are several factors

that can be described as involving competition and threat. Realistic Group Conflict Theory and the more recent Instrument Model of Group Conflict propose that intergroup attitudes and behavior reflect group interests and are based, at least in part, on the nature of and the compatibility of group goals. When group goals are compatible, positive relations are likely to result. Conversely, when group goals are incompatible and groups are seen as competing for resources, conflict and negative intergroup relations are likely to result. Intergroup conflict, in turn, promotes hostility toward the competitive out-group. This hostility may be manifested in the expression of negative attitudes toward members of the other group and in attempts to prevent the other group from succeeding. For example, it has been demonstrated that when immigrants are seen as doing well and thriving in a difficult job market, members of the host population perceive increased competition with immigrants for resources and, therefore, are more likely to express negative attitudes toward immigrants and immigration. Relatedly, Social Dominance Theory suggests that threats to the in-group's power and dominance in society can lead to discrimination and willingness to aggress against less dominant out-groups. In such cases, prejudice may be driven by a fear of loss of social status or power rather than by a fear of loss of resources.

According to Social Identity Theory, for prejudice and discrimination to arise, it is not necessary to believe that a group is threatening tangible resources. Rather, the mere categorization of people into out-groups and an in-group stimulates a motivation to maintain or achieve a sense of positive group distinctiveness. As demonstrated in numerous studies using the minimal group paradigm, strategies for doing so include biases in evaluations that favor the in-group over out-groups and discrimination in resource allocations. In such cases, the motivation for a positive group identity may lead to processes that result in prejudice toward and discrimination against members of other groups.

Similarly, a number of recent theories have suggested that prejudice and discrimination may result from threats to values. For example, Symbolic Racism Theory proposes that prejudice against African Americans in the United States is partly attributable to a belief that African Americans threaten core American values, particularly those embodied in the Protestant work ethic (e.g., value of hard work, reward based on personal merit). Similarly, research on prejudice toward homosexuals has demonstrated that

perceived threats to religious and family values may form the basis of, and are used to justify, prejudicial attitudes.

Certain types of individuals are especially likely to perceive these types of threats and, thus, are particularly likely to hold negative attitudes toward other groups. Two personality variables that have been studied extensively in this context are right-wing authoritarianism and social dominance orientation. Right-wing authoritarianism involves submission to authority, belief in traditional values, and willingness to aggress against those who do not follow traditional ways. High right-wing authoritarians tend to be threatened by a variety of out-groups, particularly those that are seen as holding nontraditional values, and to hold negative attitudes toward these groups. In contrast, social dominance orientation involves a belief in inequality and preference for hierarchically structured social systems. Individuals who are high in social dominance orientation believe that unequal social outcomes and social hierarchies are appropriate, and they are especially likely to perceive threats to their group's status, power, and resources from other groups. As a result, they are especially likely to hold negative attitudes toward, and to discriminate against, other groups.

Group conflict and perceived threat, whether tangible or intangible, are not the only factors that can lead to prejudice and discrimination. There is a large body of research to suggest that people naturally categorize individuals into groups, particularly based on salient features such as sex, race, and age. Once people do so, a number of cognitive processes serve to promote stereotypes and prejudice toward members of other groups. For example, it has been demonstrated that the operation of an illusory correlation between a distinctive out-group (e.g., Asians) and a distinctive behavior (e.g., erratic driving) can result in an unfounded negative stereotype about members of the group (e.g., belief that Asians are bad drivers). Once formed, stereotypes are difficult to change due to selective attention to, interpretation of, and memory for stereotype-consistent information. In addition, once people categorize others into groups, they tend to exaggerate the similarities within groups and the differences between groups. Thus, members of an out-group are seen as all alike and as very different from members of the in-group.

In sum, the perception that a group is threatening the in-group's way of life and competing with the in-group for resources, power, status, and positive

identity can lead to prejudice toward and discrimination against members of that group. Individuals who are high in right-wing authoritarianism and social dominance orientation are particularly likely to perceive such threats. Cognitive factors related to the categorization of individuals into groups can also promote and perpetuate negative stereotypes and prejudice.

### 3. THE EXPERIENCE OF BEING A TARGET OF PREJUDICE AND DISCRIMINATION

Individuals belonging to groups that are negatively stereotyped in society are potential targets of prejudice and discrimination. How do members of such groups experience and navigate the social world with these added concerns? Central to this experience is identity. Although considerable variation in identification exists, individuals belonging to devalued groups (e.g., ethnic minorities, physically handicapped individuals, homosexuals) tend to identify strongly with these groups. This heightened identity is the result of a combination of factors such as awareness of the devalued identity and availability of social support from group members. For many members of devalued groups, then a significant portion of self-image and identity is tied to that group membership.

Even before members of devalued groups become victims of prejudice and discrimination, the knowledge of their devalued group identity can have adverse consequences. One example of these consequences is evident in research on "stereotype threat." Stereotype threat occurs in situations in which members of negatively stereotyped groups perceive the possibility of confirming a negative group stereotype through their behavior. For example, the stereotype of African Americans as less intelligent than European Americans creates anxiety for them in academic domains. In these domains, performing poorly would serve to confirm the negative stereotype of low African American intelligence. When faced with the threat of confirming a negative stereotype, performance is impaired and, over time, disidentification with the academic domain can occur. As a classic demonstration of the stereotype threat effect, when a test is described as a measure of academic ability and race is made salient, the performance of African American students is impaired. That is, African American students perform worse

than European American students taking an identical test and also worse than African American students who are not told that the test is a measure of academic ability or for whom race is not made salient. Ironically, this is particularly the case for African American students who identify strongly with academic achievement. Extensive research has demonstrated the potential for such impaired performance related to the stereotypes of a variety of groups (e.g., women and math performance).

Because of the awareness of the potential for prejudice and discrimination, individuals belonging to devalued groups may exist in a state of attributional ambiguity. This term is used to describe a state in which members of devalued groups cannot make clear explanations for the causes of negative or positive events that occur in their lives. For example, if an African American individual receives poor service in a restaurant, it might not be clear whether this occurred due to the prejudice of a server or to the fact that the restaurant was busy at the time. This ambiguity can exist even when positive events occur. For example, if an African American student receives a high grade in a course, it might not be clear to the individual whether this grade was due to his or her ability or to the fact that the professor was concerned with appearing nonprejudiced.

This ambiguity surrounding the causes of positive and negative events can have positive or negative implications for the self. Making attributions to discrimination can protect self-esteem by allowing individuals to maintain positive self-beliefs in the face of failure in desired domains. However, making constant attributions to stable, external, and/or uncontrollable factors has been shown to be related to poor self-esteem, feelings of helplessness, and depression. Research shows that devalued group members at times will make attributions to bias, whereas in other situations they are reluctant to attribute events to prejudice. There is a tension that devalued group members must negotiate between making accurate assessments of prejudice and protecting perceptions of efficacy and control. Although the state of attributional ambiguity is described as pervasive, not all members of devalued groups will perceive and react to attributional ambiguity in the same way. Assessing the antecedents and consequences of attributional ambiguity is important in understanding the experience of being a target of prejudice and discrimination, as well as in promoting the overall well-being of these individuals.

There are other costs associated with making attributions to prejudice and discrimination, including

social isolation, physical threat, financial loss, and emotional disturbance. Not surprisingly, individuals who belong to devalued groups are not always willing to perceive or admit prejudice and discrimination. One phenomenon that captures this is called the personal/group discrimination discrepancy. Researchers have found that individuals are willing to admit discrimination against their group in society, but are much less willing to admit that they personally have experienced discrimination. One factor that consistently affects the size of the personal/group discrimination discrepancy is individuals' identification with the group in question. Researchers have found that those higher in group identification are less likely to show the discrepancy; that is, they report more similar levels of personal and group discrimination than do those who are low in group identification. Perceiving or not perceiving discrimination has implications for numerous group processes such as identification and the willingness to engage in behavior that will help to improve the position of one's group in society.

In sum, the consequences of being a target of prejudice and discrimination are pervasive and begin even before observable negative behavior from others has occurred. However, individuals who belong to devalued groups are not passive recipients of biased treatment. They actively construct their world and experiences to protect and promote well-being and overall life success.

#### **4. PSYCHOLOGICAL STRATEGIES FOR REDUCING PREJUDICE AND DISCRIMINATION**

By being aware of the processes that lead to prejudice and discrimination, we can develop strategies for promoting more positive attitudes and behavior toward members of other groups. The contact hypothesis represents one of the most prominent approaches to prejudice reduction. In general, people are afraid of the unknown, which includes groups of people with whom they are fairly unfamiliar. However, the more knowledge about, understanding for, and experience with other groups that people obtain, the less negative and hostile people are likely to be. The underlying reasoning of the contact hypothesis is that if people are exposed to members of other groups in a variety of ways, they will become more positive toward members

of those groups and less likely to display prejudicial behavior. Nevertheless, it is not sufficient just to expose people to a minimal amount of information about another group or to live next door to someone who is of a different background. Several specific conditions must be met for the contact strategy to improve attitudes and behavior toward members of other groups.

Actual cooperation among members of the different groups is essential. In other words, people from the distinct groups should not be competing against each other for jobs, status, living space, money, and so forth. Cooperation between groups can be achieved by developing a common goal that can be achieved only if members of the different groups work together. Such cooperation will also help to develop a common identity that further decreases negative intergroup attitudes and discrimination. Equal status between the groups represents another important factor that underlies the contact strategy. In other words, it is critical that members of one group do not see themselves as superior to the members of the other group. Cooperative interdependence has been found to be helpful in superseding the presence of unequal status between two groups. In cooperative interdependence, each individual, regardless of group membership, is equally important for the task accomplishment.

Frequency, duration, and meaningfulness of interactions among the members of the different groups represent a set of factors that contribute to the effectiveness of the contact strategy. Specifically, the interaction among the members of the distinct groups should be frequent, long, and significant rather than occasional, short, and irrelevant. As a result, people will gain meaningful personal information about members of the other group. Social and institutional support is the final essential condition of the contact hypothesis. Contact between groups in the manner outlined here should be strongly supported by various levels of society such as the government, educational organizations, and employers. Overall, the contact approach provides a general framework consisting of several important factors that have been shown to contribute to decreasing prejudice and discrimination among members of different groups.

The second major approach to reducing prejudice and discrimination consists of categorization strategies. Two main strategies for restructuring people's mental representations of members of other groups have been suggested. Recategorization proposes the development of one common identity (i.e., us) rather

than distinct in-group–out-group identities (i.e., us vs them). The existence of a shared identity will decrease the salience of the differences that often exist between two groups and will highlight the commonalities that the members of the two groups share. As a result, this focus on the shared aspects of a common identity will decrease the negative attitudes and discrimination between the two groups.

Decategorization offers a different approach to restructuring the perception of “us versus them.” Rather than recategorizing separate identities by developing one shared identity, decategorizing signifies the elimination of group categorization. In other words, the removal of group-based typing will lead to the individualization of the members of the other group. As a result, they will be perceived as unique persons without the negative stereotypes that are traditionally associated with that specific group. This perception of people as unique individuals will further decrease the prejudice and discrimination that were brought about by intergroup differentiation.

It is also important to address the role of individual differences in the effectiveness of strategies for the reduction of prejudice and discrimination. For example, to decrease prejudice and discrimination, it is important to develop strategies that will counter right-wing authoritarian tendencies. One strategy that has received support is the personal value confrontation technique. With this technique, the discrepancy between people's ratings of the importance of freedom versus equality and their simultaneous support for discriminatory treatment of a minority group is highlighted. When the attention of high right-wing authoritarians is drawn to this incongruity, they tend to show improved attitudes toward the minority group.

In sum, based on knowledge of the processes that underlie prejudice and discrimination, strategies to counteract these influences can be suggested.

## **5. PROGRAMS FOR REDUCING PREJUDICE AND DISCRIMINATION IN EDUCATIONAL AND EMPLOYMENT SETTINGS**

Several programs for reducing prejudice and discrimination in educational and employment settings have been implemented. As shown in [Table I](#), one way of thinking about these programs is to categorize them

**TABLE 1**  
**Programs for Reducing Prejudice and Discrimination**  
**in Educational and Employment Settings**

---

**Enlightenment-based programs**

- *Multicultural education*: Provide information on the history and cultural practices of a wide variety of groups and emphasize standards of acceptance and inclusiveness
- *Cultural diversity training*: Attempt to promote value of group differences and understanding of how behavior is influenced by people's culture and background
- *Moral development training*: Emphasize standards of behavior and values of social justice, fairness, and equality

**Interaction-based programs**

- *Intergroup dialogues*: Allow members of different groups to address areas of contention through discussion and recognition of different perspectives
  - *Cooperative learning and cooperative work groups*: Encourage cooperative interdependence among members of different groups working toward a common goal
- 

as enlightenment-based programs or interaction-based programs.

The enlightenment-based programs generally involve increasing the knowledge that people have of other groups and altering people's perceptions of their relationships with others. These programs focus on changing the representation of members of other groups (e.g., reducing stereotypes, personalizing others, developing more inclusive representations) and on highlighting values relevant to tolerance and equality. For example, cultural diversity training programs are brought into industrial settings to teach managers and employees to value group differences and to develop increased understanding of other groups. Through this process, it is hoped that individuals will become committed to identifying and preventing the operation of stereotypes and prejudice in the workplace.

Interaction-based programs focus on bringing members of different groups together to provide direct interaction among members of the groups. They exert their influence through processes associated with intergroup contact. For example, cooperative work groups bring members of different groups together to work on a common task and emphasize cooperative interdependence whereby each person is dependent on the others for successful completion of the task. In the process, it is hoped that members of the different

groups will come to appreciate the contributions of all others, come to see each other as members of the same "team," and learn about the individual characteristics of others so that they are personalized.

Despite the plethora of programs to reduce prejudice and discrimination in educational and employment settings, the effectiveness of many of these programs has not been fully evaluated. Indeed, assessment of effectiveness depends critically on the specific outcomes desired. For example, is the goal to reduce stereotypes, change group representations, or reduce tension among members of different groups? Depending on the context and on the rationale for implementing a program in a new setting, one or another of these goals may be paramount. As a result, in choosing which program to introduce in an educational or employment setting, one must first identify the specific goals desired. Once an appropriate program is implemented, a continuing process of evaluation of effectiveness can be put into place.

### See Also the Following Articles

Employment Discrimination ■ Homosexuality ■ Interpersonal Conflict ■ Stereotypes

### Further Reading

- Brewer, M. B., & Brown, R. (1995). Intergroup relations. In D. T. Gilbert, & S. T. Fiske (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 554–594). New York: McGraw-Hill.
- Brown, R. (1995). *Prejudice: Its social psychology*. Oxford, UK: Blackwell.
- Crocker, J., Voelkl, K., Testa, M., & Major, B. (2000). Social stigma: The affective consequences of attributional ambiguity. In C. Stangor (Ed.), *Stereotypes and prejudice: Essential readings* (pp. 353–368). Philadelphia: Psychology Press.
- Esses, V. M., Jackson, L. M., & Armstrong, T. L. (1998). Intergroup competition and attitudes toward immigrants and immigration: An instrumental model of group conflict. *Journal of Social Issues*, 54, 699–724.
- Fiske, S. T. (1995). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, & S. T. Fiske (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 357–414). New York: McGraw-Hill.
- Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia: Psychology Press.
- Oskamp, S. (Ed.). (2000). *Reducing prejudice and discrimination*. Mahwah, NJ: Lawrence Erlbaum.

- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). San Diego: Academic Press.
- Stephan, W. G., & Stephan, C. W. (2001). *Improving intergroup relations*. Thousand Oaks, CA: Sage.
- Swim, J. K., & Stangor, C. (Eds.). (1998). *Prejudice: The target's perspective*. San Diego: Academic Press.



# Privacy

Carol M. Werner, Barbara B. Brown, and Irwin Altman

University of Utah, Salt Lake City, Utah, USA

1. Background and Definition
  2. Importance of Privacy Regulation
  3. Loneliness, Crowding, and Crime: Consequences of Adequate and Inadequate Privacy Regulation
  4. Technology and Privacy Regulation
  5. Summary and Conclusions
- Further Reading

## GLOSSARY

**crowding** Perception that there are too many people in a situation.

**privacy as a cultural universal** The notion that all societies have developed mechanisms for regulating openness and closedness.

**privacy mechanism** Behaviors used to change the amount of openness or closedness in an interaction.

**privacy regulation** Selective control of access to the self or to one's group, that is, control over the degree of openness or closedness.

Privacy regulation refers to how individuals and groups control interactions with others, especially about confidential or intimate information or with respect to physical contact or access. All societies have developed mechanisms for regulating privacy, and effective use of these mechanisms is related to individual and group viability.

## 1. BACKGROUND AND DEFINITION

### 1.1. Definition

According to Altman in 1975, privacy regulation is defined as the “selective control of access to the self or to one’s group.” Although standard dictionary definitions equate privacy with seclusion and withdrawal, psychologists study privacy regulation processes as a dialectic wherein people ebb and flow between seeking and avoiding social contact. There are many applications of privacy theory, including privacy policies in hospitals and other businesses, privacy policies on Web sites, security screening, and planning for privacy regulation in the design and use of home and community settings. All of these applications must accommodate people’s needs to be open as well as closed or else they will be inadequate. Creating information management systems or designing settings that only allow people to close themselves off makes it difficult, if not impossible, for people to share information and/or connect with others, whereas systems and designs that encourage openness, sharing, and accessibility run the risk of preventing people from guarding their own secrets and/or screening out unwanted information from others.

“Selective control” means that people aim to manage whether, when, and with whom to be open or closed. They actively negotiate how accessible they want to be about themselves and their personal information. Decisions about openness/closedness, as well as

abilities to control degrees of openness, vary with time and circumstances. For example, among friends and acquaintances, there are some people to whom an individual is almost always available (regardless of his or her other desires, obligations, and activities), some people the individual talks to unless he or she is busy, and other people the individual avoids most of the time. With respect to strangers, the individual may ignore them completely or may be open in civil and polite ways. In some situations, the individual may eagerly seek interactions with strangers and view them as opportunities to form new friendships, interact across cultures, and/or create a more civilized community (e.g., pleasantries in a park or at sporting events). At the same time, it is important for the individual to maintain reserve with strangers until safety is ensured. The general principle is that people constantly and selectively choose how open or closed they desire to be. People often act in ways that help to achieve their goals, creating blends of openness and closedness for short or long periods of time.

“Access” refers to a two-way process of both receiving information from others and giving information to others. People monitor both what comes in and what is given out, and they decide whether to allow the information to pass. Information includes physical stimuli such as sounds, odors, visual information, and even the presence of someone at the door or on the telephone. Other information that people aim to control is economic and financial such as social security numbers, credit card numbers, and banking information. Information can also be more psychological in nature such as intimate and confidential information about people’s criminal or sexual histories, health problems, grades, and incomes. People try to maintain control over what they learn about others and what others learn about them. The importance of controlling information varies with the nature of the information, individual preferences, time, and circumstances, and people put more or less effort into privacy regulation for different reasons.

“Self or group” is included in the definition to emphasize that sometimes individuals are the focus of analysis and sometimes dyads and groups are the focus. That is, individuals regulate interaction with other individuals or groups, and dyads and groups often act in coordination to regulate interaction with outsiders. For example, privacy regulation within a marriage involves regulation between the spouses, between the spouses and their respective in-laws, and between the spouses and their friends or outsiders.

In sum, according to Altman,

Privacy is an interpersonal boundary-control process, which paces and regulates interaction with others. Privacy regulation by persons and groups is somewhat like the shifting permeability of a cell membrane. Sometimes the person or group is receptive to outside inputs, and sometimes the person or group closes off contact with the outside environment.

Thus, privacy regulation is dynamic, and people constantly monitor their needs and opportunities and then decide what levels of openness and closedness to seek or accept.

## 1.2. Privacy Mechanisms

People use many “mechanisms” or strategies for achieving their desired levels of openness and closedness. Some mechanisms are verbal or paraverbal (e.g., intonations of words), whereas others are nonverbal (e.g., interpersonal distance, eye contact). Many mechanisms involve the physical environment; examples include opening or closing doors, choosing more sociable or more secluded settings, and creating territories by claiming and defining places and their boundaries. Many mechanisms are technological; examples include answering machines and caller ID that allow an individual to selectively answer or ignore telephone calls, filters on e-mail that allow blacklisting of “spam,” and home-centered filters on television channels and Web sites to control what information children can access. Mechanisms that facilitate openness include instant messaging, videoconferencing, and the ready availability of cell phones.

Individuals can use privacy mechanisms alone or in combination to seek out or exclude others and information about others. With respect to the intimacy level in an interaction, sometimes a subtle facial expression is sufficient and sometimes a facial expression, physical distance, and changing the topic all are needed to achieve the desired level of intimacy. Similarly, with respect to physical contact, sometimes a single mechanism is sufficient and sometimes additional mechanisms are needed. Some individuals who are very skilled at privacy regulation can achieve contact with a telephone call and/or can achieve needed solitude without offending others, thereby leaving open future opportunities for interaction. As an example of the latter, consider the difference between “Sorry, I’m busy right now” and “Sorry, I’m busy right now but I’ll call you when I’m free.” The second version achieves

short-term solitude while maintaining opportunities for future openness.

## 2. IMPORTANCE OF PRIVACY REGULATION

### 2.1. Importance

Privacy regulation serves critical individual and interpersonal functions. Individuals need to be able to regulate how much stimulation they receive and how much information others have about them. They need to control their physical environment and keep control over resources such as food, time, and space. Individuals and groups that can be easily intruded on, stolen from, forced to disclose secrets, and/or violated in other ways can lose a sense of self. Similarly, individuals and groups that cannot connect with others for personal or professional goals are isolated and unable to participate in interpersonal activities that are essential for functioning effectively in society. Individual and group viability is an important consequence of effective privacy regulation.

### 2.2. Privacy Regulation as a “Cultural Universal”

Because of its importance to effective functioning, privacy regulation is acknowledged as a “cultural universal,” meaning that every society has developed mechanisms so that individuals and groups can regulate interactions with others. Particular privacy mechanisms vary from one group to another, and use of privacy regulation mechanisms is usually guided by cultural norms, customs, and styles of behavior. Examination of societies around the globe shows that they vary in their preference for being relatively open or closed, but even the most open societies have mechanisms for closing themselves to others and even the most closed ones have ways of opening themselves up to others.

Most societies treat homes as places where individuals can withdraw, knowing that they can develop selected relationships and will rarely experience unwanted visitors. Visitors are expected to call ahead or announce themselves in other ways and then wait to be invited inside. In some situations, residents signal their availability to guests. For example, in traditional Egyptian Nubian homes, decorative plates are mounted on the homes’ exteriors but are removed

when the families are in mourning. The presence or absence of the plates signals to visitors whether they would be welcome. The Navajo in the southwestern United States have a custom that gives control over visitors to residents: Visitors do not knock, but they make themselves visible and wait for residents to invite them in. In some societies or for some groups of people, it is acceptable to use “keep out” mechanisms if people so choose, and no offense is taken when people leave the room, go on extended trips, or simply ignore visitors. In other groups, these and other “keep out” mechanisms are acceptable if people are busy working but are less acceptable if people are involved in other activities such as entertainment and loafing. The statement that someone “needs space” affirms the idea that people can legitimately vary between degrees of openness and closedness.

There are many privacy regulation mechanisms for the control of confidential information, and their use depends on many factors. Interpersonal distance controls access to some personal information (e.g., odors such as alcohol on the breath, physical details about age such as gray hair and wrinkles). Controlling verbal information is quite common as people choose to disclose or not disclose confidential information. In some groups or for some people, not only refusing to disclose but also actively lying is acceptable as a privacy strategy. Although people may use evasion or lying to control information, in many societies (e.g., the Javanese in Indonesia) lying is rarely needed because confidentiality is respected and it is considered rude to inquire too much about other people’s business.

In 2002, Petronio suggested that privacy rules and mechanisms often change and evolve so that they work most effectively for each group. She noted that many groups negotiate both general and specific privacy regulation rules. An example would be a group that shares confidential information and understands that the information should not be disclosed to the general public but negotiates specific details such as whether spouses or particular friends could be told some part of the story. Thus, the group co-owns the information, agrees to cooperate in protecting it, but negotiates particular boundaries around different parts of the information.

As computers make sharing of confidential information easier, mechanisms are being developed to control who has access and for what reasons. Privacy policies on Web sites promise that credit card information is protected from theft, that one can choose whether to be contacted, and that one’s contact information is not shared without permission. An important difference

between U.S. and European Union rules lies in the “opt-in” versus “opt-out” wording of these policies. In the United States, a company assumes that it can use or even share contact information for promotional purposes unless consumers explicitly notify the company to keep their identities confidential. In contrast, in the European Union, the opposite is assumed, and consumers must notify the company of their willingness to share identity information.

### 3. LONELINESS, CROWDING, AND CRIME: CONSEQUENCES OF ADEQUATE AND INADEQUATE PRIVACY REGULATION

#### 3.1. Dynamics of Privacy Regulation

As illustrated in Fig. 1, privacy regulation is a dynamic and ongoing process. People develop a desired level of contact with others, compare that desired level with their actually achieved level, and activate privacy regulation mechanisms (if needed and possible) designed to reach their desired levels of contact. When people desire more contact than they are experiencing, they feel lonely and select mechanisms that will bring them into contact with others. When people desire less contact than they have achieved, they describe themselves as crowded or overwhelmed and select mechanisms that will reduce their levels of contact.

Judgments about crowding and loneliness are very subjective and can be based on interaction partners, topics of conversation, and whatever alternative activities might be available. An individual can feel lonely in a crowd of people if the particular people the individual wants to be with are absent. The process can be very fluid. An individual can join a group and feel that his or her social contact is optimal, spend some time with that group but then begin to feel crowded or lonely, and then move to a different group and feel that social contact is optimal again. Problems with privacy regulation can occur when individuals or groups lack the behavioral and interpersonal skills needed to seek out or avoid contact with others or when the physical or social environment makes privacy regulation difficult.

#### 3.2. Viability

The ability to manage privacy is related to viability, that is, the ability to thrive, achieve goals, grow, and adapt to a changing environment. Research suggests that individuals and groups that do a better job of regulating interactions are more viable. Measures that tap viability have varied and include psychological variables such as satisfaction with the environment, satisfaction with social relations, improved task performance, greater task persistence, less stress, and other evidence that people are successful in achieving their short- and long-term goals. Several lines of research

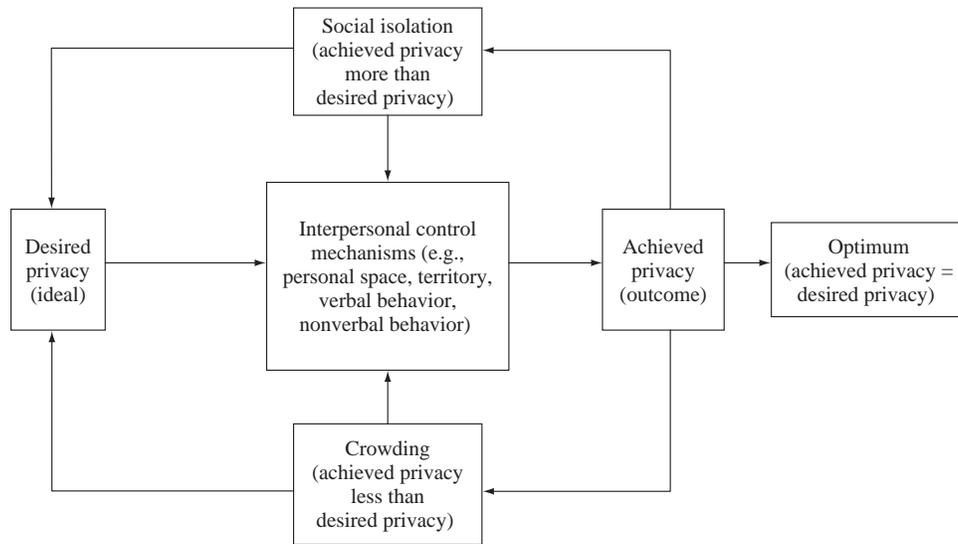


FIGURE 1 The process of privacy regulation.

have addressed relationships between privacy regulation and indexes of viability. These include loneliness, crowding, family and community relations, burglary and crime, and interaction stress in offices, prisons, hospitals, and other public settings. Typically, a combination of intrapersonal and sociophysical factors contribute to viability. For example, in one study, when pairs of naval recruits began an isolation mission by clearly establishing and maintaining separate territories that gave order to and allowed control over daily activities, they were more likely to complete the mission successfully. In contrast, pairs who did not create separate spaces did not perform as well on daily assignments and were more likely to abort the mission before the completion date. Thus, viability was related to successful management of privacy.

Another study asked dormitory residents what strategies they used for privacy regulation, that is, what they did when they wanted to make contact with others and what they did when they wanted to avoid contact. These strategies were related to one index of viability: whether the students chose to return to school the following year (excluding academic expulsions). Again, viability was related to effective strategy use. Students who remained in college were more likely to use both opening and closing mechanisms and to state that the mechanisms were successful. Students who dropped out reported using fewer of the successful mechanisms and also reported using other mechanisms that did not help them to achieve their desired levels of contact.

### **3.3. Loneliness and Privacy Regulation**

Loneliness is usually considered to be a problem stemming from a lack of individual skills, but solutions include both skills training and restructuring of the physical environment. Clinical and counseling psychologists have developed training programs designed to help people learn how to make and maintain friendships, and some of these programs have been quite successful. Equally important is research suggesting that physical and social environments play a role in supporting or discouraging contact. For example, one university study found that foreign students could regulate interactions with family members inside their apartments but had difficulty in making friends with residents in the courtyard. One explanation is that the foreign students had difficulty because the customs

and strategies for making contact in the United States were different from those in their countries of origin; for example, the Asian students did not feel comfortable in making casual “small talk,” whereas that was how U.S. students tended to initiate interactions. The housing director began to provide additional settings in which U.S. and foreign students could interact. Another study found differences in friendship patterns in inner-city public housing developments to be related to the physical environment. These researchers found that having green spaces close to residents’ apartments was related to increased informal contact among neighbors, larger social networks among adults, and higher rates of child–adult contact and interaction in the outdoors. In essence, residents were drawn to green spaces, and this increased proximity led to informal contacts that then led to deeper friendships.

The importance of the physical environment in providing opportunities for friendship formation is well known from research on proximity in housing developments as well as from research on public gardens as a context for friendship formation. A growing trend in community development is to use design features that invite people to spend time in community settings. These include streets that invite walking (e.g., storefronts at the sidewalk instead of behind parking lots, safe tree-shaded streets with attractive architecture, few cars or low speed limits, things to look at along the way, proximity to shopping and other services) and the availability of parks and green spaces where children and adults can congregate. These studies combine to suggest that coupling social skills training with supportive physical and social environments is all important for helping people to avoid loneliness.

### **3.4. Crowding and Privacy Regulation**

One of the most well-researched areas in privacy regulation is crowding. Unlike being “cramped,” where one person has too little space to accomplish tasks, crowding always involves more people than one desires or too little space for those people in a setting. Crowding occurs when an individual or group feels too accessible to others. For example, the phrase “two’s company, three’s a crowd” captures the idea that it is not the sheer number of people but rather one’s goals in an interaction that determines whether one experiences crowding. Similarly, being part of a crowd at a football game not only is not crowding but actually part of

the fun of being a fan is the presence of numerous others.

Theories and research on crowding have focused on “control” and the lack of control that accompanies the experience of crowding. Loss of control is an important construct because it connects the work on privacy regulation—and the theory’s central idea of selective control—with research on stress and stressors. Stressful events are often characterized by a loss or lack of perceived control. Sometimes, people experience crowding stress because they cannot control the amount or type of stimulation they receive from others such as crowd noises that overwhelm them. Related to this is crowding stress that occurs from uncontrolled interaction with others such as too many people at a party and too many neighbors in a housing complex. Sometimes, stress occurs because people interfere with or frustrate each other’s goal attainment, as might occur when too many people try to get on a subway or freeway, when a third person thwarts dyadic intimacy, or when there is no space to sit at the park or on the beach. Another common basis of crowding stress is resource based and occurs when there are too many people for the available food, water, and space. All of these bases of crowding are subtle variations on the idea that other people can undermine people’s sense of control over their own goals as well as over their interactions with others.

Studies conducted among college and university dormitory residents revealed that crowding was related to negative effects on friendship formation, lower satisfaction with and commitment to their academic institution, and lower grades. Housing three people in a room resulted in unexpected social dynamics. Male students tended to withdraw from the situation and leave the room to study at the library or spend time with friends. Female students tended to form “two against one” coalitions that left one student feeling alone and rejected. These problems were reversed when students were housed two to a room.

Social dynamics are also a problem in large dormitories (i.e., 32–40 residents) compared with both dormitories providing differentiated spaces and smaller dormitories (i.e., 6–20 residents). Sharing a dormitory with large numbers of residents makes it difficult to learn individual names and differentiate one student from another. It also makes it difficult to avoid social contacts when walking down the hall or using the shower. Initially, students in large residence halls may try to assert control and regulate their interactions, but eventually they stop trying as much to do

so. In one experiment, students in large dormitories reported a variety of social and educational consequences. They spent less time socializing, began to ignore other residents, complained about greater difficulty in controlling interactions, and said that they had more unwanted interactions. They also reported evidence of learned helplessness such as putting less effort into trying to control interactions with others and finding quiet places to study.

Research also shows that although people can tolerate high density for short periods of time, sustained high density for several weeks is stressful by itself and can exaggerate other stressors, including the stress of daily hassles. To add to these problems, friendships often deteriorate under the stress of high-density living, and people lose opportunities for social support that would ordinarily help to buffer them against stress.

In addition to helping to understand relationships between density and stress, these studies have been valuable in showing design features that give more control to residents and reduce crowding stress. For example, dormitories are now designed as suites or apartments to create smaller subgroups of residents. Breaking up long hallways into separate areas and providing kitchens, bathrooms, and social areas for each group gives residents manageable numbers of interactions, gives more control over who uses their spaces, reduces competition for resources, and restores order to interactions. Suite or apartment designs house six or more students together, with central common areas surrounded by separate sleeping areas. The common living room and bathroom encourage residents to function as a group, whereas the separate sleeping rooms each house two students, allowing them to withdraw from the group as needed but also allowing them to form closer relationships with a smaller number of residents. In theory, each cluster or pod then becomes a viable social unit that can interact with other similar clusters in the larger dormitory building.

Research has also examined how to optimize functioning in homes and apartments, regardless of whether residents are related, involved, or living together as roommates. Some groups develop effective rules that allow individuals to regulate interactions with others. People can use open and closed doors and verbal messages (e.g., “I need to study now”) to signal their accessibility and inaccessibility. In many cases, design features can support such regulation. People can withdraw to their bedrooms, can study, or can use the bathroom if they desire to be alone.

Research indicates that homes and apartments with highly differentiated spaces (i.e., more small rooms with separate access via hallways) give residents more control over their interactions (i.e., allow them to decide whether to be open or closed to others) and reduce crowding stress. The ability to avoid contact must be complemented by opportunities for making contact. In most apartments and homes, people can gather in common areas, such as living rooms and family rooms, when interaction is desired. Often, people are attracted to rooms with activity opportunities such as television sets, games, and stereos. These areas become “behavioral focal points,” that is, spaces whose design, location, and function invite people to gather together.

Another design feature that supports privacy regulation is the “privacy gradient,” that is, the idea that places in homes vary in their physical accessibility to outsiders. This physical accessibility should parallel residents’ desired accessibility. Some rooms and spaces are close to the entrance, whereas others are deeper in the home’s interior and so are less accessible to outsiders. The front yard, main entrance, and entryway all are fairly accessible to anyone who desires contact. Living rooms, dining rooms, and guest bathrooms tend to be formal areas where strangers and visitors might be invited. Places such as the kitchen, family room, bedrooms, and family bathrooms are typically used exclusively by residents. Managing interactions may be facilitated when the physical layout of the home or apartment puts public areas closer to the entrance and private areas far from the entrance and less accessible to visitors.

### 3.5. Privacy Regulation in Other Settings

Research on residential crowding has been paralleled in prisons, homeless shelters, hospitals, and similar settings, and the results have generally been similar. Living in undifferentiated spaces with numerous others is stressful. Aggression and crime can be high, and residents feel vulnerable and unable to control interactions. For example, in dormitory-style prisons, there were high aggression rates as well as high stress-related illness rates. Redesigning these spaces with simple barriers that gave each inmate a separate area around a bed reduced stress levels considerably. Consistent with the idea that privacy regulation involves openness and closedness, the prisons did not completely enclose each prisoner. Walls were low enough and entryways

had no doors, so that guards could ensure safety for each prisoner and monitor for illegal activities.

### 3.6. Crime and Defensible Design

Public housing in the United States has become notorious for its physical unattractiveness, lack of upkeep by residents and staff, and high crime rates that include drug abuse, minor and major property crimes, gang warfare, and even murders and drive-by shootings. For more than 50 years, social scientists have tried to understand the bases of this disorder so as to correct it. A central issue is the extent to which the housing developments themselves undermine positive social relationships. Comparisons of different configurations indicate that developments with more “defensible designs” yielded more viable communities. Design elements allow residents to differentiate legitimate users from outsiders, encourage residents to use outdoor spaces and become protective of their own territory, and allow residents to use passive surveillance to monitor legal and illegal activities. In this way, defensible design can be viewed as a privacy regulation mechanism. There is no simple panacea to problems associated with many housing developments, but some essential design features include limiting the numbers of residents so that they can recognize one another and distinguish residents from outsiders (this is often accomplished with relatively low numbers of floors and limited numbers of apartments on each floor and with separate entrances and separate elevators), windows that allow surveillance of the outside area, outside gathering areas that provide activities and seating for multiple age groups, attractive settings to increase pride in place, and clear differentiation of the private setting from the public street.

Most of the research on defensible design has been done in public housing developments. These ideas have been replicated in middle-class suburban neighborhoods, where homes and neighborhoods with more defensible space are less likely to be burglarized. Neighborhoods are less likely to attract burglars when they appear to have stronger territorial claims, that is, when the homes are personalized and the homes and neighborhood convey the impression that residents feel a sense of ownership over the space and would react to the presence of suspicious strangers. The roads appear to be less public (i.e., fewer traffic signs of all types), sites allow surveillance from one home to another, residents appear to spend time on the block, and intruders are often challenged by residents or reported

to the police. Sense of territorial control is also apparent at each home site. The home and yard are personalized with names and house numbers, and there is often evidence that someone is at home (e.g., toys and tools are in the yard, mail and garbage are picked up). Surveillance is clearly possible because shrubs do not obscure the entrance or the windows, and homes are sited so that neighbors can observe each other's homes. When burglars are asked to indicate which homes they are not likely to target, they usually indicate the homes where they would expect the resident or a neighbor to challenge them if they tried to enter.

In sum, the physical environment serves important privacy regulation functions. Environments that provide clear differentiation between public and private areas are less likely to be intruded. Neighborhoods and housing developments whose designs support neighborly friendships also support creation of neighborhood territories and defensible spaces. These spaces also contribute to enhanced control over who can enter and who cannot as well as to what people can do in these spaces.

## **4. TECHNOLOGY AND PRIVACY REGULATION**

### **4.1. Communication Technology and Privacy**

As people increasingly use technology to connect them to the world, people are also developing technological mechanisms that allow them to control their accessibility and inaccessibility to others. People use cell phones, telephones, fax machines, e-mail, and the World Wide Web 24 hours a day to contact friends and business associates. They can meet new people, learn new ideas and philosophies, try on new personas, and even fall in love using Web-based exchanges. They can also send pictures through their telephones and the Web. Television is expected to allow Web access that will expand the ability to select programs, movies, sports events, and so on. Sophisticated cameras provide surveillance to increase safety in parking garages, on university campuses, and elsewhere, and cameras are used to cite motorists for speeding and other violations.

Each of these opportunities comes laced with issues essentially involving tensions between openness and closedness. How and when do people control their accessibility to others? How can people trust the honesty of others they meet online? If people have cell

phones, should they be available 24 hours a day? Does owning a telephone and having telephone service mean that telephone solicitors can contact people, or can people decide who can call them and who cannot? Do Web and television access mean that children can view material that is intended for adults? How can children be protected from Web-based crime, including adults who use chat rooms as a base for contacting and harming children? When does surveillance become intrusive?

Extensive discussions of these and related issues have begun to lead to a variety of solutions. This article addresses proposals in the United States, although other countries are dealing with similar issues. Efforts are ongoing to find ways of modifying and improving these control systems so that they provide opportunities for openness while protecting people from serious harm as well as minor inconveniences. Just as different but effective privacy mechanisms have evolved in different cultures, privacy mechanisms are continuing to evolve as new technologies open the world to people. Some strategies that allow selective access already exist such as caller ID on telephones, unlisted telephone numbers, listed numbers that omit home addresses, and Web search engines that screen children from pornography but allow adults wider access.

At the same time, marketers selectively target particular consumers using information gathered on the Web, grocery store "preferred customer" spending patterns, and credit card files and other sources. Advertisers seek to use this personal information to target marketing to the most appropriate and receptive consumers. They emphasize the importance of free and open access to consumers as a way of maintaining a vital economy. Knowing individuals' family sizes and configurations, buying habits, business and recreation preferences, annual incomes, and other information allows companies to target consumers with promotional materials. Some consumers enjoy receiving this material, regardless of whether it comes through the Internet, telephone, or mail, whereas others consider these contacts to be intrusions and seeking ways of preventing such marketing. Privacy regulation strategies include state and federal legislation as well as commercially available screening systems and personal controls.

Faxes and telephone calls are regulated by the Telephone Consumer Protection Act of 1991. When fax machines first became widely available, marketers sent unsolicited promotional materials. Because consumers could show evidence of harm—the recipient had to pay for paper and ink, and incoming offers tied up their fax machines—federal rules supported

recipients over senders. Faxes could not be sent without permission or an existing relationship (e.g., organizations could fax conference announcements to members). A more restrictive rule, requiring recipients' written permission before faxes can be sent, was under review at the time of this writing.

Similarly, telephone solicitors created sufficient problems (e.g., calling at dinnertime; taking financial advantage of people) that federal policies and supportive legislation have been put into place to allow people to put their names and telephone numbers on "do not call" lists, with no charge to consumers but with hefty penalties to marketers for violations. The policy was challenged in court and was on appeal at the time of this writing. This list is for interstate commerce, and each state must enact its own "do not call" list to regulate local marketing calls. Some states want consumers and marketers to share the cost of this service, indicating their belief that telephone solicitors have a right to make calls and that consumers should pay to assert their right not to be called.

E-mailed solicitations (derisively called "spam") have been more difficult to stop, in part because there is less apparent harm to consumers and marketers claim that unsolicited e-mails could be seen as a necessary nuisance. Also, many senders immediately change their e-mail addresses or otherwise conceal their identities, making it difficult to screen future mailings or locate the senders if the behavior is declared illegal. Evidence of harm has been claimed by Web administrators whose server equipment must handle large numbers of unsolicited and unwanted e-mails daily. These administrators are pressuring the federal government to stop spam.

#### **4.2. Health Care Privacy**

The appropriate blending of openness and closedness is also a central issue in health care privacy. Patients may wish to be accessible by sharing test results among several doctors or by receiving information about new medications and medical procedures for their particular conditions, they may need an efficient way in which to get billing information to their insurance company, they may wish to receive test results on the telephone, and so on. Therefore, delimited openness of medical information can be beneficial. On the other hand, patients may wish to disseminate only information about particular medical conditions, may be uninterested in hearing from marketers, may want to withhold some information from insurance companies (even to the point of paying privately for some tests so that

insurance companies will not perceive them to be risky clients), may want to receive test results only on a confidential telephone line (rather than having a coworker take a message), and so on. A major concern has been that patients' fears about control over information might make them unwilling to be open and honest with their doctors, thereby putting their health at risk. Patients might not disclose illegal or embarrassing behaviors, past health problems, health problems of relatives, and other private concerns, lest this information be widely disseminated without their explicit permission. People may opt out of genetic tests that can yield important information because they fear that employers might get the information and use it for insurance coverage or employment decisions.

Federal legislation begun during the Clinton administration and modified during the Bush administration has addressed some of these issues. The Health Insurance Portability and Accountability Act of 1996 (HIPAA, modified in 2002) requires a single signed consent to allow access in three domains: "treatment" (provision of health care and related services), "payment" (submitting insurance claims and determining eligibility), and "health care operations" (broadly defined sharing of information). The bill has been criticized for being coercive (people in health plans often must sign to receive treatment) and for the vague and broad scope of the third provision. Others have commended the plan for bringing a national uniform rule to these issues and for increasing access as well as control. For example, it is now easier for doctors to share patients' medical records, and there are fewer insurance approvals to sign (due to the single signature requirement). Control over information is more specific under this rule. Psychological and psychiatric records cannot be shared without a separate signature specifically for that information. With respect to test results, doctors must now ask how to contact the patient and whether leaving a message is acceptable. Also, without explicit permission, hospitals cannot even disclose whether someone is a patient, let alone disclose health status.

It has taken many years and much discussion to develop and enact HIPPA. Balancing access with confidentiality is a difficult task, especially when widespread use of computerized databases makes it easy to share confidential information and easy to inadvertently or deliberately share information that should be protected. Some patient advocates consider HIPPA to favor the medical system over individuals' rights to privacy, and reform efforts are under way. This is an issue that will receive ongoing attention as people seek

to achieve desired openness and closedness under changing circumstances.

## 5. SUMMARY AND CONCLUSIONS

### 5.1. Overview of Importance of Privacy Regulation

Privacy regulation is the selective control of access to the self or group. All societies have developed strategies through which their members can regulate interactions with others, although each society sets different standards of openness and closedness and has evolved different mechanisms for regulating privacy.

Privacy regulation involves many different kinds of events—from people standing too close in subways to burglaries being committed in neighborhoods, from concealing information about one's test scores to lying about medical conditions to obtain insurance. Privacy regulation also involves managing relationships to achieve a desired level of intimacy, perhaps by disclosing secrets to deepen a relationship or by avoiding listening to disclosures to maintain the relationship at a less intimate level. Privacy regulation involves choosing the levels of openness and closedness that one desires and then trying to achieve those levels by activating various privacy regulation mechanisms. Individuals and groups that can regulate their privacy are more viable and more likely to achieve their goals. When desired privacy is not achieved, people often complain of stress.

### 5.2. Prospects for the Future

Privacy regulation has been practiced and refined for centuries, yet people still have difficulty in managing interactions. This is due in part to differences between individuals in expectations about how open people should be and how to convey this through privacy regulation mechanisms. These problems occur in part due to cultural differences. As people immigrate to new countries, residents and new immigrants must recognize differences in privacy goals and strategies. Possibly, once people are aware of these differences, they can work on accommodating each other's preferences or on recognizing the cues used to signal whether the other wishes to be more open or more closed. In a changing situation, privacy regulation mechanisms and desires can also be changed.

Another source of problems is the rapidly changing technological environment and pressures to use

technology for more openness. Psychological research on privacy suggests the following principles for guiding the use of these technologies:

1. *People should have reasonable control over others' access to them and people's access to others.* People should be able to proactively and reactively open and close themselves to others as they desire and see fit. This principle highlights privacy as a dialectic process of opening and closing rather than closing alone. It also emphasizes the importance of both a proactive and a reactive control capability. People should be able to proactively initiate openness or closedness rather than just react passively to the world.

2. *People may need a new or enhanced repertoire of regulatory mechanisms to control openness and closedness during the technological era.* Traditional face-to-face communication uses particular verbal, paraverbal, environmental, and cultural mechanisms. Although some of these mechanisms may apply during the technological arena, additional mechanisms—unique to new modes of communication—might be necessary.

3. *A dynamic privacy regulation system, with the ability to shift desired openness and closedness as circumstances change, will continue to be necessary and may require innovative applications of technology.* For example, some people like the anonymity of chat rooms, whereas others fear being lied to and deceived. To address this, some Web chat rooms may continue to be wide open and allow people to adopt personas and play out fantasies. In other chat rooms, participants may insist on knowing background information on each other and may agree to post authenticated vitae as the price for participating.

4. *Awareness of the importance of privacy regulation must be at the forefront of technological innovations.* Mechanisms that allow flexibility in achieving openness and closedness will be most successful.

### See Also the Following Articles

Environmental Stress ■ Personal Space ■ Residential Satisfaction and Perceived Urban Quality

### Further Reading

- Altman, I. (1975). *The environment and social behavior: Privacy, personal space, territory, crowding*. Pacific Grove, CA: Brooks/Cole.
- Altman, I., & Chemers, M. M. (1980). *Culture and environment*. Pacific Grove, CA: Brooks/Cole.
- Altman, I., Vinsel, A., & Brown, B. B. (1981). Dialectic conceptions in social psychology: An application to social

- penetration and privacy regulation. *Advances in Experimental Social Psychology*, 14, 107–160.
- Baum, A., & Epstein, Y. (Eds.) (1978). *Human response to crowding*. Hillsdale, NJ: Lawrence Erlbaum.
- Baum, A., & Valins, S. (1979). Architectural mediation of residential density and control: Crowding and the regulation of social contact. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 131–175). New York: Academic Press.
- Brown, B. B. (1987). Territoriality. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (Vol. 1, pp. 505–531). New York: John Wiley.
- Evans, G., Lepore, S. J., & Schroeder, A. (1996). The role of interior design elements in human responses to crowding. *Journal of Personality and Social Psychology*, 70, 41–46.
- Lepore, S. J., Evans, G. W., & Schneider, M. L. (1991). Dynamic role of social support in the link between chronic stress and psychological distress. *Journal of Personality and Social Psychology*, 61, 899–909.
- Margulis, S. T. (Ed.) (2003). Contemporary perspectives on privacy: Social, psychological, political [special issue]. *Journal of Social Issues*, 59(2).
- Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. Albany: State University of New York Press.



# Productivity

Robert D. Pritchard

University of Central Florida, Orlando, Florida, USA

1. Introduction
  2. Defining Productivity
  3. The Importance of Productivity
  4. Improving Productivity
  5. Approaches to Productivity Improvement
  6. Measuring Productivity
  7. Purpose of Measurement
  8. Specific Approaches to Measuring and Improving Productivity
- Further Reading

## GLOSSARY

**effectiveness** A measure of outputs compared with goals or standards; examples of effectiveness measures would include actual billable hours divided by the standard or expected amount of billable hours and actual percentage of returning customers compared with the goal of returning customers.

**efficiency** A measure of outputs divided by inputs; outputs are the products or services of the organization such as radios produced, customers satisfied, and/or clients seen, whereas inputs can be all inputs needed to produce the outputs such as labor, equipment, supplies, energy, and buildings.

**feedback** Information provided about employee behavior and outputs; it can be in many forms, including providing information on the amount or quality of work completed, subjective ratings of performance by managers, information provided informally by peers and subordinates, and a person's own perceptions about his or her work done.

**goal setting** A process whereby the subordinate and his or her supervisor jointly agree on a specific quantitative level of output that the subordinate will produce during the next work period (e.g., increasing revenues by 4% in the next quarter); less formal goal setting might be plans to make improvements or general intentions to perform better in the future.

**incentives** Outcomes of value that are provided to employees based on their levels of performance; most common are monetary outcomes (e.g., bonuses, incentive pay) and outcomes that have monetary value (e.g., vacations, sports equipment).

**performance** A general term that can refer to the amount or quality of the outputs produced, an evaluation of the outputs produced, or some combination thereof; it can be applied to an individual or to a much larger unit such as the total organization.

**performance appraisal** A process whereby one's supervisor, and sometimes one's peers and subordinates, evaluate the performance of an individual and provide this evaluation as feedback to that person; it is typically done with a series of ratings on various dimensions of performance such as technical knowledge and oral communications skills, but it can also include objective data on the quantity or quality of work done by the individual.

**performance management** A general term that includes all techniques used to influence and improve the performance of employees in organizations; such techniques include training, feedback, control systems, goal setting, and incentive systems.

**productivity** A term that can be defined in multiple ways, with the most traditional being outputs divided by inputs, usually expressed in monetary terms; it is also commonly defined as efficiency (outputs over inputs) plus effectiveness (outputs relative to a standard or goal).

*productivity improvement* Improvement in productivity measures over time.

*role clarification* Techniques used to make the expectations of the organization clear to the individual; examples include discussions about expectations with one's supervisor, formal systems that identify the relative importance of different parts of the work, and information systems that indicate what the priorities are for making performance improvements.

Productivity is an important topic, and the field of psychology has made many contributions to understanding and improving productivity in organizations. Productivity, especially improving productivity, has important benefits ranging from global effects to effects on individuals' quality of life. There are a number of different definitions of productivity, and it is important for communicating with others to be clear on what definition is being used. Psychologists focus on improving productivity by influencing people's behavior, especially through techniques such as feedback, motivation, goal setting, incentives, and role clarity. Measuring productivity is complex, and this article describes both the purpose of the measurement and the criteria for good measurement. Once measured, productivity is normally used as feedback, so the article summarizes criteria for optimal feedback design. Finally, the article notes specific approaches to measuring and improving productivity.

## 1. INTRODUCTION

### 1.1. History

Labor economists originally used productivity as a formal term during the early 1900s. During the 1950s and 1960s, the United States was the most productive country in the world, and there was little doubt that this state of affairs would continue. However, by the 1970s, American superiority in productivity was being threatened by productivity in other countries, and this was seen as a major concern that would lead to serious long-range economic problems. For example, the success of the Japanese in U.S. automobile and steel markets led to plant closings and lost jobs, major U.S. companies were in severe financial difficulties, and what had historically been a trade surplus became a trade deficit that was increasing.

### 1.2. Productivity as a National Issue

Although many causes were responsible for these problems, loss of clear U.S. superiority in productivity was seen as one of the major ones. Because of this, productivity and productivity improvement became issues of great social and political importance, not only in the United States but in all industrialized nations. This led many other disciplines, besides economics, to focus on productivity issues, and one such discipline was psychology.

## 2. DEFINING PRODUCTIVITY

Defining productivity is a difficult task. Since the 1970s, behavioral scientists of many types have worked in the area of productivity, resulting in a proliferation of definitions of productivity. However, the vast majority of definitions fall into one of three categories.

### 2.1. Efficiency

The first category defines productivity as an efficiency measure where outputs are divided by inputs. Thus, the number of outputs generated, such as refrigerators, potatoes, or university graduates, is divided by the inputs used. Such inputs usually include raw materials, labor, energy, and so forth. Typically, both outputs and inputs are measured in terms of cost, so that the two can be divided to form a ratio that is interpretable across different settings. This definition is used by most economists and by some behavioral scientists.

### 2.2. Combinations of Efficiency and Effectiveness

The second definition of productivity is a combination of efficiency and effectiveness. Efficiency is the output/input measure used in the first definition. Effectiveness is the ratio of outputs relative to some standard or objective. For example, the number of trucks produced relative to a specific quantitative goal is an effectiveness measure. Those taking the approach that productivity should include both efficiency and effectiveness argue that both are necessary for a complete definition of productivity. One such definition is that productivity refers to how well an organization uses its resources to achieve its objectives.

### 2.3. Broader Definitions

The third type of definition is the most broad one and essentially includes as productivity any characteristic that makes the organization function better. This approach includes efficiency and effectiveness but also includes factors such as quality of output, work disruptions, absenteeism, turnover, and customer satisfaction.

### 2.4. Definitions and Clear Communications

It is important for anyone working in the area of productivity to be aware that these different definitions exist. Some researchers and practitioners are very rigid in their choice of definition and do not accept that the other definitions are “productivity.” Thus, for effective communication, the same definition of productivity should be used by all.

## 3. THE IMPORTANCE OF PRODUCTIVITY

Productivity and productivity improvement are important because they influence many aspects of our lives. They have effects that range from global effects to those on specific individuals.

### 3.1. Global and National Benefits

Increased productivity means generating the same goods and services with less inputs. Thus, it is a way in which to conserve societal or global resources ranging from trees to human labor. Increased productivity also allows for more outputs to be available for the same inputs. Thus, we can be closer to a global society of plenty while using less of our global resources. At the national level, there are clear ties between productivity and important economic outcomes such as controlling inflation. For example, the real cost of goods is influenced by productivity because productivity improvement results in producing the same goods for lower costs. Assuming a roughly constant profit margin, the real cost of goods decreases as productivity increases.

### 3.2. Benefits for Industries and Specific Organizations

Productivity and productivity improvement are also important at the levels of the industry and the individual

firm. If productivity improvement of an industry or a firm is higher than that of its competitors, that industry or firm will more likely survive and prosper. Costs and prices are lower, thereby making those products and services more competitive in the marketplace. This leads to higher sales, greater profits, and more jobs.

### 3.3. Benefits for Individuals

Productivity improvement can also increase the quality of our lives. Productivity gains will lead to efficient use of our time and more leisure time, and high productivity is a key to advancement in organizations. Equally important, people like to be productive. It is a central aspect of self-fulfillment and self-respect.

## 4. IMPROVING PRODUCTIVITY

The major reason why people focus on productivity is to improve productivity. It is by improving productivity that the gains from the global level to the individual level can occur.

## 5. APPROACHES TO PRODUCTIVITY IMPROVEMENT

There have been two major approaches to improving productivity. The first is through a change in technology. Some economists argue that most or all major improvements in productivity can be traced to technological improvements. The other approach is behavioral. Behavioral scientists argue that new technology can improve productivity only if organizational personnel use it effectively; thus, technological improvements also require a behavioral component. Furthermore, behavioral scientists argue for the importance of issues such as performance measurement, motivation, feedback, goal setting, incentives, and role clarity. It is here that the field of psychology has made the greatest contributions.

## 6. MEASURING PRODUCTIVITY

To improve productivity, it is important not only to measure it but to measure it well. However, productivity measurement is very complex, and a number of issues must be considered in designing a productivity measurement system.

## 7. PURPOSE OF MEASUREMENT

The first step is to decide on the purpose of the measurement because that decision will affect how the measurement is done. One purpose in measuring productivity is to compare large aggregations of organizations with each other. Examples of this purpose include comparing the United Kingdom's economy with Japan's economy and comparing the electronics industry with the automobile industry. Another purpose is to evaluate the overall productivity of individual organizations for comparison with each other or with some standard. A third purpose is as a management information system or control system. Here the focus is on a single organization, and the measurement deals with the functioning of the human/technological system. Finally, productivity could be measured for use as a motivational tool. This is the typical purpose of the organizational psychologist. The assumption is that if individuals change their behavior appropriately, productivity will increase.

### 7.1. Criteria for a Good Productivity Measurement System

There are many specific issues that must be considered in measuring productivity by the psychologist or organizational behavior specialist. These include ensuring that all important aspects of the work are measured, having reliable and valid measures that are cost-effective to collect, using measures that are understood by all, using measures over which the personnel have control, ensuring that the measurement system will be accepted by both the managers and the people doing the actual work, and ensuring that measures for each unit are aligned with the strategic objectives of the broader organization.

In most approaches to productivity improvement, the productivity measurement system is used for feedback to the personnel doing the work. The assumption is that such feedback will help to produce the behaviors that will increase productivity. However, to be most effective, the feedback system must meet an additional series of criteria beyond those for the productivity measurement system. Feedback must be given in a timely manner, be understood by those receiving the feedback, contain both descriptive information (e.g., the unit finished 15 orders) and evaluative information (e.g., 15 orders is very high performance), give information on all important aspects of the work, account

for the varying importance of different aspects of the work, provide an overall productivity score, identify improvement priorities, and allow for unit personnel to plan new strategies for doing the work.

Research has indicated that such feedback can be very effective in improving productivity. For example, one approach developed by Pritchard and associates measures productivity and uses these measures for feedback. It is called the Productivity Measurement and Enhancement System or ProMES. Figure 1 shows the effects of feedback due to ProMES as shown from field studies on 55 organizational units located in seven different countries. The nature of the organizations varied, and the research was conducted by nine different research groups. The horizontal axis represents time beginning with baseline (prior to the use of feedback) and continuing through the implementation of feedback. The vertical axis represents the overall effectiveness score, which is an index of overall productivity. Once feedback was introduced, there were large increases in productivity.

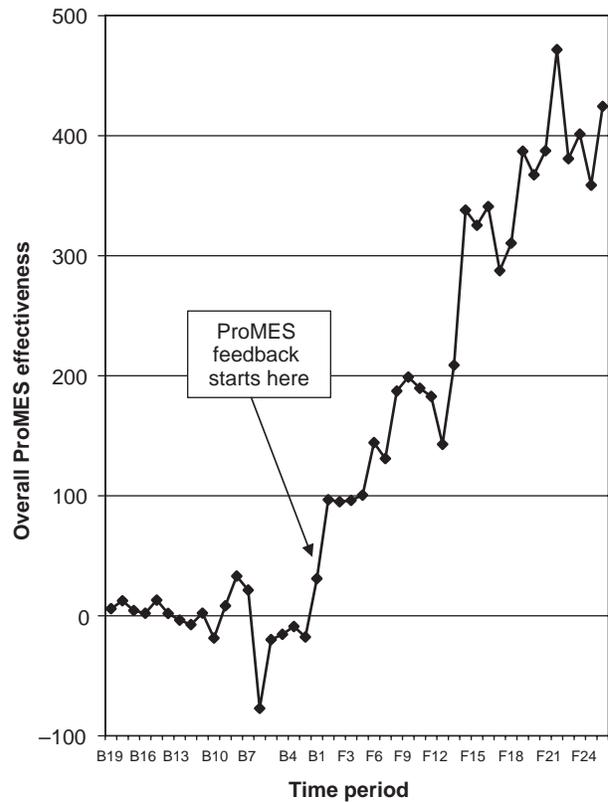


FIGURE 1 Effects of ProMES feedback on productivity. Reprinted from Pritchard and colleagues (2002).

## 8. SPECIFIC APPROACHES TO MEASURING AND IMPROVING PRODUCTIVITY

Psychologists and other organizational specialists have proposed a number of other specific techniques for measuring productivity that focus on using productivity measures as feedback to help personnel improve their productivity. The best-known examples are the Balanced Scorecard and Six Sigma.

The Balanced Scorecard was originally designed as an aid to top management. It stressed that other measures besides financial measures are important to overall organizational effectiveness and that including these nonfinancial measures would result in a more “balanced” evaluation. It has since been broadened and is now used for all levels of the organization. Six Sigma was originally developed for feedback and productivity improvement in manufacturing settings. The central idea is that processes with multiple steps, such as assembling cars and refining petroleum, require that each step be done with extremely high reliability for the final product to meet required specifications. Measures of each step in the process are taken, and efforts are made to increase the reliability of each step.

Finally, there is a class of practices that are designed to improve performance and thereby productivity. These practices, which have come to be called “performance management,” include performance appraisal, other forms of feedback, goal setting, incentives, and role clarification.

## See Also the Following Articles

Competence at Work ■ Groups, Productivity within  
■ Work Motivation

## Further Reading

- Campbell, J. P., & Campbell, R. J. (Eds.). (1988). *Productivity in organizations*. San Francisco: Jossey-Bass.
- Dasgupta, T. (2003). Using the Six Sigma metric to measure and improve the performance of a supply chain. *Total Quality Management*, 14, 355–366.
- Guzzo, R. A., Jette, R. D., & Katzell, R. A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. *Personnel Psychology*, 38, 275–291.
- Kaplan, R. S., & Norton, D. P. (1996). *Translating strategy into action: The Balanced Scorecard*. Boston: Harvard Business School Press.
- Kendrick, J. W. (1984). *Improving company productivity*. Baltimore, MD: Johns Hopkins University Press.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57, 705–717.
- Pritchard, R. D. (1992). Organizational productivity. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3, pp. 443–471). Palo Alto, CA: Consulting Psychologists Press.
- Pritchard, R. D., Holling, H., Lammers, F., & Clark, B. D. (Eds.). (2002). *Improving organizational performance with the Productivity Measurement and Enhancement System: An international collaboration*. Huntington, NY: Nova Science.
- Sink, D. S. (1985). *Productivity management: Planning, measurement and evaluation, control and improvement*. New York: John Wiley.



# Pro-environmental Attitudes and Behavioral Change

Henk Staats

Leiden University, Leiden, The Netherlands

1. Introduction
  2. Environmental Problems as Social Dilemmas
  3. The Relation between Attitudes and Behavior
  4. How to Effect Pro-environmental Behavior Change: The Role of Psychology
  5. Considerations Regarding the Effectiveness of Interventions for Pro-environmental Behavior Change
  6. Conclusions
- Further Reading

## GLOSSARY

**attitude** The general evaluation of an object; in the theories described in this article, it refers to the general evaluation of a behavior.

**intervention technique** A method to change behavior in a certain direction; in this article, it refers to a method to change anti-environmental behavior to pro-environmental behavior.

**pro-environmental behavior** Behavior that is relatively favorable for the environment in comparison with behavior that serves the same primary function (e.g., using public transportation instead of driving a car).

**social dilemma** The tension between individual interests and collective interests.

To understand pro-environmental behavior and be able to influence it, it is essential to apply knowledge

from attitude-behavior theories. However, the central characteristic of pro-environmental behavior can be understood only if knowledge from social dilemma theory is applied simultaneously.

## 1. INTRODUCTION

Large-scale environmental problems have been on the political agenda for more than 30 years but continue to threaten the earth's inhabitants. Clearly, some of the problems are particularly stubborn and defy easy solutions. Basic to an understanding of environmental problems is the fact that they are ultimately phenomena created by human behavior and, thus, are to be solved by human behavior. Gradually, this has become recognized in a field that was traditionally dominated by experts from the physical sciences such as chemists, biologists, and ecologists.

This article gives an overview of contributions from social/environmental psychology that deepen insights into the human dimension of environmental problems and that are helpful in designing interventions that change behavior in a pro-environmental direction. First, the article describes a number of research traditions that are highly relevant for understanding behavior that affects the environment. These deal with attitude-behavior relations, moral considerations, and

habits. Overarching these perspectives is the conceptual framework of social dilemmas, a paradigm that is essential for understanding the nature of environmental problems. The article then describes what psychology has to contribute when considering ways of changing behavior in a pro-environmental direction.

## **2. ENVIRONMENTAL PROBLEMS AS SOCIAL DILEMMAS**

Essential in thinking about the origin of many environmental problems is their social nature, originating from a fundamental tension between acting in the service of direct individual interests and acting in the service of long-term community interests. Many societal issues have this structure, and it is almost essential to an organized collective that individuals sacrifice, to some degree, their individual interests in favor of the interests of the community. Consider, for example, the principle of paying tax to finance collective institutions and services such as police, health care, and infrastructure for transportation. From an individual perspective, it would be profitable to evade the obligation of paying taxes. But when everyone would succeed in doing this, no public good would become available and everybody would be worse off. This is the core of the social dilemma that lies at the heart of many environmental problems. It was described in 1968 by Hardin as “The Tragedy of the Commons.” That publication has become emblematic for the analysis of environmental problems.

Different in comparison with the taxpayers’ dilemma is that environmental problems are often delayed in time and that effects are not necessarily manifest at the place where the problem is created. Consider the greenhouse effect, one of the most serious environmental problems of our time. One reason why the greenhouse effect is difficult to solve is that effects are expected to be delayed some 50 years. Also, it will become most threatening in parts of the world that have hardly contributed to the problem but that will be most vulnerable to the expected rise in sea level.

The social dilemma paradigm has proven to be a fruitful area of social-psychological experimental research, providing information about which factors are likely to influence the trade-off between choosing for the private interest and choosing for the collective. Important factors include the payoff structure (i.e., how much the individual loses by acting in the

collective’s interest), communication about the decision among group members, expectations about the others’ cooperation, perceived efficacy (i.e., how important the individual contribution is in determining the outcome of the group), and identifiability (i.e., the degree to which a person is known to have decided in favor of individual interests vs collective interests).

## **3. THE RELATION BETWEEN ATTITUDES AND BEHAVIOR**

The social dilemma framework helps one to understand how behavior can be differentially driven toward individually valued outcomes or collectively valued outcomes such as the quality of the environment. However, it gives only a limited account of how behavior results from cognitive and motivational processes. Three theories are specifically important in the context of environmental behavior.

### **3.1. Theory of Reasoned Action**

One of the most influential models in the social-psychological literature is the theory of reasoned action (TRA) developed by Fishbein and Ajzen. The TRA postulates that behavior is the result of three main factors: attitude toward behavior, subjective norm, and behavioral intention. Attitude and subjective norm are postulated to cause behavioral intention, which in turn causes behavior.

In the TRA, Fishbein and Ajzen confined the attitude concept to the general evaluation of a behavior. Other theories have no restrictions as to the nature of the attitude object and, thus, may consider ideas, persons, things, issues, ideas, or other entities. However, these more general theories have fared less well when it was important to understand and predict behavior. Therefore, at the time, Fishbein and Ajzen’s theory came as a much-needed innovation in the domain of attitude theories.

The attitude is measured with a number of adjectives (e.g., good–bad, pleasant–unpleasant, positive–negative) that are generally statistically combined to obtain a general judgment of the performance of a certain act.

Subjective norm is the perception of the individual that, in general, other people who are important to the individual want him or her to perform, or not to perform, that behavior. Who is important may vary across behaviors, but often this will include family, friends,

and neighbors. Other influences may stem from a religious institution to which a person feels attached, a political party, and the environmental movement. It is the individual's overall impression of how to behave, according to this collective of reference persons and groups, that is defined as the subjective norm.

The third concept is behavioral intention, that is, the deliberate plan to perform the behavior. According to the theory, the intention is the result of a decision process based on the attitude and the subjective norm, and it is the best predictor of behavior.

Finally, behavior is the performance of a certain act. In the TRA, behavior is usually confined to acts that are observable by others such as separating garbage and buying ecological food.

### **3.1.1. The Backgrounds of Attitude and Subjective Norm**

In the TRA, the concepts of attitude and subjective norm both are explained by a combination of two elementary concepts. The attitude toward a behavior is caused by beliefs about the likelihood of salient outcomes of performing a behavior, weighted by the evaluation of each of those outcomes. Salient outcomes are the outcomes that immediately come to the mind of the individual when thinking of the behavior. This combination of belief strength and evaluation of each outcome was formalized by Fishbein and Ajzen in the following formula:  $A_{act} = \sum B \times E$ . The formula states that the attitude toward a certain behavior ( $A_{act}$ ) is the sum of the products of beliefs (subjective estimate of likelihood of occurrence) and evaluations (valence) of all the behavioral outcomes that are considered by the individual.

Subjective norm, like the attitude concept, is composed of two subconcepts that explain which persons and groups are responsible for the normative pressure to perform, or not to perform, a certain behavior. The first concept is that of normative belief. A normative belief indicates the extent to which a person thinks that a specific referent person or group wants him or her to perform a behavior. Usually, with each behavior, the individual holds a number of normative beliefs. Referent persons will be family members (e.g., spouse, children, parents) who are important for many behaviors, whereas the influence of other persons or groups may vary more strongly with the behavior. For some behaviors, such as recycling of organic waste, the opinion of an environmental organization may be quite influential, whereas the opinion attributed to this

group will be irrelevant when deciding whether to make a visit to one's parents' house over the weekend. Each normative belief ( $NB$ ) is weighted for the so-called motivation to comply ( $MC$ ), that is, the degree to which an individual allows this referent person or group to exert influence on him or her. Subjective norm ( $SN$ ) is the sum of the products of all normative beliefs and the corresponding motivation to comply. Expressed as a formula,  $SN = \sum NB \times MC$ .

### **3.1.2. Correspondence: A Major Condition for Application of the TRA**

Correspondence entails that all of the components of the model are measured at the same level of specificity. Correspondence must exist on four criteria: action, target, context, and time. Action refers to the behavior itself. For example, if the behavior that is being investigated is recycling, all concepts of the theory must be specified as recycling and not as recycling in the measure of behavioral intention and as pro-environmental behavior in the measurement of the attitude. Similarly, one should specify the target (also called object sometimes) consistently. If one is interested in predicting the recycling of organic waste and not of used paper, aluminum cans, or glass, one should specify each component of the model as pertaining to the recycling of organic waste. And again, concerning context, one should decide whether one wants to investigate people's recycling behavior of organic waste in their homes, in their workplaces, or perhaps in all of the places where they spend time. Finally, concerning time, one must decide whether one wants to investigate people's recycling of organic waste next week (e.g., because there will be a special organic waste recycling drive organized by an environmental organization), the coming year, or forever from now on. Note that it is a matter of choice by the investigator at which level of specificity all of the components of the model are measured. Note also that the often used concept of environmental concern can be considered to be a very general attitude that is not correspondent with specific behavioral measures. Therefore, relationships between environmental concern and specific behaviors will necessarily be weak.

## **3.2. The Theory of Planned Behavior**

The TRA assumes that the behavior under investigation is under volitional control, that is, that people believe that they can execute the behavior whenever they are

willing to do so. Gradually, the TRA was used more often for the study of behaviors for which control was a variable factor. For that purpose, the TRA was complemented by a component that was named perceived behavioral control. This concept represents the extent to which people believe they are able to perform the behavior because they have adequate capabilities and/or opportunities or are lacking in these. It is very easy to see that this factor can substantially improve the generality of application of the model because there are many behaviors that need specific skills or external facilities. For example, recycling is virtually impossible if no collection system is available, and abandoning private cars is often impractical, at the least, when public transportation functions poorly.

The successor of the TRA that incorporates volitional control is the theory of planned behavior (TPB). Similar to attitude and subjective norm, the joint effects of two subconcepts determine behavioral control. The first subconcept consists of so-called control beliefs, that is, the estimated likelihood that each of a number of specific factors will facilitate or impede execution of the behavior. An example might be the following control belief: "I can go to my office by bus given the distance from the bus stop to my home." The second factor is called perceived power, that is, a judgment of the degree of facilitation or impediment that each specific control belief represents. An example might be the following: "The distance from the bus stop to my home makes it very easy/very difficult to go to my office by bus." The formula combining the two subconcepts of likelihood of control and perceived power is identical to that of the attitude and subjective norm concepts:  $PBC = \Sigma C \times P$ .

Both the TRA and the TPB have been used repeatedly for investigating specific environmental behaviors such as changing travel mode, water conservation, recycling, and green consumerism. In general, the models have proven to be useful in understanding the behavior, with important contributions of perceived behavioral control.

### **3.3. Norm Activation Theory**

The attitude concept, as used in the TPB, is meant to be a general evaluation of a behavior. This implies that all of the specific evaluative criteria that guide a decision to adopt, or not to adopt, that behavior should determine the attitude score. This does not seem to be the case given that research has shown repeatedly that

moral considerations are not effectively covered by the attitude concept. Environmental behavior, because of its social dilemma character, clearly has a moral dimension in that acting in favor of the environment/community is generally perceived as morally superior to acting out of self-interest. The development of moral considerations and their relation to behavior is described in Schwartz's norm activation theory (NAT). The central concept is personal norm, defined as intrinsically motivated self-expectations regarding morally appropriate behaviors. Norm activation generates a feeling of personal obligation to perform a specific behavior. According to the NAT, norm activation occurs under the influence of four situational activators and two personality trait activators. Norm activation starts with the awareness of need. It involves the extent to which a person's attention is focused on the existence of a person or a more abstract entity (e.g., the environment) in need. Apart from being aware of need, the potential actor must define the situation at hand as one in which he or she feels some responsibility for the consequences of the needy party's welfare. Norm activation is further enhanced by efficacy, that is, the extent to which a person recognizes actions that might alleviate the need. The actions that are recognized as such determine which specific norms are activated. Finally, ability refers to the actor's perception about whether he or she possesses the resources or capabilities needed to perform the focal behavior.

In addition, the NAT proposes that two personality traits influence activation of personal norm: awareness of consequences and denial of responsibility. Awareness of consequences refers to a person's receptivity to situational cues of need. Denial of responsibility refers to a person's inclination to deny responsibility for the consequences of his or her behavioral choices for the welfare of others. The four situational activators and the two personality traits determine whether or not a behaviorally specific personal norm becomes activated. Specific behaviors, such as burning garden waste, recycling, buying lead-free gasoline, using off-road vehicles in the environment, energy conservation, and littering, have been studied by means of the NAT. In general, just parts of the whole model have been used. Most studies show that personal norms significantly influence pro-environmental behavior, whereas there is less agreement about the predictors of personal norms and the exact process through which personal norms become activated.

### 3.4. Habits

All of the approaches to understanding environmental behavior described in this section assume that people process relevant information to decide to perform an environmentally relevant behavior in a pro- or anti-environmental manner. However, people often rely on behavioral routines or habits. Habits are especially likely to develop when behaviors are displayed frequently and in a stable context. Habitual behavior is defined as behavior that is displayed automatically on the presence of a goal, that is, a direct goal-action link that is not preceded by consciously developed intentions. Automatically executed behavior has three main characteristics in that (a) it can be executed fast, (b) its execution cannot be controlled, and (c) it is very efficient (i.e., it allows other cognitive processes to be executed simultaneously). Care should be taken not to equate habitually performed behavior with all behaviors that are performed in the past, even when the conditions of high frequency and context stability are met. Behavior executed in the past can influence future behavior through its effect on intentions, and in that case it is unlikely that the behavior is a habit. However, some studies have shown that the strength of intentions to influence behavior depends on the degree to which behavior has been performed in the past. When behavior has been frequently performed in the past, the effect of intentions on behavior is smaller. In such cases, where past behavior changes the intention-behavior relationship, it is very likely that past behavior indicates that habits play a role.

Clearly, the automatic execution of behavior has important advantages. But it has drawbacks as well in that people who have established habits pay less attention to information that might be important for changing behavior. For example, the information that is issued by a railway company about improved services is likely to be ignored by people with strong car travel habits, even though it might be very advantageous for them. Given that people with strong habits cannot easily be influenced by information, an important question then becomes how to change habitual anti-environmental behavior.

## 4. HOW TO EFFECT PRO-ENVIRONMENTAL BEHAVIOR CHANGE: THE ROLE OF PSYCHOLOGY

Applied psychology can play a role in establishing more environmentally sound living patterns. This

section gives a summary of the potential contribution of psychology for the development of intervention methods. The methods range from obligatory to voluntary systems, but all need the application of psychological knowledge to understand and optimize effects.

### 4.1. Laws, Rules, and Regulations

It seems to be a straightforward solution to communal problems to implement rules, regulations, or laws that prohibit the actions that cause the problem. However, from a psychological perspective, there are several issues that need mentioning. First, regulations will fail when the targeted population is not convinced of the necessity of the policy measure. In that case, strict enforcement will be necessary and costly. If strict enforcement is not applied, the targeted population will not adhere to the regulation. In both cases, the measure will not have the favorable effects that are desired. Second, a policy measure will be accepted according to its perceived fairness. Fairness means that people will realize that the policy measure targets the right actions and punishes offenders according to the severity of their offenses. Third, from these characteristics and based on laboratory research on social dilemmas, it can be predicted that people will be most willing to endorse policy measures when they are in favor of the proposed measure but are very uncertain what other people's attitudes and behavior will be. These psychological effects of perceptions, attitudes, and uncertainties and their relationships were formalized in the structural goal expectation theory developed by Yamagishi in 1986. Fourth, implementation of policy measures, even if they are successful, is not an easy way out of the social dilemma. To develop legislation, political support for measures protecting the environment is necessary. To create political support, environmental awareness and concern among voters, and willingness to sacrifice personal advantages for long-term collective benefits are necessary.

### 4.2. Changes in the Physical Environment

Measures that are generally less dependent on political approval are changes in the physical environment that facilitate pro-environmental behavior or impede anti-environmental behavior. These kinds of measures can work in three different ways.

**4.2.1. Direct Influence on Behavior**

Changing the layout of a road (e.g., by creating additional curves or narrowing lanes) will reduce automobile speed and emissions directly.

**4.2.2. Influence on Habits That Control Behavior**

Many environmental behaviors are habitually performed. Changing the physical context of the behavior means that these habits no longer apply and information must be sought about how to achieve the behavior's goal in these changed circumstances. This renewed need for information creates possibilities for pro-environmental behavior change, for example, by showing that the daily commute to work can now be completed faster by public transportation than was the case during the period when the car commute habit was developed.

**4.2.3. Influence on the Convenience of Performing Pro-environmental Behavior**

In many cases, pro-environmental behavior is impeded not by negative attitudes but rather by weak positive attitudes combined with a lack of perceived behavioral control. For example, most people are in favor of recycling garbage but are discouraged by the time and effort it takes to bring their recyclable garbage to a recycling station. Creating drop-off facilities for recycling paper, bottles, and other recyclables at close range from home may strongly increase the volume of recyclable materials.

**4.3. Monetary Incentives**

It might seem as though the most straightforward way in which to solve environmental problems is to let polluters pay for the damage they do to the environment. In general, this may be an effective strategy that eliminates the social dilemma structure of environmental problems. Research has shown that people are indeed sensitive to changes in the payoffs and that the number of egoistic choices (as compared with choices for the collectivity) is reduced when egoistic choices are less profitable. However, the following subsections review a number of complicating factors pertaining to the psychology of justice, the perception of monetary changes, and the dependence of monetary incentives on information to accomplish the desired effect.

Moreover, financial measures run the risk of weakening intrinsic motivation, that is, the motivation to act in a pro-environmental manner out of deeply felt concern for the state of the environment.

**4.3.1. Policy Measures**

Financial measures, in particular, can be evaluated differently depending on the fairness principle that is considered appropriate. Three principles are applied: those of equity, equality, and need. Equity implies that the damage is compensated by the actor in proportion to the damage that he or she causes. Equality implies that people contribute to the same extent, for example, that they pay the same amount for having their houses connected to the city's sewer system regardless of the amount of polluted water they produce. Need implies that people with more financial resources assume a larger share of the financial burden. It is not always clear which principle should apply in a specific situation. This political disagreement constitutes a reason for the absence or delay of appropriate measures.

**4.3.2. Price Perception**

Price perception is subject to many different influences that can substantially change the relation between actual costs and willingness to take pro-environmental action. A finding that is at the core of Tversky and Kahneman's prospect theory is that people are more sensitive to potential losses than to potential gains. For that reason, it is advised to frame financial consequences of anti-environmental behavior as a potential loss instead of framing financial consequences of pro-environmental behavior as a potential gain. The complexity of consequences of financial measures may cause the use of heuristics. For example, the calculations necessary to decide whether it is cost-efficient to install double-glazing are difficult because it is necessary to simultaneously consider changing consumption, changing energy prices, costs of a loan, and/or lack of profit from a different investment. Consumers, and even entrepreneurs, are rarely able to take all of the relevant factors into account and may simply evaluate a decision on the basis of nominal costs before and after installment.

**4.3.3. Financial Measures**

Measures such as subsidies, loans, discounts, and other price changes do not automatically lead to changes in

behavior. Being aware of those price changes is crucial, and this implies the importance of informational interventions that precede or accompany the introduction of financial measures. Several experiments have demonstrated that attractive financial propositions are successfully implemented only when sufficient attention is given to making the programs known. To increase participation, it is often more prudent to spend money on information campaigns than to further increase the financial attractiveness of propositions.

#### 4.3.4. Target Recurring Behavior

When preparing financial measures that target recurring behavior, one should bear in mind that the motivation to perform pro-environmental behavior is changed from intrinsic (out of concern for the environment) to extrinsic (for the monetary advantage). For one-time decisions, this might not be an issue. However, for recurring behavior, a relapse to anti-environmental behavior might occur once the financial gain from acting in a pro-environmental manner is removed. Pro-environmental behavioral maintenance then becomes problematic. Another unwanted effect that may stem from financially favorable energy-saving devices is the rebound effect. The knowledge that energy, or water, is so sparsely used due to new efficient devices might invite people to be less careful in their use of these devices. Having lights on throughout the night and showering longer might be invited by the installment of energy-saving lamps and water-saving showerheads, respectively. This phenomenon helps to explain why reductions in energy use through new efficient devices are generally not as favorable as foreseen in the calculations of technicians.

### 4.4. Information Strategies

As has now become clear, information is indispensable for implementing other interventions. It is helpful in increasing acceptance for policy measures, in announcing that physical changes in the environment are forthcoming, in announcing that financial support can be received for the implementation of energy-saving measures, and so on. Apart from supporting other measures, information can be used on its own for a large number of purposes, all ultimately intended to change behavior in a pro-environmental direction. Prominent functions of information include the following:

1. Information may create insight into the structure of the dilemma of environmental behavior. Laboratory research on social dilemmas has pointed out that people are more willing to cooperate when they are made aware of the dilemma structure of environmental problems.

2. Information may increase problem awareness and emotional concern. Both are important, but they are important to different degrees for different target groups. Groups of people who are motivated and capable of processing information thoroughly are better served by more factual information. Knowledge about the environmental problem—its nature, causes, and consequences—is an important condition to induce pro-environmental behavior, according to Schwarz's norm activation theory. For people who are relatively lacking in ability and motivation to process information, messages that stress the affective consequences of environmental degradation may be more influential.

3. Information can teach people relevant skills that facilitate the performance of pro-environmental behavior. People should learn how to dispose of chemical waste and how to compost organic waste before such pro-environmental behavior is applied. Many studies point out that the relative ease of performance is a crucial factor in pro-environmental behavior.

4. Information may tackle anti-environmental habits by reminding people of their pro-environmental attitudes. Using prompts to remind people not to drive fast, to lower thermostat settings well before going to bed, and so on can be successful in that these prompts encourage behaviors with which people agree.

5. Information may be used to address possible personal benefits of pro-environmental behavior. Sometimes, wrong estimates are made regarding the costs and benefits of pro-environmental behavior. This may occur, for example, when someone estimates the financial consequences of home insulation. Wrong estimates can be caused by outdated information (among other reasons), and this is especially likely in the case of habitual behavior when less attention is given to new information. This neglect of information may, for example, impede a choice for public transport even though travel conditions have improved greatly compared with the past when the choice to drive private automobiles was made.

6. Communication can reduce social uncertainty. Knowing that others are also willing to incur the costs of pro-environmental behavior convinces people that their efforts will not be in vain. This confidence in the efficacy of shared efforts will increase people's

willingness to behave in a pro-environmental manner. It should be noted, of course, that this mechanism works only when others (are willing to) behave pro-environmentally. If this is not the case, communication strategies should focus on other determinants such as knowledge, concern, and ability.

## 5. CONSIDERATIONS REGARDING THE EFFECTIVENESS OF INTERVENTIONS FOR PRO-ENVIRONMENTAL BEHAVIOR CHANGE

Changing behavior is sometimes a difficult task, and this certainly applies to changing environmental behavior in a pro-environmental direction. The social dilemma nature of environmental problems makes personal benefits of change, and thus of pro-environmental change, less likely. Nevertheless, people are willing to change to some extent. The categories of instruments described in the previous section may be effective in inducing change. But what does effectiveness actually mean? Contrary to what one would think intuitively, effectiveness is not a unitary concept. In what follows, this section adheres to the evaluation system proposed by De Young in 1993. According to De Young, effectiveness must be broken down into five dimensions:

- *Reliability*. Is an intervention able to effect the desired change in the target group? Moreover, is it able to do this every time? This criterion comes closest to what one would ask of an intervention program. However, an answer may be subtler than the questions imply. For example, the technique may be able to change individual behavior to some extent and may be more successful with some individuals than with others. The question about repeated use implies that the technique is not expected to create a lasting change that is exhibited each time the behavior can be displayed.

- *Speed of change*. How fast are changes in behavior accomplished? Of course, the sooner, the more impact (aggregated over time). However, some interventions can be expected to have their effect only after some time. The installation of expensive energy-saving equipment, such as a new central heating furnace, will usually be done when the old equipment is worn out. For research on effectiveness of intervention programs, it constitutes a challenge to be able to discover

changes initiated by the program but effected over a longer term. However, this would be desirable to know so as to be able to give a good estimate of the value of an intervention program.

- *Particularism*. This is the degree to which an intervention must be tailored to the characteristics of a target group. This ranges from an intervention that is universally applicable to an intervention that is uniquely designed for an individual. Money is an example of a motivator that is considered to be generally applicable because it retains its value regardless of the person intervening, whereas personal attention seems to be at the other extreme because it clearly matters who is giving personal attention. An intervention that is universally applicable and effective is, of course, a great tool because it can be administered much more efficiently. However, in general, a degree of tailoring to characteristics of the target group is necessary to increase the effect of an intervention. Favorable in that respect is a recent development that allows messages to be tailored to the target person or group in an automated process by which personalized information is sent to individuals whose characteristics have been inventoried earlier by means of a questionnaire and stored in an electronic database. Based on this information, a computer program automatically compiles an informative message that is closely tuned to the receiver. This technique combines the reach of mass media with the advantages of personalized advice; therefore, it seems promising.

- *Generality*. To what extent does an intervention technique influence behaviors that are not directly targeted but are conceptually related? And how closely related should they be to be influenced by the intervention? In general, it has proven to be difficult to create a degree of transfer from one behavior to the next. This is considered one of the two main challenges for an environmental psychology that aims to contribute to substantial changes in environmental behavior.

- *Durability*. To what degree are effects of intervention techniques lasting? Achieving durability of pro-environmental change constitutes the second main challenge. Especially with intervention techniques that focus on voluntary participation, it has been difficult to create substantial effects that last beyond the period when the intervention is executed. Examples of long-lasting changes are scarce. However, it seems that some techniques are more promising in this respect than are others. Especially explicit commitment (e.g., by a pledge done in public to

adhere to the goals of the intervention) seems to offer possibilities of behavioral maintenance after the intervention has expired.

## 6. CONCLUSIONS

Environmental problems are essentially problems of human behavior. Psychology can help one to understand behavior and propose ways of changing behavior. These are the main tenets of environmental psychology. This article has given a brief overview of influential theoretical perspectives and of the way in which psychology can help to create or improve pro-environmental behavior. It seems that a considerable knowledge base has been built over the past 30 years but that much more scientific attention, creativity, and persistence are needed, especially in developing interventions that change behavior. The rate of pro-environmental changes, and especially the durability of changes, is problematic. This is very understandable when considered from the perspective of social dilemmas. Given the existing structures in society, it is difficult to forgo individual interests for environmental improvements that are uncertain, may take a long time to come about, and may even be mainly beneficial to others. Possibly, psychology must dedicate more attention to the key actors in society—politicians, managers, and people who lead and represent large organizations. To this point, most attention has been on understanding and influencing people as consumers and citizens. For the most part, interventions have aimed at influencing people in their homes and sometimes in the workplace. In general, no attempts have been made to influence decisions made at city halls, in the boardrooms of multinational companies, or in the courts. Nevertheless, the bottom-up strategy of influencing people as consumers and citizens may profitably be complemented by attempts to understand people in specific, rare, and influential roles. This might help to bring about structural changes in society that would make it easier for everyone to behave in accordance with the needs and demands of the environment.

## See Also the Following Articles

- Attitude Measurement ■ Attitudes ■ Conservation Behavior
- Environmental Design and Planning, Public Participation in
- Restorative Environments

## Further Reading

- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, 52, 28–58.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, 31, 169–193.
- De Young, R. (1993). Changing behavior and making it stick. The conceptualization and management of conservation behavior. *Environment and Behavior*, 25, 485–505.
- Geller, E. S. (2002). The challenge of increasing proenvironment behavior. In R. B. Bechtel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 525–540). New York: John Wiley.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162, 1243–1248.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47, 262–291.
- Manstead, A. S. R., & Parker, D. (1995). Evaluating and extending the theory of planned behavior. In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 6, pp. 69–96). Chichester, UK: Wiley.
- Schwartz, S. H. (1977). Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 221–279). New York: Academic Press.
- Staats, H. (2003). Understanding pro-environmental attitudes and behavior: An analysis and review of research based on the theory of planned behavior. In M. Bonnes, T. Lee, & M. Bonaiuto (Eds.), *Psychological theories for environmental issues* (pp. 171–201). Aldershot, UK: Ashgate.
- Stern, P. C. (2000). Psychology and the science of human–environment interactions. *American Psychologist*, 55, 523–530.
- Stern, P. C., & Oskamp, S. (1987). Managing scarce environmental resources. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (Vol. 2, pp. 1043–1088). New York: John Wiley.
- Verplanken, B., & Aarts, H. (1999). Habit, attitude, and planned behaviour: Is habit an empty construct or an interesting case of goal-directed automaticity? In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 10, pp. 101–134). Chichester, UK: Wiley.



# Prosocial Behavior, Development of

Nancy Eisenberg and Adrienne Sadovsky

Arizona State University, Tempe, Arizona, USA

1. Introduction
  2. The Development of Prosocial Behavior
  3. Gender Differences
  4. Cultural Differences
  5. Contextual Factors
  6. Dispositional Factors
  7. The Origins of Prosocial Behavior
- Further Reading

prosocial behavior, and prosocial actions are affected by contextual factors as well as by individual differences in temperament or personality. Prosocial tendencies are likely due to genetic factors as well as to socialization both within the family and in the child's larger social world.

## GLOSSARY

**altruism** Prosocial behavior that is motivated by sympathy or moral values/concerns rather than by egoistic factors (e.g., concrete rewards, social approval, elimination of vicarious distress to make oneself feel better).

**empathy** An affective response based on another person's emotion or condition that is similar to what the other person is experiencing or would be expected to experience.

**meta-analysis** A procedure that statistically combines the findings from a number of different studies to determine whether a general finding holds across various empirical studies.

**prosocial behavior** Voluntary behavior that is intended to benefit another person.

**sympathy** Other-oriented concern that is often based on empathy.

Prosocial behavior starts to develop early in life and plays an important role in moral development. There are cultural and gender differences in some types of

## I. INTRODUCTION

Prosocial behavior, or voluntary behavior intended to benefit another person, is of obvious importance for enhancing human relationships and for the smooth functioning of society. Of particular importance is altruism, which sometimes is defined as prosocial behavior that is motivated by sympathy or moral values/concerns rather than by egoistic factors (e.g., concrete rewards, social approval, elimination of vicarious distress to make oneself feel better). Unfortunately, it is often difficult to differentiate the two; indeed, some psychologists (and philosophers) claim that real altruism does not exist. Thus, most of the existing empirical work involves the study of prosocial behaviors that may or may not be altruistically motivated.

Some theorists, such as Batson, believe that altruism usually stems from an individual experiencing another's emotional state (i.e., empathy) or other-oriented concern (i.e., sympathy). Other theorists, such as Eisenberg, believe that prosocial behaviors (especially altruistic ones) can be engendered by internalized moral values

as well as by sympathy. Still others, such as Cialdini, believe that prosocial behavior is motivated by hedonism; for example, people help to improve their own moods. Empirical studies have demonstrated that there are associations between prosocial behaviors and both sympathy and measures of individuals' moral reasoning and values. Empathy and sympathy are also believed to inhibit cruelty toward others, although it is not clear whether prosocial behavior and aggression are highly negatively related (although both likely are affected by empathy and sympathy).

## **2. THE DEVELOPMENT OF PROSOCIAL BEHAVIOR**

Even infants can experience concern for others and have the capacity to perform prosocial behaviors. By 10 to 14 months of age, infants frequently express agitation when viewing others in distress. Self-distressed reactions when viewing others in distress decrease in frequency during the second year of life, and by 18 months of age infants sometimes make efforts to comfort (e.g., by patting a person) or help other people who are upset or sad, especially their parents. In addition, even toddlers frequently try to assist adults with household chores.

According to Hoffman, toddlers do not fully distinguish between their own inner states and other people's inner states; consequently, their attempts to help are often egocentric (e.g., a toddler may bring a distressed peer to his or her own mother rather than the peer's mother). With age, children's prosocial actions become more situationally appropriate and sensitive to others' specific needs.

Meta-analytic reviews have shown that the frequency of children's prosocial behavior, including sharing, comforting, and instrumental helping, tends to increase with age across childhood, although there is little difference in the frequencies of prosocial behavior between younger and older adolescents. Age differences tend to be greater in studies with experimental/structured designs than in those with naturalistic or correlational designs, and they also tend to be greater when prosocial behavior is measured with self- or other-reports than when it is measured with observations. Moreover, although the consistency of prosocial behavior across contexts is low during the early years of life, it becomes more consistent with age and there are relatively clear individual differences in prosocial tendencies by late childhood and adolescence. However, there is some

consistency of prosocial tendencies over time, even from the preschool years. For example, preschool children who spontaneously engage in sharing behaviors that involve some cost (e.g., giving up a toy) are more sympathetic and prosocial as adolescents and adults.

Younger and older children also differ in the reasons they report for their prosocial actions. Older children are more likely to say that they assist for apparently altruistic or other-oriented motives, whereas younger children are more likely to assist to obtain material or social rewards. Furthermore, with development, children are more likely to resolve hypothetical moral dilemmas regarding assisting others at a cost to themselves by referring to higher level moral justifications. For example, preschoolers tend to verbalize primarily hedonistic reasoning and some needs-oriented (primitive empathic) reasoning. By school age, some children begin to express concern about gaining approval, enhancing interpersonal relationships, and behaving in stereotypically "good" ways. During late elementary school or thereafter, children start to verbalize reasoning reflecting abstract principles, internalized affective reactions (e.g., guilt or positive affect about the consequences of their behavior for others, living up to internalized principles), self-reflective sympathy, and perspective taking. From childhood into early adolescence, children's needs-oriented reasoning increases, whereas hedonistic reasoning decreases. For most children, reasoning concerning role taking, internalized norms/rules/values, generalized reciprocity across members of society, internalized affective reactions based on concern about the consequences of their behavior for others, and positive affect related to internalized values and living up to those values is uncommon during elementary school and increases with age from adolescence into early adulthood. Nonetheless, people of all ages sometimes express low-level, hedonistic moral reasoning. Age-related changes in children's prosocial moral reasoning are of interest because higher level reasoning likely reflects developmental changes in the goals and motives underlying children's prosocial behavior and has been correlated with the tendency to engage in prosocial behavior.

## **3. GENDER DIFFERENCES**

Gender stereotypes regarding sex differences in prosocial behavior are common in American culture and probably in numerous other cultures. In general, girls and women are expected to be more caring and to

engage in more prosocial behaviors than are boys and men. Consistent with the stereotype, in meta-analytic analyses, effect sizes indicate that girls reliably display greater prosocial tendencies than do boys. For example, in 1998, Eisenberg and Fabes found that girls evidenced considerably greater kindness and consideration for others than did boys ( $d = .42$ ). Likewise, girls shared and donated more than did boys, although not to such a marked extent ( $d = .13$ ), and also provided somewhat more instrumental help ( $d = .14$ ). In additional meta-analyses, investigators found that girls exhibited greater empathy and sympathy than did boys, at least on some types of measures (e.g., self-reports, some observed behavioral measures). In 1999, Fabes and colleagues noted that these sex differences in prosocial tendencies become stronger during adolescence as compared with early and middle childhood. Thus, there is support for a developmental change in the expression of prosocial behaviors.

In contrast, examination of sex differences in adults' prosocial behaviors has provided mixed results. In 1986, Eagly and Crowley found that men, but not women, generally exhibited considerably greater helping ( $d = .34$ ). However, many of the studies with adults involved helping strangers and performing instrumental tasks (e.g., changing a tire). These authors suggested that men's greater helping likely reflects the conventional gender role-rooted expectation for men to enact agentic behaviors or to "rescue" others (i.e., behave chivalrously) and that the findings likely differed for prosocial behaviors involving both caring and recipients who are known.

Consistent with Eagly and Crowley's speculations, women have been found to display greater prosocial tendencies than have men with regard to other indexes of prosociality. For example, women deliver messages containing greater emotional support and report more sympathy and empathy than do men, both on questionnaires and after being exposed to empathy-inducing stimuli, although they generally do not differ on physiological or facial measures of empathy. Furthermore, adolescent girls and adult women generally score higher on caring and other-oriented moral reasoning.

It is important to emphasize that the strength of sex differences fluctuated greatly across studies included in the aforementioned meta-analyses. Myriad factors, such as the design of the study, account for some degree of the difference between women and men. Nonetheless, the data do support the legitimacy of the stereotypes regarding sex differences in prosocial

behaviors, at least when taking into account the type of helping behavior. However, the finding that sex differences in empathy and sympathy are stronger for self-reports than for other methods of assessment is consistent with the notion that people likely perceive a greater gender difference in prosocial tendencies than actually exists and respond accordingly.

#### 4. CULTURAL DIFFERENCES

Prosocial tendencies appear to differ considerably across cultures. For example, in some traditional cultures such as the Ik in Africa, aggression and cruelty are commonplace, whereas in some other cultures, caring, helping, and supportive behaviors are fairly commonplace. Anthropologists have found that children in traditional cultures tend to assist family members and neighbors more than do children in industrial communities. Even in industrialized societies, children from collectivist cultures that emphasize the well-being of the larger social group (e.g., China) appear to be more prosocial with peers than do children from individualist cultures (e.g., the United States). However, it is not known whether these cross-cultural differences hold when unfamiliar people, rather than family members, neighbors, and friends, are the targets of aid.

The reasons for observed cross-cultural and subcultural differences in prosocial and cooperative behavior are probably multiple and complex. Clearly, the value placed on the well-being of others and on prosocial behavior varies across societies and likely is taught to children by parents and teachers as well as communicated through social norms. Some evidence suggests that these expectations are communicated in everyday life. For example, in cultures where children are expected to be cooperative and help others, often as part of their chores (e.g., when caring for siblings), children tend to be relatively prosocial. Moreover, prosocial actions tend to be viewed as more obligatory in collectivist societies than in individualist societies. Also in collectivist societies, privileges and social acknowledgment in the classroom are dependent on group accomplishments rather than individual accomplishments, so that children are reinforced for helping one another.

#### 5. CONTEXTUAL FACTORS

Children's and adults' prosocial behaviors vary considerably as a function of the specific context. For

example, people are more likely to help if they see someone else—even a stranger—modeling prosocial behavior. Moreover, there is evidence that children and adults help more when they are in a good mood and help less when they are sad or depressed (especially in the case of older children and adults). Furthermore, factors such as the clarity of cues related to the other person's need, the cost and benefits associated with a particular prosocial action, and the salience of social norms regarding prosocial behavior appear to influence whether people choose to help or share with others.

## 6. DISPOSITIONAL FACTORS

Some people are more prosocial than others. Individual differences in prosocial behavior have been linked to a variety of aspects of personality or other dispositional characteristics. For example, as already noted, people who are prosocial are relatively prone to experience empathy and sympathy for others. In addition, prosocial individuals tend to feel a sense of responsibility for their actions and for others and also feel competent to assist. During childhood, prosocial tendencies also have been linked to being well regulated, socially competent, emotionally positive, and sociable, although nonassertive and less sociable children may be especially likely to perform compliant prosocial behaviors that are requested by their peers (in part because they have difficulty in asserting their own will). Thus, even within cultures and across the sexes, characteristics of individuals seem to contribute to their willingness and ability to assist others.

## 7. THE ORIGINS OF PROSOCIAL BEHAVIOR

Findings in twin studies support the common assumption that there is a genetic component to prosocial behavior and the tendency to empathize with others. Nonetheless, individual differences in children's socialization experiences in the home, school, and community are linked to individual differences in their prosocial behavior.

Prosocial children (and adults) tend to be those with parents who model and value prosocial actions and who apply ethical principles (especially those related to caring and helping) to a broad spectrum of

humanity. Parents of prosocial children also tend to use reasoning (inductions) when they discipline their children (e.g., pointing out the consequences of children's behavior for others) and seldom resort to physical punishment or threats of punishment. In addition, these parents tend to be authoritative and democratic in their parenting style (i.e., are supportive and use inductions while also clearly communicating their expectations for appropriate behavior) and provide opportunities for their children to assist others. In general, parents of prosocial children also appear to promote the development of children's sympathy, and sympathy appears to motivate altruistic caring behavior.

Similarly, prosocial and cooperative behavior also can be fostered in school-based programs that emphasize rational discipline, cooperation, and prosocial values and that include activities designed to heighten children's sympathy and understanding of others and their feelings. In addition, a positive relationship between children and their teachers is associated with children's consideration for others and children's empathy with unfamiliar peers. Finally, there is a modest positive relation between children's viewing of prosocial television programming and their prosocial behavior. Thus, it appears that individual differences in prosocial behavior are due to a variety of influences in the home, media, and community.

### See Also the Following Articles

Emotion ■ Gender and Culture ■ Stereotypes

### Further Reading

- Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Lawrence Erlbaum.
- Davis, M. H. (1994). *Empathy: A social psychological approach*. Madison, WI: Brown & Benchmark.
- Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, 100, 283–308.
- Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In N. Eisenberg (Ed.), *Handbook of child psychology*, Vol. 3: *Social, emotional, and personality development* (5th ed., pp. 701–778). New York: John Wiley.
- Eisenberg, N., & Mussen, P. (1989). *The roots of prosocial behavior in children*. Cambridge, UK: Cambridge University Press.
- Fabes, R. A., Carlo, G., Kupanoff, K., & Laible, D. (1999). Early adolescence and prosocial/moral behavior: I. The role of individual processes. *Journal of Early Adolescence*, 19, 5–16.

- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge, UK: Cambridge University Press.
- Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin*, 126, 703–726.
- Oliner, S. P., & Oliner, P. M. (1988). *The altruistic personality: Rescuers of Jews in Nazi Europe*. New York: Free Press.
- Schroeder, D. A., Penner, L. A., Dovidio, J. F., & Piliavin, J. A. (1995). *The psychology of helping and altruism: Problems and puzzles*. New York: McGraw-Hill.
- Whiting, B. B., & Whiting, J. W. M. (1975). *Children of six cultures: A psychocultural analysis*. Cambridge, MA: Harvard University Press.
- Zahn-Waxler, C., Robinson, J., & Emde, R. N. (1992). The development of empathy in twins. *Developmental Psychology*, 28, 1038–1047.



# Psychological Assessment, Standards and Guidelines for

Dave Bartram

SHL Group, Surrey, United Kingdom

1. Introduction
  2. International Standards and Guidelines
  3. National Standards and Guidelines
  4. Conclusion
- Further Reading

Standards and guidelines in assessment cover three related areas: the people who use tests, the tests the people use, and the process of assessment within which tests are used. Both nationally and internationally, all three of these issues have been addressed in various ways.

## GLOSSARY

**psychological assessment** Any assessment of an individual's psychological characteristics carried out using procedures that are standardized and repeatable; this can include structured interviews and standardized work sample assessment exercises as well as more traditional psychological tests.

**psychological testing** Standardized procedures whereby quantitative measures of psychological attributes or characteristics are obtained.

**psychological tests** Instruments with known psychometric properties designed to measure one or more psychological characteristics; the psychometric properties will include measures of reliability and validity as well as data to aid interpretation.

**standards and guidelines** Two kinds of criteria that differ in that standards are generally more prescriptive than guidelines, whereas guidelines tend to be advisory and informative.

**test user** Anyone who is responsible for the use of a psychological test in an assessment process.

## 1. INTRODUCTION

Standards and guidelines have been developed to cover all aspects of test design, development, and use. In many countries, test publishers or distributors were the first to develop qualification criteria, guidelines, and standards to regulate, advise, and control test users. National psychological associations and international bodies concerned with testing and test use (e.g., European Federation of Psychologists Associations [EFPA], International Test Commission [ITC]) have also played a major role in developing and promulgating standards and guidelines. These standards and guidelines have addressed issues associated with test development, the qualification of test users and test distribution, proper test use and best practice, and test taker rights and responsibilities.

Standards and guidelines cover all sorts of psychological testing and assessment methods—from tests of ability and aptitude to tests of personality and motivation, from paper-and-pencil tests to computer-based

and Internet-delivered tests, from single instruments to complex 2- or 3-day “assessment centers.”

A survey of 48 countries, carried out jointly in 1996 by the ITC and EFPA, suggested the following:

- Approximately 79% of test users operate in the educational field, 12% in the clinical field, 9% in work and organizational settings, and less than 1% in the forensic area.
- The majority (86%) of test users are nonpsychologists.
- The clinical and forensic areas are the only ones where nonpsychologists do not outnumber psychologists as test users.
- Only 41% of all users have received any specific training in test use, with the lowest rates of training in the educational area and the highest rates in the clinical area.
- Psychologists have only slightly higher rates of training than do nonpsychologists.

Overall, the survey showed the need for more training for all test users and revealed marked differences in patterns of response among different countries. Of particular interest were the big differences among countries in the power and role of the professional psychological associations in relation to setting and enforcing standards.

## **2. INTERNATIONAL STANDARDS AND GUIDELINES**

In the relatively recent past, it was possible to think of individual countries as “closed systems.” Changes could be made in terms of best practice, procedures, and laws affecting the use of tests in one country without there being any real impact on practice or laws in other countries. People tended to confine their practice of assessment to one country, and test suppliers tended to operate within local markets by adapting prices and supply conditions to the local professional and commercial environments.

This situation has changed. Individual countries are no longer closed systems. Psychological testing and assessment is an international business. Many test publishers are international organizations that sell their tests in a large number of countries. Many of the organizations using tests for selection and assessment at work are multinationals. In each country, test suppliers and test users are likely to find differences in practices related to access, user qualification, legal constraints on test use, and so on.

### **2.1. The International Test Commission**

The ITC is an association of national psychological associations, test commissions, publishers, and other organizations committed to promoting effective testing and assessment policies and to ensuring the proper development, evaluation, and use of educational and psychological instruments. The ITC has been responsible for a number of guideline development projects. In each case, the process has been one of ensuring widespread international agreement on the content of the guidelines.

#### **2.1.1. ITC Guidelines on Adapting Tests**

These provide 22 guidelines for adapting psychological and educational tests for use in various different linguistic and cultural contexts. These guidelines fall into four main categories: the cultural context, the technicalities of instrument development and adaptation, test administration, and documentation and interpretation. Adaptation is an area of major importance as tests become used in more and more countries and as tests developed in one country get translated or adapted for use in another country. Adaptation needs to consider the whole cultural context within which a test is to be used.

#### **2.1.2. ITC Guidelines on Test Use**

These guidelines were published in 2000 and have been widely adopted by countries around the world as the basis for their own local standards. They focus on good test use and cover both ethical issues and best practice process. They provide a common international framework from which specific standards, codes of practice, qualifications, user registration criteria, and so on can be developed to meet local needs. By 2004, these guidelines were available in 14 different languages.

The competencies defined by the guidelines are specified in terms of assessable performance criteria. These competencies cover the following:

- Professional and ethical standards in testing
- Choice and evaluation of alternative tests
- Test administration, scoring, and interpretation
- Report writing and feedback
- Rights of the test candidate and other parties involved in the testing process

### **2.1.3. ITC Guidelines for Computer-Based and Internet-Delivered Testing**

Work is in progress on the development of international guidelines in this area. These were scheduled for publication in 2004.

## **2.2. European Initiatives**

### **2.2.1. The EFPA Test Review Criteria**

The EFPA represents 30 European countries. Its Committee on Tests and Testing published the EFPA standards for the review of psychological tests in 2001. These standards, developed from British and Dutch work on test reviewing, are now being used as the basis for test reviewing across Europe and provide the basis for a test registration procedure.

Within Europe, new legislation on data privacy has had a major impact on testing and the use of tests. This has resulted in good practice in assessment becoming embodied in law and has led to the development of guidance to support this within European countries. Increasingly, it is being recognized that there is a need for standards and guidelines to be considered at a European level to match the demands arising from European Union directives, especially those that relate to fairness and equity issues in the employment area and to the transferability of professional qualifications among countries.

### **2.2.2. Other European Initiatives**

The European Association of Psychological Assessment (EAPA) has published a set of Guidelines for the Assessment Process. The European Association of Work and Organizational Psychologists (EAWOP) has established a working group to focus on Europe-wide issues relating to testing and assessment in the area of work and employment. Work has started on the development of European standards for the certification of competence in the use of tests for assessment in the work and organizational field. This work is being carried out by a joint EFPA–EAWOP task force and could lead to a European qualification of competence in testing.

### **2.2.3. Other International Initiatives**

International standards and guidelines have been developed in some other areas that relate to psychological testing and assessment. For example, an international

task force of leading practitioners in the field developed Guidelines and Ethical Considerations for Assessment Center Operations. These guidelines have been updated a number of times.

## **3. NATIONAL STANDARDS AND GUIDELINES**

International guidelines and standards have both influenced national initiatives and been influenced by them.

### **3.1. The United States**

There are a large number of organizations and committees involved in work that relates to testing and assessment practice in the United States. It is not possible to mention all of these here. However, some key ones are described in what follows.

#### **3.1.1. The “Joint Standards”**

The Standards for Educational and Psychological Testing (the 1985 version of the standards issued jointly by the American Educational Research Association [AERA], the American Psychological Association [APA], and the National Council on Measurement in Education [NCME]) have been widely adopted around the world as the authoritative definition of technical test standards. After a lengthy revision and consultation process, a new version of these influential standards was published in 1999. Many countries' psychological associations were expected to adopt these as the successor to the earlier edition.

#### **3.1.2. Test User Qualification**

Over the past few years, there has been some progress in the United States on the issue of test user qualification. In 2000, the APA Task Force on Test User Qualifications (TFTUQ) produced guidelines that inform test users and the general public of the qualifications that the APA considers to be important for the competent and responsible use of psychological tests. In this context, the term “test user qualification” is being used to indicate a desired level of competence rather than the award of some certificate or license or the outcome of a credentialing process (as is the case in the United Kingdom). The report suggests that the old three-level APA system (A, B, C) for classifying test user qualifications that was first defined in 1950 is now

obsolete and needs to be replaced by one where the competences required are defined in relation to the context, instrument, and use to which it will be put.

### **3.1.3. Computer-Based Testing**

The Association of Test Publishers (ATP) has produced technology-based guidelines for the testing industry. The goal of the proposed guidelines is the international adoption by companies involved in technology-based testing. The guidelines are for use as an aid in the development, delivery, and use of computer-based certification examinations as well as psychological testing in general. The guidelines cover applications for use on the Internet and various multimedia computer strategies used to deliver, administer, and score tests.

### **3.1.4. Internet-Delivered Tests**

In 2001, the APA set up a task force on psychological testing and the Internet to assess the extent and range of testing on the Internet and to review current practices. It aims to define and outline the issues raised by Internet testing, including test validity, administration, confidentiality of test taker and test results, test taker authenticity, ethical interpretations of test results (e.g., feedback), psychological dynamics of Internet testing (e.g., proclivity of being more revealing when taking Internet-based tests), copyright infringement, psychometric equivalence (e.g., comparability of tests results), license issue of the psychologist (e.g., crossing state lines), and interpretations based on limited assessment information.

## **3.2. European Countries**

There is a great deal of work going on in Europe in relation to the development and encouragement of best practice in psychological assessment. Some major examples are described in what follows.

### **3.2.1. The United Kingdom**

The United Kingdom was the first country to establish a test user certification procedure accredited by the national psychological association, the British Psychological Society (BPS). Both psychologists and nonpsychologists can obtain certificates in test use from the BPS. These qualifications are subject to stringent quality control or “verification” procedures and currently provide recognition of qualification in test use relating to occupational assessment. Similar

qualifications for educational testing became available in 2004. All of these qualifications are based on a set of performance standards defining the competence required of a test user and a set of guidelines defining how this competence should be assessed.

The BPS has a standardized process for the review of tests and has published a number of volumes of test reviews focused on assessment in the work and organizational assessment area. These are designed as “consumer” reviews for test users rather than being for a more academic audience.

All of the BPS procedures and services relating to testing and test use are now managed within a single organization, the Psychological Testing Center. This organization manages the delivery of test user certification and test reviewing and operates a Web site that provides information, advice, and guidance for test developers, test users, and test takers.

The BPS has produced a set of guidelines for assessment centers and standards for computer-based testing. The BPS has also produced guidelines on 360-degree feedback in collaboration with the Chartered Institute for Personnel and Development, the University of Surrey at Roehampton, the SHL Group, and the Department for Trade and Industry.

The British Standards Institute has issued a standard (BS7988) covering technology-based assessment. This standard focuses on many of the technical issues involved in the delivery of assessments by computer and over the Internet and is concerned mainly with educational testing and tests for licensing and certification. Unlike the ATP guidelines, this standard currently excludes areas of psychological testing (e.g., personality assessment).

### **3.2.2. Germany**

In Germany, the national standards institute has published a “norm” setting out standards for assessment for job selection (DIN33430). These standards set out good practice for the process of assessment and include requirements relating to test quality and test user competence.

### **3.2.3. Finland**

Guidance on good practice in assessment has been produced in Finland as a consequence of recent legislation relating to the European Union Directive on Data Privacy. This directive has had implications for all European countries on how test data can be used, the rights of test takers over their data, and what is done with the data.

Like the German standard, the Finnish guidance describes best practice for assessment process as a whole (in employment settings) rather than just for psychological testing. Recently, the Finnish Psychological Association developed test user certification procedures for psychologists and nonpsychologists based on this guidance.

### 3.2.4. Sweden, Norway, and Denmark

In Sweden, an institute, Stiftelsen för Tillämpad Psykologi (STP), was set up in 1996 to provide test user qualification certification, test reviewing, the quality audit of organizational testing policies, and an ombudsman function to which test takers may appeal. Test and test use quality assurance procedures build on the ITC Guidelines on Test Use and on the BPS test review and test user certification models. The STP also accredits organizational test policies and testing processes. Guidelines for organizational policies on test use are based on the model provided in the ITC Guidelines on Test Use.

More recently, Det Norske Veritas (DNV) has initiated a project to establish test user certification procedures in Norway, Sweden, and Denmark. DNV, one of the leading worldwide registrars for certification of ISO 9000 and related standards, is working in conjunction with the Norwegian Psychological Association, the STP, and representation from Denmark.

### 3.2.5. The Netherlands

In the Netherlands, Commissie Testaangelegenheden Nederland (CoTAN), a committee of the Dutch Psychological Association, has a long history of producing reviews of tests covering all areas of assessment.

More recently, the Institute for Best Test Practice (4TP) was set up with a set of aims and objectives very similar to those of the Swedish STP. The Dutch institute aims to improve the quality of test use in occupational and workplace assessment. It is making use of the ITC guidelines and the experience of both the BPS and the STP in setting standards with which users, instruments, and practices must confirm. The institute has developed procedures for examining the competence of test users, for reviewing tests (using the EFPA criteria), and for auditing organizations.

A guiding principle of 4TP is the need for transparency. First, the standards adopted must be open and

available for all to see. Second, all of the results of examinations carried out by the institute will be published on the Internet—registers of certified test users, reviewed tests, and quality audited organizations. Therefore, everybody interested in the quality of people, instruments, and practices will be able to access this information. In this way, it is hoped that the institute will provide a practical way of implementing quality and transparency.

## 4. CONCLUSION

The challenge for the future is to bring together the various standards and guidelines from different countries and organizations and to secure international agreement on the key issues. Work by bodies such as the ITC, EFPA, and EAWOP has made considerable progress during the past few years in moving this forward.

### See Also the Following Articles

Intelligence Assessment ■ Personality Assessment  
 ■ Psychometric Tests ■ Psychophysiological Assessment

### Further Reading

- American Psychological Association. (n.d.). [www.apa.org](http://www.apa.org).  
 Association of Test Publishers. (n.d.). [www.testpublishers.org](http://www.testpublishers.org).  
 Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment, 12*, 62–71.  
 Bartram, D. (1998). The need for international guidelines on standards for test use: A review of European and International initiatives. *European Psychologist, 3*, 155–163.  
 Bartram, D. (2001). The development of the International Guidelines on Test Use: The International Test Commission Project. *International Journal of Testing, 1*, 33–54.  
 Bartram, D., & Coyne, I. (1998). Variations in national patterns of testing and test use. *European Journal of Psychological Assessment, 14*, 249–260.  
 British Psychological Society. (n.d.). [www.psychtesting.org.uk](http://www.psychtesting.org.uk).  
 Buros Institute. (n.d.). [www.unl.edu/buros/](http://www.unl.edu/buros/).  
 European Federation of Psychologists Associations. (n.d.). [www.efpa.be](http://www.efpa.be).  
 Honkanen, H., & Nyman, K. (2002). *Handbook of good practice in personnel assessment*. Helsinki: Finnish Psychological Association.  
 International Test Commission. (n.d.). [www.intestcom.org](http://www.intestcom.org).  
 Institute for Best Test Practice. (n.d.). [www.4tp.nl](http://www.4tp.nl).



# Psychological Rehabilitation Therapies

Karen G. Langer

New York University Medical Center, New York, New York, USA

1. History
  2. Populations Served
  3. Treatment Considerations
  4. Psychological Issues
  5. Therapeutic Approaches
  6. Conclusion
- Further Reading

and/or self-sufficiency have been partially or extensively compromised temporarily or long term; interventions are aimed at restoring or improving level of independence or self-sufficiency, enhancing personal functioning and quality of life, or addressing the physical, architectural, social, cognitive, communicative, and/or other barriers to everyday life.

## GLOSSARY

**acquired disability** A loss, impairment, or decline from the previous level of functioning or bodily integrity, sustained due to illness, trauma, or injury.

**brain dysfunction** A brain disturbance that can be caused by illness or injury and that results in disruption to life-sustaining autonomic functions or to motor, sensory, appetitive, affective/emotional, and/or cognitive functions subserved by the brain areas involved.

**cognitive remediation** Provision of treatment strategies geared to improve cognitive abilities directly or to teach compensation strategies for cognitive skills weakened by brain dysfunction and help people deal with changes in cognitive functions.

**congenital disability** A functional or physical impairment originating from a developmental event or deviation during or prior to birth of nonhereditary origin.

**rehabilitation** An interdisciplinary treatment approach, devoted to the optimization of physical and psychological functioning, where independence, personal autonomy,

Psychotherapy has among its goals establishing a new, or restoring a prior, sense of identity to patients whose sense of self has been deeply threatened by mental or physical trauma. Along with decreasing emotional suffering, the restoration of identity may be a key outcome of psychotherapeutic treatment in rehabilitation. Patients' losses include not only the physical aspects of disability but also the symbolic and psychological losses that ensue. Cognitive deficits may develop as a result of brain dysfunction, and these may also be targets of remediation techniques designed to compensate for, or directly strengthen, those functions that were weakened. Emotional sequelae, including depression, anxiety, terror, anger, blunting, constriction, instability, lability, and agitation, may also be foci of treatment. Broader issues involving psychosocial and interpersonal needs, physical dependence, personal care routines, medical comorbidities, vocational disruptions, and financial losses all can have impact. All that forms the fabric of quality of life may need to be

considered in treating patients with disabling loss in rehabilitation.

## 1. HISTORY

Among the early developments in the field of rehabilitation were concerns for the holistic treatment of people with physical disabilities. Dr. Howard Rusk, the pioneer of rehabilitation medicine as a specialty within medical practice, was dedicated to the notion that persons served were best viewed as people within a complex of not only physical but also psychological and social facets. The latter would include a focus on the emotional effects of disability and the social implications as well as on the vocational and avocational concerns that need to be considered.

The field of rehabilitation psychology was an outgrowth of the field of rehabilitation medicine in the sense that psychologists devoted themselves to the study and treatment of persons with acquired or congenital disabilities. Psychological treatment and psychotherapeutic treatment of persons with disabilities are now considered to be an integral part of rehabilitation medicine and behavioral medicine programs. Psychological rehabilitation draws heavily from fields such as rehabilitation, clinical psychology, neuropsychology, health psychology, behavioral medicine, counseling psychology, and developmental psychology. Practitioners come to the field from one of these areas and gain the specialized training necessary to confront the unique challenges encountered with rehabilitation of patients in psychological treatment.

## 2. POPULATIONS SERVED

Psychologists serve in adult and pediatric settings that include all ages of the life span—pediatric, adolescent, general adult, and geriatric. The populations served in rehabilitation and medical settings involve a great diversity of conditions. These include traumatic injuries of all kinds and illness-related disabilities such as neurological disorders, neuromuscular disorders, cancers, debility due to deconditioning from prolonged illnesses, amputations, spinal cord injuries, brain dysfunction, head traumas, and orthopedic conditions—in short, any conditions that may result in physical limitations or disabilities. Patients may have sustained multiple injuries in ski accidents or falls or may have been involved in car or diving accidents, resulting in spinal

cord or head injuries. They may have been victims of natural disasters or terrorist attacks, collapsed buildings, gunshot injuries, burns due to fires or explosions, or other tragic events. Patients may have undergone back surgery for spinal conditions such as stenosis or may have undergone hip replacement surgery due to osteoarthritis or degenerative joint disease. Others may have more chronic conditions such as Parkinson's disease, lupus, and multiple sclerosis.

Onset may be sudden such as in stroke, myocardial infarction, and Guillain–Barre syndrome. Alternatively, sudden onset (e.g., lower extremity amputation) can sometimes follow a prolonged period of declining function/disability and pain or even progressive surgeries. Time factors may play a role in patients' experience of disability, including chronicity and duration of condition, age at onset, and congenital versus acquired disability. For some patients, the longer the condition has existed, the more time they have had to adjust; for others, the pain or disability may have persisted for a long time, so that the stress can be additive or cumulative. Other factors to be considered include the nature of onset (e.g., traumatic vs nontraumatic, accident vs illness related), risk-to-life factors (e.g., initial brush with death, terminal nature of condition), and prognostic descriptors (e.g., progressive, stable, remitting, expected to improve). Also, the treating clinician must consider the question of awareness, that is, what the patient has been told, whether the patient knows the condition, and the patient's ability to be aware of implications based on emotional factors such as denial.

Psychological treatment in rehabilitation often deals not only with the stresses imposed by disability and altered life circumstances but also with issues of hospitalization and medical tests and procedures. Posttraumatic stress disorder may be an aspect of patients' emotional responses as they relive the accidents or traumas, sometimes in the form of flashbacks or nightmares. Naturally in the course of continuing treatment, psychologists may deal with a broad range of evolving and changing responses to the whole spectrum of effects of disability and its life-altering impact.

## 3. TREATMENT CONSIDERATIONS

### 3.1. Assessment

Psychological intervention often follows an assessment that can range from 1 hour to several hours and

generally taps all pertinent aspects of emotional response to the loss, change, or trauma. Emotional reactions may include anger, anxiety, depression, agitation, emotional lability, emotional stability, and even apparent indifference. Signs of affective blunting or constriction may be noted and need to be distinguished from illness-specific features such as the mask-like face (i.e., masked facies) often associated with Parkinson's disease and prosodic deficits (i.e., aprosodia) often associated with right hemisphere brain dysfunction. Included in this assessment is a focus on psychosocial history, vocational history, health habits (e.g., dietary, exercise, smoking, alcohol, recreational drugs), medical history, premorbid psychological functioning and history of treatment, suicide risk, ability to communicate emotional needs based on affective constriction and defensive styles, awareness of losses, emotional denial, cognitive ability to understand and appreciate the losses, prior risk-taking behaviors, educational history, and living arrangements. Also tapped are features of mental status and cognitive function as indicated through cognitive or neuropsychological testing during the process of psychological assessment.

### 3.2. Setting

The setting in which the patient is treated is also meaningful. Early phases of rehabilitation, which are often acute care or hospital based, may engender needs that are centered around medical or nursing needs, personal care routines, bodily functions, and/or food intake. Treatment often involves psychoeducation along with cognitive-behavioral or psychodynamic approaches to help educate a frightened patient who is unfamiliar with medical/hospital routines and terminology. Issues of dependence on caregivers—often strangers—for basic personal needs can engender feelings of helplessness, shame, anger, and/or depression as the patient grapples with losses to dignity, self-esteem, and sense of control over bodily functions. The treating psychologist must be familiar with the physical status and needs of the patient. By conveying knowledge of the medical/physical needs and a sense of competence, the experience becomes more normalized for the patient, and expectations are addressed from both educational and psychological perspectives.

### 3.3. Nature of Stressors

Disability itself produces stressors that are cumulative or interactive. Physical losses, including motor deficits

and sensory losses, affect all aspects of mobility, elimination functions, and activities of daily living. Speech or swallowing deficits may ensue, affecting quality of life in terms of communication and basic appetitive pleasure. Personality changes or cognitive losses may have far-reaching effects on sense of self, interactions with family and others, and ability to resume prior activities such as work and leisure activities. The Americans with Disabilities Act of 1990 offers guidelines and legal protection to people with disabilities regarding access to public services, transportation, employment, and so forth. Yet the modifications and the special equipment needs, as well as their visibility, may also be stressful. The fact is that a physical disability can involve disruption not only to the body and its functions but also to more psychological entities. In fact, among the most devastating losses experienced by the patient may be losses to aspects of the self. The role of cultural factors and attitudes is also critical in shaping the individual's responses to a disability and the responses of others in that subculture. Nonetheless, familial and individual differences are many and may have even more meaning to the patient than to the subculture alone.

### 3.4. Developmental Considerations

For congenital conditions producing physical or cognitive disabilities, there are complex issues pertaining to the development of sense of self involving body image and sense of mastery. These may be particularly salient for the child who is growing up with a physical disability. Patients may grapple with guilt regarding the impact of their disabilities on their families. In some instances, especially if there are genetic components to a disabling illness, there may be added stress on the parents' marriage and concerns for genetic transmission. Some adults with lifelong disabilities have come to develop special skills that help them to adapt in the world, including social skills, navigational skills, and personality features. These skills, which are highly prized for their adaptive value, may form a core part of the identity in a very positive way. Identification with disability may then be very strong and complex. Patients with exacerbation of long-standing disabilities (e.g., postpolio syndrome) often find it difficult to take advice from less seasoned staff members who might not know their conditions as intimately as do the patients.

Other developmental considerations pertain to stage of life, such as when the disability occurred during the

life span, prior experience with illness or disability in oneself or others, and the availability of supports. The latter is crucial in working with an aged population where many family members, who could be of support, may have already died or become disabled themselves.

#### **4. PSYCHOLOGICAL ISSUES**

The issues that are confronted by the patient and psychotherapist are many and include mortality, mutilation, loss of limb, helplessness, loss of body functions, alterations in sexual functioning, loss of psychological entities such as sense of competence and mastery, loss of personal identity and ways of knowing oneself that may be shattered or threatened by physical injury, loss of job, loss of one's sense of manhood or womanhood, and role alteration within the family or with friends. Patients are frequently frightened of repeating trauma and may have flashbacks or dreams of impending disaster anticipated with devastating illness. In end-stage illness, pain management is a concern, as is worry about losing one's mental faculties, whether due to the illness or due to medication. Sometimes, there are issues of unawareness, and these are among the most challenging ones because the patient might not even realize the severity of the situation. This unawareness is often due to cognitive dysfunction itself but is sometimes due to significant emotional denial that blocks the awareness and the ability to make sound judgments regarding care or planning needs. There may be overestimation of abilities that can pose ethical challenges to the treating therapist with regard to issues of competence. These often do not require legal/judicial intervention because there might also be the possibility of improvement in the patient's status.

Psychological treatment of the patient in rehabilitation is aimed at the emotional sequelae of the event onset or ongoing situation of loss. Reactions such as depression, anxiety, anger, agitation, emotional lability, and instability of mood are very common. The supportive normalization of certain reactions is a key aspect of forming the therapeutic bond and educating patients, particularly less medically sophisticated patients, about what to expect. For many others, a more aggressive approach to treatment of the significant psychic distress and emotional disorders is indicated. The notion of losing control not only over physical functions and body parts but also over the emotional response system can be very threatening

for patients who previously took pride in emotional coolness. Patients who used denial or avoidance as defensive styles prior to the onset of their illnesses/disabilities will continue to use these defenses in dealing with the disabilities. Denial, as a response to disability, poses a significant challenge to treatment not only by psychotherapists but also by other treatment providers such as physicians, rehabilitation staff, and family members or caregiving aides.

The degree to which there is distress and suffering is crucial to consider in any psychological treatment. A depressive reaction may be chronic whereby the disability and its privations persist over years or decades. Facilitating adaptation to disability may then be a realistic goal for the psychologist. Adaptation, unlike "acceptance," means that an individual is coping with and dealing with the situation to the best of his or her abilities. Adaptation does not imply that the person has accepted a notion of permanence or resignation, and it certainly does not imply that the person has lost hope of improvement (although he or she might very well have). Restoring identity to the patient whose physical or cognitive losses have compromised his or her identity is an integral part of rehabilitation and therapy.

#### **5. THERAPEUTIC APPROACHES**

Psychological treatment may evolve from crisis intervention to approaches that address deeper issues as the patient emerges from the acute phase of shock or traumatic stress. In time, therapies using more depth or psychodynamic models may be chosen to work on emotional issues, or cognitive behavioral approaches may be used to deal with depression, anxiety, or other emotional responses. Stress management techniques are employed to help relieve anxiety, tension, and posttraumatic stress. Other models, such as existential explorations of meaning of loss and life, can be selected for use by the therapist as indicated.

Once adapted to the needs of the setting and the patient, traditional psychodynamic approaches are helpful to the patient in dealing with new or unresolved needs and conflicts that may pertain directly to the disability. Issues regarding dependence and care, having developmental roots, or reflecting challenges to the personality structure will often surface as the patient grapples with unwelcome changes in physical functioning. The impact on well-established patterns of interactions and roles within family and other social structures assume primary importance to the

patient. Thus, psychodynamic approaches can be very useful in helping patients, as well as staff and family, to understand the basic emotional features that are threatened by the disability (e.g., the need for self-reliance or control) or are exacerbated by the disability (e.g., fears about abandonment or loss of love). Historical roots of needs or conflicts that have remained dormant prior to the onset of disability may be more amenable to exploration since the onset and may allow for unexpected resolution. Support, empathy, and genuine caring can be integral parts of the healing process in adjusting to the trauma of loss and disability and of the conflict/need resolution process in using the psychodynamic approach to treat the patient in rehabilitation.

Stress management and relaxation techniques, alone or in conjunction with the other forms of psychological treatments offered, are very useful in the treatment of the medical or rehabilitation patient.

Other psychological treatments may involve behavioral management strategies employed when behavioral disruptions occur. These disruptions can be induced by a stress reaction in the context of usually more limited coping skills. Alternatively, they can occur as a result of restricted access to the patient's previously used coping skills due to the limitations of disability itself. There may also be underlying psychiatric disturbance that is exacerbated or aggravated by the psychic trauma of disabling losses, leading to psychic regression or decompensation. Sometimes, there may be primary effects of brain dysfunction on behavior (e.g., initiation, sexual disinhibition, problem-solving skills, memory, attention/concentration, impulsivity, agitation, aggressive responses, adynamia, apathy). These effects may disrupt behavior to a greater or lesser degree, and when they are seen in the social sphere or with family, they can be very disturbing. Often, the psychologist treats the patient's behavior through a combination of psychotherapy and behavioral conditioning principles. Behavioral management meetings may be organized to help staff implement individualized approaches to behavioral management issues, whether due to significant cognitive dysfunction or the patient acting out under stress.

Because some patients might not even realize a need for psychological treatment, they may appear to be uncooperative, defensive, unmotivated, and/or agitated. Judgment and safety may be affected by lack of awareness or by impulsive behavior. Patients might not even fully realize the implications of their condition, or may not yet realize the extent of the change in their behavior or emotional responses. These are challenges

to treatment and may put the burden of goal setting on the treating therapist until these patients can better acknowledge changes and appropriate treatment goals.

With regard to the cognitive impairments that may accompany certain disabilities, specifically those involving brain function of various etiologies, there is a tradition of cognitive remediation treatments. Such treatments are frequently adjunctive to treatments offered by other disciplines, such as occupational therapy and speech therapy, and are often integrated into the patient's physical therapy program. Therapy is tailored to the nature of the deficit, so the patient may be treated for basic disturbances. Deficits are measured by using neuropsychological assessment tools if possible. If deficits are too severe, they are measured by observational assessment and by comparison with scales of recovery after brain injury. Neuropsychological deficits may involve arousal level (i.e., hyperaroused or hypoaroused), attention and concentration, orientation, thought coherence, visual perception, visuospatial skills and visual attention, reasoning, coherence, judgment, planning, initiation, response latency, organization, inhibition of response and problem solving, all aspects and types of memory, and language functions (with the latter being treated primarily by speech and language pathologists), among others. Basic functions, such as reading, writing, copying, and arithmetical skills, may be affected by any of these deficits.

For brain dysfunction, model systems have been designed to integrate the cognitive-behavioral treatments within the entire rehabilitation treatment program throughout the day. Various brain injury facilities have established treatment programs that integrate the conceptualizations of neuropsychological recovery after brain injury. These programs may exist ideally within a dedicated unit, but they also may be effective in "units without walls." These are designed to have the complete treating staff integrate the recommended cognitive approaches within their disciplines, tailored to the individual cognitive, behavioral, physical, and emotional losses of the patient with diffuse or traumatic brain injury.

Cognitive remediation treatment may target skills lost or weakened by brain dysfunction and may involve either direct remediation of the weakened skill or compensatory strategies that help the patient deal with and compensate for the losses (e.g., training in the proper use and organization of a log/notebook for the patient with memory storage/retrieval difficulties). There is some debate in the literature regarding the

efficacy of compensatory versus direct remediation strategies. Both individual remediation and group remediation have been used in different settings. Psychological treatments have also been used in novel ways to enhance physical functioning (e.g., biofeedback techniques to improve aspects of motor control in certain patients).

Treatment of comorbid psychiatric or substance abuse issues may pose a challenge in the medical rehabilitation setting. However, these issues may arise with a patient who has a preexisting personality disorder (with aggressive anger, impulsivity, risk-taking, and/or acting-out behaviors). In fact, the personality trait may perhaps be expressed in more exaggerated form due to the emotional stress of loss or to the direct or secondary effects of brain injury. Sometimes, particularly in the aftercare or outpatient setting, the patient may even be referred to the psychologist specifically for treatment of these disturbances (e.g., impulsivity, anger/aggression management). The impact on significant others and family may be profound, especially if there is a change or worsening of a preexisting problem or a new emergence of a disturbing behavioral symptom. Family members may feel at a loss to deal with these disruptions, and counseling may be beneficial.

## 6. CONCLUSION

Depending on the setting, training, and need, therapists may take a more active or directive role. Some may become involved in coordination on behalf of patients for social services and/or vocational needs; with agencies, the challenge is to help the patients rebuild their lives. Others may be more involved in the team approach. The key is realization that the patients' needs may be complex and interdependent. Often, treatment is not offered in a vacuum and ideally is offered within a coordinated system, however tightly or loosely defined, because patients might not be able to negotiate through other aspects of life planning due to devastating change and loss or to the cognitive disruptions (if any). The specific approach and training of the therapist may dictate whether psychodynamic, cognitive-behavioral, stress management, cognitive remediation, or more counseling/educational-type approaches are used, or are useful, for the needs of patients.

Rarely are the challenges of identification more apparent than in working with persons who may be of similar age, background, and/or vocation and who are

reasonably healthy until some catastrophic loss occurs. Psychologists might need to use the identification process in a manner that does not cloud patients' individuality and uniqueness of response, so that the therapists maintain the therapeutic stance needed to work effectively while retaining subjective understanding. The task of simultaneous subjectivity and objectivity is, of course, not unique to the practice of psychotherapy in a rehabilitation setting. In the work of psychological rehabilitation, however, there is a constant and humbling realization of one's own humanity and vulnerability. This sense of personal vulnerability may tug at the heartstrings. With time, experience, supervision, and self-awareness, this task is made easier for psychotherapists. Of course, the remarkable ability to adapt may be a humbling experience for clinicians working with patients rallying against all odds to cope with tasks of daily life that have become monumental. The inspiration derived can fuel professional growth, enhance therapeutic work, and enrich personal meaningfulness of the many values that form part of therapists' identity.

## See Also the Following Articles

Coping ■ Stress

## Further Reading

- Americans with Disabilities Act (1990). 42 USC 12101nt.
- Christensen, A-L., & Uzzell, B. (Eds.). (1999). *International handbook of neuropsychological rehabilitation*. New York: Kluwer Academic/Plenum.
- Eisenberg, M. G., Glueckauf, R. L., & Zaretsky, H. H. (Eds.). (1999). *Medical aspects of disability: A handbook for the rehabilitation professional* (2nd ed.). New York: Springer.
- Frank, R. G., & Elliott, T. R. (Eds.). (2000). *Handbook of rehabilitation psychology*. Washington, DC: American Psychological Association.
- Langer, K. G., Laatsch, L., & Lewis, L. (Eds.). (1999). *Psychotherapeutic interventions for adults with brain injury or stroke: A clinician's treatment resource*. Madison, CT: Psychosocial Press.
- Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.
- Marinelli, R. P., & Dell Orto, A. E. (Eds.). (1999). *The psychological and social impact of disability* (4th ed.). New York: Springer.
- Rosenthal, M., Griffith, E. R., Kreutzer, J. S., & Pentland, B. (Eds.) (1999). *Rehabilitation of the adult and child with traumatic brain injury* (3rd ed.). San Diego, CA: Academic Press.



# Psychological Skills Training in Sport

Michael J. Mahoney and Benjamin P. Chapman

University of North Texas, Denton, Texas, USA

1. Introduction
  2. Cognitive Skills
  3. Emotion Self-Regulation Skills
  4. Motivational Skills
  5. Social and Interpersonal Skills
  6. Training Psychological Skills
  7. Conclusions
- Further Reading

## GLOSSARY

**anxiety** An experience characterized by fear, doubt, or worry; at a physical level, it is often felt as a racing heart, a nervous stomach, jitteriness, or jumpiness.

**arousal** A physiological state characterized by heightened activity of the sympathetic division of the autonomic nervous system; it prepares an organism for action and is a common, if not necessary, component of most athletic experiences.

**efficacy expectations** The belief that one is capable of performing a given behavior in such a way as to bring about a desired outcome; experienced athletes may hold higher efficacy expectations about their abilities to perform various skills and tasks in sport.

**flow** A state of consciousness in which one's attention and concentration are focused on an activity with precision and clarity; the performance of the activity is experienced as effortless and pleasurable and is often superior by objective standards.

**individual zone of optimal functioning (IZOF)** The intensity magnitude of a given biopsychosocial state in which an individual athlete performs best; IZOFs exist for the behavioral, cognitive, affective, motivational, bodily, operational, and communicative domains.

**imagery** The internal experience of sensory stimuli that are not externally present; for instance, visualization involves seeing something through "the mind's eye" and is a skill often used by athletes to calm anxiety, focus form, and facilitate performance.

**mastery experiences** Episodes of behavior in which a desired outcome is achieved through the actor's own effort; they lead to a sense of accomplishment, and when they accumulate, they facilitate a sense of competence and encourage efficacy expectations.

**mental rehearsal** A set of activities wherein an athlete imagines himself or herself performing some skill or task or behaving in a desired fashion; although this may involve imagery, elite athletes may mentally rehearse other components of an experience such as what it will feel, sound, and even smell like to perform an action.

**psychological skills** A set of techniques and strategies used to optimize athletic performance; they are largely nonphysical in nature.

**self-instruction** A process that encompasses internal reminders, reassurance, and self-talk; it may be used as a self-regulatory tool and is often effective in the remediation of some of the negative emotional states that athletes commonly experience.

**self-monitoring** The process of attending internally to one's ongoing cognitive and emotional processes as well as externally to one's behavior and performance.

**self-regulation** A process that is dependent on self-monitoring and involves evaluating one's degree of success in achieving desired goals as well as implementing both internal and external feedback to make necessary corrections; self-monitoring and self-regulation in sport performance appear to be skills associated with athletic success.

Psychological skills in sport constitute a set of techniques and strategies used to optimize athletic performance. Psychological skills have traditionally been neglected in favor of physical skills in assessing athletic performance. However, increasing evidence suggests that psychological skills may be just as important as their physical counterparts, particularly in elite-level athletes. Psychological skills in sport comprise different practices and strategies across the realms of behavior, cognition, emotion, motivation, and interpersonal functioning. Some techniques may facilitate the enhancement or development of various skills in athletes; however, much of what is known about psychological skills has been gleaned from studies of elite athletes, and research on skills training is not yet extensive. Therefore, caution is recommended in attempting to train psychological skills in a "one size fits all" fashion. However, interventions may be more confidently "targeted" to a specific group or, better yet, can be "tailored" to a specific individual.

## 1. INTRODUCTION

Andrea shakes out the tightness in her arms and shoulders while she waits in the shadows outside the gym's side entrance. She peeks around the corner and is hit by a gusty winter breeze from the open doors across the floor. The murmuring crowd looks like a tidal wave, rising high up into the rafters, waiting to crash down on her. She has competed in the championships before, but never in front of this many people.

Suddenly, the loudspeaker crackles, and the master of ceremonies announces a 5-minute break to check equipment before Andrea is due to compete. Her heart surges, and suddenly her limbs feel stiff and her skin feels clammy. She had been all ready, psyched up, and in the perfect rhythm. Now, she will get cold and go

out there and make a fool of herself in her biggest meet ever.

Andrea thinks of her coach. He is not there, but she knows what he would say. She takes a deep breath to slow down her heart, remembering how well she has done in big meets before. Her practices have been going better than ever lately. Didn't she tell her friend just last night that her confidence was at an all-time high? Andrea closes her eyes and imagines what she will be doing in 5 minutes, what her body will feel like as it goes through the motions that are by now second nature to her. Her heart begins to slow, but at the same time, she bounces on the balls of her feet, does pushups off the wall, shakes her legs out. Blood returns to her limbs, replacing the clammy, paralyzed feeling. She takes another deep breath.

Andrea's name buzzes over the loudspeaker, but it seems distant, just loud enough to reach her. She walks into the gym, concentrating only on the equipment and imagining what her performance will feel like. The crowd begins to disappear; she hears only the rhythm of her own breathing and sees only the equipment. With a calm mind, she steps into the competition area, her body feeling like a compressed spring that is warm and ready for action.

In a minute, it is all over. Andrea steps back and looks up. The judges nod approvals, but she already knows that her performance was a personal best. The crowd's cheers spill down on her now like fresh air, no longer a suffocating tidal wave. She thinks of her coach and smiles.

This scenario is one example of many performance situations that confront athletes. Naturally, the situational demands of athletics vary largely depending on the sport. Is it an "open" sport, where the challenges and demands are constantly changing, or is it a "closed" sport, where the tasks are performed against a relatively stable backdrop (e.g., the green on that troublesome fourth hole does not suddenly change its size or slope, despite the frustrated impressions of golfers)? Is the athletic endeavor an individual one, such as wrestling or weightlifting, or a team game, such as soccer or basketball?

Research on elite athletes has always studied the psychological dimensions of athletic performance. What makes an athlete a good athlete? Given equal physical capabilities, what makes one athlete better than another? What makes a good athlete an extraordinary one? Contemporary sport psychology emphasizes a variety of factors that may influence performance differently across individuals and situations. For a

time, however, the presumably stable personality traits of athletes were thought to play the largest role in their success at sport.

### **1.1. The Athletic Personality**

During the 1960s and 1970s, research examined personality differences between athletes and nonathletes as well as between successful athletes and unsuccessful athletes. Was there a general “athletic personality”? Were athletes more extraverted than their nonathletic peers? Were elite athletes a “different breed” from their less successful counterparts—more disciplined and aggressive? To explain the personality type associated with good athletes, Morgan proposed a “mental health” model suggesting that athletic success is inversely related to psychopathology. Other evidence gathered through the use of conventional personality inventories suggested that successful athletes tended to be more assertive, dominant, aggressive, self-sufficient, and vigorous; to have a higher need for achievement; and to be less anxious than less successful athletes.

Although such personality factors seemed to explain some of the differences between more successful athletes and less successful athletes, research of this sort fell out of favor by the mid 1970s. Sport psychologists grew more interested in differences between individuals rather than group commonalities. Successful athletes vary widely in characteristics such as pain tolerance, optimal arousal levels, and anxiety, and these factors interact in different ways across different athletes in different situations. So, if there is no general “athletic personality” solely responsible for one’s performance in sport, what accounts for athletic success?

### **1.2. Psychological Skills in Sport**

The answer appears to lie in the broad array of psychological skills reported by successful athletes. Surveys of elite athletes and studies of peak performance reveal a variety of mental, emotional, and motivational techniques that appear to account for a large part of success in sport. Evidence indicates that elite athletes employ mental imagery and bodily feedback, generate and maintain motivation, handle performance anxiety, concentrate, and retain self-confidence more effectively than do their less successful counterparts. Furthermore, such skills may be at least partly modifiable.

This article considers psychological skills in five broad categories: those skills that are primarily bodily

and behavioral, those that are cognitive, those that involve primarily emotions, those that concern motivational states, and those that encompass interpersonal functions in sport. Bodily and behavioral skills relate to controlling one’s physiological functions and actions such as regulating one’s bodily state and readiness and acquiring and polishing motoric routines that are necessary for successful performance. Cognitive skills encompass a broad class of activities such as using mental imagery, monitoring one’s thoughts and cultivating effective self-talk, and attending and concentrating in optimal ways. Emotional skills typically involve the regulation and optimization of affective states, including the ability to control anxiety and foster positive mood and feelings about oneself. Motivational skills provide athletes with inducement or impetus to excel and may involve things such as goal setting and responding to social facilitation. Finally, interpersonal skills relate to one’s smooth interaction with others in sport and include communication abilities and the regulation of emotion in a team environment. Many skills function across several of these domains. The current classification scheme is intended only for ease of presentation and description and not as a discrete taxonomy.

### **1.3. Bodily and Behavioral Control Skills**

Behavioral and bodily control skills operate at the physical level to enhance athletic success. Two primary skills involve regulating one’s level of arousal and practicing or physically rehearsing motor skills.

#### **1.3.1. Physiological Arousal Regulation**

Physiological arousal refers to a state of heightened activity within the autonomic nervous system. High levels of arousal bring on the so-called “fight or flight” response, a reaction to threat and stress inherited from humans’ distant mammalian forebears. Athletes may experience physiological arousal through the distinctive “charged-up” feel of adrenaline, increased flow of blood to skeletal muscles, increases in heart rate and respiration, and so forth.

It is important to note that such physical sensations often accompany the emotional experience of panic or anxiety. Although anxiety usually involves a strong component of physiological arousal, the two phenomena are not identical. An athlete may be in a state

of high physiological arousal but not be anxious; for instance, he or she may be angry. Similarly, an anxious athlete may experience low levels of physiological arousal, particularly if he or she is exhausted but still worried.

Some have suggested that, at an optimal level, arousal may actually diminish anxiety and/or serve to focus selective attention. For these reasons, different athletes often prefer different levels of arousal. One useful way in which to understand this relationship is from the perspective of an individual zone of optimal functioning (IZOF), an idea proposed by Hanin.

An IZOF corresponds to the intensity magnitude of a biopsychosocial state within which a given athlete feels most capable of executing the skills that are necessary for success in his or her sport. Examining autonomic arousal from this perspective, athletes must discern the levels of arousal at which they perform best. This may vary across behaviors, skills, and situations. Athletes who are aware of their somatic IZOF are able to detect when they are over- or under-aroused and then can take corrective measures to regulate their arousal.

The ability to self-regulate arousal levels is critical for athletic success. This may take the form of last-minute mental or physical rituals to increase or decrease arousal, as discussed previously, or may be a more gradual process drawn out for hours (or even days) prior to a competition. In fact, the regulation of arousal can be important even during practice sessions. The ability to “turn on” or “turn off” the autonomic nervous response is largely a learned skill that is discussed later in the article.

Finally, different sports require different levels of physiological arousal. Games that are dependent on fine motor skills—archery, golf, shuffleboard, billiards, and so forth—clearly require a relative degree of calmness for effective skill execution. Sports that require strength, gross or explosive motor movements, and combative maneuvers—football, wrestling, sprinting, boxing, and so forth—generally necessitate higher levels of physiological arousal, although again optimal zones vary by individual athlete. It is also possible to be overaroused even in such sports; for instance, a soccer player may lose what is called “touch” in his or her passes and shots, instead kicking the ball into the stands.

### **1.3.2. Behavioral Practice**

Motor learning of sport-specific skills requires physical repetition. Different degrees and types of repetition

increase different skills at different rates in different athletes. But whether skill acquisition is sudden or gradual, the adherence to some sort of practice regimen is nearly always necessary for the learning, maintenance, and continued development of athletic skill.

One consideration is the amount of practice. Early research on the amount of practice in motor performance suggested that plateaus or ceilings sometimes may be reached, but modern work on deliberate (i.e., effortful) practice suggests that sustained practice generally leads to continued improvement, even over long periods of time. The relationship between practice and skill acquisition may or may not be linear and is usually a function of the athlete, the skill, and the degree of learning that has already taken place.

The variability or breadth of practice is also an important consideration. To learn a complete repertoire of sport-specific motor skills, an individual must practice the complement of skills that the sport requires. Sometimes, it is also beneficial to practice under a variety of conditions, particularly if the conditions of competition vary (as in changing weather patterns or variable lighting). Thus, varying the motor behaviors themselves and the conditions under which they are executed may lead to greater degrees of overall skill acquisition.

Finally, performance on complex motor tasks may be improved by decomposing them into simpler component movements. An athlete practicing a partial movement may focus attention and effort more thoroughly. Thus, proficiency acquired at these task parts may then translate into greater proficiency at the whole task. Part task practice typically works best for behaviors in which component parts are not highly interdependent; complex behaviors in which components are more highly organized and coordinated may benefit less from decomposition into separated constituents.

## **2. COGNITIVE SKILLS**

Cognitive skills are those that primarily involve mental processing. This section discusses three categories: attention and concentration, mental rehearsal and imagery, and self-monitoring.

### **2.1. Attention and Concentration**

Within the realm of sport, attention and concentration are important in many ways. Athletes need selective

attention to extract relevant features from the field of activity. Likewise, divided attention may be called on to follow diverse streams of action, particularly in open sports. In the earlier example, Andrea must monitor processes inside herself, such as her bodily feelings and sensations, as well as the events in the competition area. Without good divided attention skills, she might have missed her name when it was called over the loudspeaker.

Attention and concentration skills involve focusing mental effort on task-relevant cues and filtering out distractors. Some evidence has linked attention to optimal levels of arousal, and it has been suggested that the superior attention skills of elite athletes are related to their ability to regulate their own arousal levels. Optimal levels of arousal may narrow or focus attention on task-relevant details in the environment. Conversely, under- or overaroused athletes may find themselves to be distractible. But attention may also be focused too narrowly, with less successful athletes overfocusing on minute details of competition and not seeing “the big picture.” In addition, less successful athletes may tend to allow their attention to focus on negative perceptions of themselves or their performance.

In this vein, Moran has classified challenges to attention and concentration as arising from either external distractors or internal distractors. External distractors arise from the environment and generally involve things such as crowd noise, equipment and playing conditions, verbalizations and strategic interference from opponents, and visual distractions. Internal distractors involve things such as misplaced or excessive thoughts about the past or future, waning motivation and fatigue, and anxiety. Attention and concentration skills permit athletes to screen these distractors and focus on relevant cues and situations developing during a practice session or competition.

Also relevant is Nideffer’s model of attention, which proposes that attention may be either internally directed or externally directed and be either broad or narrow in scope. Different athletes may have preferred attention styles, and different sports may have different attentional demands. For instance, in soccer, an individual must direct attention largely externally and in a broad fashion to teammates, opponents, and other outside task demands. Optimal attention is a function of how well the individual’s preferred attention mode fits that demanded by the sport. From a psychological skills perspective, successful attention during sport may be contingent on an athlete’s ability to detect and implement the sort of attention required.

Some research has also found that elite athletes tend to use associative, as opposed to dissociative, attentional strategies. Associative strategies involve an active comparing of task demands with one’s physical capacities—a continual comparison of external and internal perceptions. Dissociative strategies, on the other hand, involve separating external task performance from kinesthetic and somatic feedback, that is, “shutting out the pain” or distracting oneself from unpleasant bodily sensations. For instance, the fact that elite distance runners report using associative attention (whereas mediocre runners report using dissociative strategies) suggests that superior performance may in part be a function of matching athletes’ physical abilities to external task demands—a process that is difficult, if not impossible, if athletes have “tuned out” their bodily signals.

The ultimate form of concentration occurs during episodes of peak performance labeled as flow. Peak performances are episodes in which athletes report complete engagement with their immediate tasks, sometimes yielding results of an incredible nature. Several features of peak performance episodes appear to be noteworthy. Athletes in flow report that all things irrelevant to the immediate task tend to recede from awareness. The spectators, distracting thoughts, and indeed even pain may disappear from athletes’ perceptions. All available resources are focused unswervingly on the athletic performance. Episodes of peak performance are characterized by such supreme concentration that even an awareness of time may be diminished or lost. Finally, athletes reporting flow experiences often say that they feel as though their performance is executed with an automated ease.

## 2.2. Mental Rehearsal and Imagery

Another skill that clearly contributes to athletic success is mental rehearsal. Mental rehearsal refers to the practiced imaging of successful athletic performance. Imagery refers to the specific thoughts, sounds, smells, visual array, and sensations that an individual experiences when he or she conjures up a mental experience of something. Mental rehearsal and imagery, particularly of specific skills, may facilitate motor learning. Repeatedly imagining successful athletic execution may enhance an individual’s certainty that he or she is capable of performing a given skill. In contrast, evidence suggests that although successful imagery is related to better performance, imagining unsuccessful outcomes tends to be a detriment to performance. Elite athletes

also report that the images they employ are extraordinarily clear, vivid, and multisensory—nearly lifelike.

Imagery is not confined to visualization. Many elite athletes report using a holistic form of imagery in which they imagine the sounds, feels, and even smells of a situation. Elite athletes also report using “first-person” imagery, that is, imagining successful outcomes from their own perspectives. Less skilled athletes may be more apt to use external imagery in which they view themselves performing some action from the outside as if they are watching a video of themselves.

### **2.3. Self-Monitoring**

The execution of physical skills is only the beginning of athletic performance. Success in sport also involves monitoring one’s performance, evaluating it accurately against expectations and task demands, and then changing or maintaining it accordingly. Also required is the awareness and regulation of thoughts, feelings, and even physiological arousal. Self-monitoring is the ability to accurately perceive and evaluate performance (or internal experience) against task demands, whereas self-regulation is the ability to manipulate one’s behavior (or internal states) accordingly.

Some evidence suggests that elite athletes are better able to self-monitor and self-regulate than are their less successful peers. In evaluating performance against demands, individuals with high mastery of the tasks at hand (i.e., more successful athletes) seem to notice and use feedback from both positive and negative aspects of performance. In contrast, individuals who are low in task mastery (i.e., less successful athletes) seem to benefit only from focusing on the successful dimension of their performance. This does not imply that to become a better athlete, one must immediately begin focusing on performance failures in addition to successes. On the contrary, such a focus may be counterproductive to self-confidence and inappropriate in less advanced athletes.

In addition to evaluating one’s behavioral outcomes against those called for in a sporting situation, self-monitoring often involves tracking thoughts. Negative thoughts often exert deleterious effects on performance. In the earlier example, Andrea experiences doubt about her ability to deal with the unexpected delay in the form of a sudden thought. The awareness of such thoughts, called negative self-statements, is extremely important in athletic performance, particularly because many people are not immediately conscious of them.

## **3. EMOTION SELF-REGULATION SKILLS**

This section focuses on two areas that are closely related to functions of emotion: the relationship of anxiety to athletic performance and the nature of self-confidence and self-efficacy.

### **3.1. Anxiety and Personal Meaning**

For many years, performance anxiety in sport was considered to be debilitating, that is, to be avoided at all costs. More recent perspectives suggest that such blanket conclusions are somewhat oversimplified. The nature and effects of anxiety vary highly across individuals and situations. To be sure, performance anxiety often hampers otherwise talented athletes, and some athletes manage it better than do others. The worried emotions that are characteristic of anxiety often have cognitive and physiological counterparts that may be controllable through self-talk or relaxation. But perhaps most important in the management of anxiety is the meaning that athletes attach to it.

For some athletes, anxiety represents a fundamental fear of failure that has deep personal ramifications. Many athletes attach their identity or self-worth to their athletic abilities and performance. In this case, failure in sport is tantamount to failure as persons, a shortcoming that is much more profound and fundamental than merely losing a game or performing poorly during a competition. Poor performance, from this perspective, is a destructive event to be feared and guarded against at all costs, yet its mere potential is so threatening that athletes can never ignore or forget it. In instances such as this, anxiety is debilitating precisely due to the meanings attached to it.

Although most athletes report that high levels of anxiety are not pleasant, anxiety may be facilitative as well. Research with some elite athletes suggests that they view their anxiety as an ally or a companion. For them, insufficient anxiety may be disconcerting because it signals a lack of interest or engagement in the task at hand. For these athletes, management of anxiety involves “pacing”—allowing a slow buildup of anxiety, beginning days or even weeks before a competition, that spurs them to maximal performance. Rather than eliminating anxiety entirely, athletes may be aided by developing an understanding of the function that anxiety plays for them. Once this has been achieved, it will be easier to know how to regulate anxiety so that it

may be kept at levels that are facilitative, but not inhibitory, of performance.

### 3.2. Self-Confidence and Self-Efficacy

Self-confidence is a general feeling of competence or certainty in one's ability to succeed. Self-efficacy refers to the belief that one can successfully perform a specific action. Both are critical to athletic performance. Because self-confidence is so vague and general, the descriptions in this section note the components that recent research suggests constitute self-confidence in sport as well as a general framework for understanding the role of self-confidence in athletes.

Vealey proposed a model in which self-confidence in sport may be seen as a product of nine dimensions: mastery, demonstration of ability, physical/mental preparation, physical self-presentation, social support, coaches' leadership, vicarious experience, environmental comfort, and situational favorableness. These nine areas can be grouped into three domains: Achievement (involving mastery of sport skills and the demonstration of an athlete's ability), Self-Regulation (encompassing an athlete's degree of physical/mental preparation and feelings about physical self-presentation), and Climate (including an athlete's sense of social support, perceptions of coaches' leadership, favorable vicarious experiences, and comfort with the environment and degree of situational favorableness). To the extent that these factors arise or are arranged in a more advantageous way, an athlete will feel more confident in sport.

At a broader level, self-confidence is not merely the absence of doubt or the rational certainty of success. Like anxiety, self-confidence may be deeply tied to athletes' personal meaning systems. Relative to their less skilled peers, elite athletes may experience a more flexible connection between their self-worth and their sport performance. The result may be a decrease in felt performance pressure because their value as persons is not totally invested in any particular performance. In other words, elite athletes may have more ability to remain "faithful to themselves" (as suggest by *confidere*, the Latin root of "confidence") regardless of the outcome of their sport performance.

Successful athletes also generally hold high efficacy expectations. Efficacy expectations are anticipations of achieving desired outcomes or goal states. Efficacy expectations may be increased through the praise and

assurance of others as well as through techniques such as mental rehearsal and imagery. But efficacy expectations are usually developed through repeatedly performing a graduated series of tasks more and more successfully. This process is called mastery.

When an individual performs some action successfully, it serves as a mastery experience. A mastery experience involves executing a behavior that leads to a desired goal state, with the result being an increase in the athlete's confidence that he or she can perform the action successfully in the future. Practice in sport allows athletes to build self-efficacy through potential mastery experiences—but in a safe environment where failure does not result in the loss of competitions. Successful performances in prior competitions may be even better mastery experiences because they involve achieving goal states under conditions similar to future competitions. In the earlier example, although Andrea has never performed before such a large crowd, she has competed previously under conditions that were similar in other ways.

## 4. MOTIVATIONAL SKILLS

This section discusses two types of motivation common in athletics: goal setting and social facilitation. Goal setting usually serves long-term purposes, whereas social facilitation is generally an external motivator on which some athletes capitalize.

### 4.1. Goal Setting

Goal setting marshals and directs an athlete's energies and efforts either in the short term, as in Andrea's intention to win her immediate competition, or in the longer term, as in her plans to increase her conditioning during the off-season. Successful athletes generally report using personally meaningful goals. Choosing goals with little personal relevance is likely to produce diminished effort. In pursuing their goals, elite athletes also tend to report a healthy balance between their athletic pursuits and other dimensions of their lives. Athletes who drive themselves single-mindedly may generate momentum in the short term but often tend to ultimately lose balance in their lives. Because of this, they may be more susceptible to slumps in motivation and performance.

In general, research suggests that performance benefits more from specific and/or realistic goals than from vague and/or unrealistic goals. Likewise, some evidence suggests that intrinsically motivated athletes

may improve when allowed to set their own goals, whereas extrinsically motivated athletes may improve when others set their goals. Another area of research has examined the effectiveness of goals aimed at task mastery versus goals aimed more externally at competitive superiority. Individuals motivated by task mastery challenge themselves in ways that develop their skills, whereas individuals motivated primarily by outcomes may choose goals or activities that are easily accomplished to avoid threats to self-esteem. It is unclear whether task mastery goals lead to better performance than do outcome-oriented goals, but athletes who employ task mastery goals do seem to experience greater increases in self-efficacy than do those who use outcome-oriented goals.

#### **4.2. Social Facilitation**

Another motivational source in athletics lies in social facilitation. Social facilitation is a process whereby performance is increased due to the presence of others. It is one of the oldest areas of sport psychological research. Better performance in the presence of competitors or spectators has been explained by classical psychological concepts such as drive theory. Spectators tend to increase arousal, and drive theory proposes that arousal increases the performance of dominant responses. This is a desirable state of affairs when dominant responses are correct and well learned, as in familiar tasks. But the presence of spectators may hamper performance when dominant responses are incorrect or underlearned, as in novel or unfamiliar tasks. Therefore, the presence of others may or may not facilitate performance. Not surprisingly, research on drive theory and social facilitation in sport has produced some inconsistent results. Social facilitation effects appear to vary across real-life sports, where motor behaviors are more complex than those studied in laboratories.

Interviews with elite athletes suggest that the ability to draw inspiration from a crowd or the presence of opponents may be highly individualized. For some, the attention may be energizing. Such individuals generally benefit from some level of awareness that others are present and watching. In this case, the crowd may become a task-relevant cue in competition, and athletes may develop the ability to capitalize on social facilitation by attending to the crowd. Performing in front of individuals in practice may help to develop the skill of drawing energy from a crowd as well as provide practice in partitioning attention among motor

performance, internal cues, and the external importance of the crowd. For other athletes, knowledge that they are the center of attention becomes anxiety provoking. Such athletes may benefit more from skills training that minimizes the scrutiny felt from the presence of others.

### **5. SOCIAL AND INTERPERSONAL SKILLS**

A broad array of social and interpersonal skills may be brought to bear in athletics. Smooth and effective interactions with others—not just teammates but also coaches, officials, and even opponents—may facilitate higher degrees of performance. This section considers two areas that are relevant to interpersonal skills: communication and interpersonal and intragroup anxiety in a team context.

#### **5.1. Communication**

Being able to communicate in sport is often critical to success. This is true even in the case of individual sports, where an individual must still interact with officials, coaches, and opponents. In general, performance will be maximized to the extent that an individual can effectively encode and deliver feedback and messages to teammates, coaches, and officials as well as decode and incorporate received information.

In team sports, delivering encouragement and technical feedback to teammates is often essential. Sending understandable messages at appropriate times permits teammates to correct deficiencies or other performance impediments, facilitating better team play. Being able to communicate openly and effectively with coaches during training and practice also permits the ready flow and interchange of necessary (and often vital) information about an athlete's functioning across biopsychosocial states. Finally, communicating appropriately with officials during the heat of competition is a particularly difficult, but nonetheless indispensable, skill.

Communication also involves receiving messages. Often, feedback may be delivered from teammates, coaches, opponents, or officials in an emotional or insulting manner. The ability to evaluate feedback for its truthfulness and usefulness, and then to glean helpful elements while rejecting unhelpful parts, is often critical to improving sport performance. Athletes who "can't take criticism," or who incorporate all critiques

of their performance without discrimination, may be missing out on valuable opportunities to benefit from appropriate guidance.

In addition to articulating thoughts clearly, communication patterns also affect the emotional climate. Belittling a sensitive or anger-prone teammate might not be an effective performance enhancer for the team, and communicating with a high level of tension in an already pressurized competition might serve only to increase the anxiety of some teammates. In some instances, the contagious effect of high-level emotions may have an overall positive impact on team play. In other cases, it may be disastrous, as in the case of raising individuals' anxiety levels.

### **5.2. Interpersonal and Intragroup Anxiety**

The uniquely social nature of sport may serve to heighten anxiety. Although individual traits or dispositions contribute to anxiety, Hanin described two types of anxiety that are also partially created by the team or interpersonal environment: interpersonal and intragroup anxiety. The two are related and, whether operating alone or in conjunction, may exert deleterious effects on individual and team performance.

In interacting with a training partner, a coach, or some other individual, interpersonal anxiety may arise. Interpersonal anxiety is typically a state in which an athlete fears negative evaluation from, is distrustful of, or is uncomfortable with the partner of an interaction. This may occur commonly with an unfamiliar or new coach or in communicating with a higher status team member. There is some indication that an athlete may feel less or no interpersonal anxiety in interacting with a cohort of equal or lower status because social comparison works favorably for the athlete in such cases.

A related form of anxiety, intragroup anxiety, may also exist in team contexts. In intragroup anxiety, the broader social milieu of the team itself affects an athlete's subjective anxiety. Intragroup anxiety may arise as a result of things such as rejection or poor treatment by teammates, a lack of group cohesiveness or poor team atmosphere, a lack of experience, and poor self-confidence. Athletes who are naturally more anxious possess greater degrees of negative emotionality and are more concerned with maintaining socially desirable appearances and behaviors. They may be more prone to intragroup anxiety. Lower status group members may be particularly susceptible to this type of anxiety.

The anxiety management skills discussed in the next section apply to interpersonal and intragroup anxiety. But in addition, trainers, teammates, and coaches may play important roles in regulating the interpersonal and intragroup anxiety of individual athletes. To the extent that familiarity and comfort can be enhanced between the members of a dyad, interpersonal anxiety may decrease and performance may increase. In a broader team setting, a mature group climate and healthy group dynamics may be encouraged by perceptive and effective coaches.

## **6. TRAINING PSYCHOLOGICAL SKILLS**

Some evidence suggests that arousal regulation, behavioral practice strategies, attention and concentration, mental rehearsal, self-instruction, appropriate goal setting, and communication skills can be trained. However, in contrast to conventional psychotherapy where accountability has long been an issue, research on psychological skills training in sport is still in its infancy. Much of what is known about psychological skills has also been gleaned from studies on elite athletes, and the techniques and strategies observed in elite athletes may or may not be effective in less skilled athletes. Thus, the consequences of attempting to teach a general canon of psychological skills in a "one size fits all" approach may range from success for some athletes, to wasted effort for other athletes, to perhaps even detrimental effects for some individuals. Therefore, overzealous or indiscriminate training of all skills with all individuals is discouraged in favor of interventions that take into account an individual athlete's personal characteristics, needs, sport, and levels of physical, cognitive, and emotional development. With this caveat in mind, this section reviews some commonly used protocols and techniques for developing psychological skills.

### **6.1. Arousal and Anxiety Regulation**

Because a large component of anxiety is physiological overarousal, the following considerations are relevant to the training of both anxiety and physiological arousal regulation. For anxious and overaroused individuals, various forms of relaxation training are the interventions of choice.

One form of relaxation training is called progressive muscle relaxation. Progressive muscle relaxation involves systematic decrements in the tension of the body's skeletal muscles. Typically, an individual flexes and then relaxes one muscle group at a time, often using visual imagery to facilitate the effect. Breathing becomes a focus, with special attention paid to a slow and steady rate of breath and deep inhalations and exhalations. When executed properly, these processes slow heart rate and decrease neuronal firing in the skeletal muscles, with parallel arousal decrements in the sympathetic division of the autonomic nervous system.

The gist of progressive muscle relaxation may be learned in a single session with a skilled facilitator, but mastery requires regular practice. A script or tape-recording may be employed to guide the progressive muscle relaxation, often beginning with arm and shoulder muscles and proceeding downward (although other variations are common). Employing tapes enables coaches, trainers, and athletes to institute relaxation practice without the presence of a trained facilitator, allowing relaxation sessions at any time or location and at less expense than bringing in a consultant. A variety of relaxation tapes are commercially available.

Another training mechanism for arousal regulation is biofeedback. Biofeedback involves electronic monitoring of an athlete's physiological signs, especially heart rate, pulse, galvanic skin response, and muscle tension. This information is subsequently used to guide progressive relaxation. With biofeedback data, problem areas in the relaxation process can be pinpointed and targeted with precision. After biofeedback training, an athlete often gains control over physiological processes that were previously beyond his or her ability to regulate. This permits an athlete to relax with greater ease and efficiency.

Biofeedback training generally occurs over multiple sessions. Initially, an athlete is introduced to the equipment and basic principles of biofeedback. He or she is shown, in visual terms, the effect of small bodily changes on heart rate and other indicators of arousal. Initially, an athlete may find it difficult to exert control over bodily processes that occur automatically. However, over successive practice sessions in a biofeedback laboratory, he or she learns how the manipulation of subtle and previously ignored physiological operations may affect levels of arousal and felt anxiety. Eventually, the biofeedback equipment is no longer needed, and an athlete can effect the necessary alterations to his or her arousal level in other contexts.

Meditation training may also reduce arousal and anxiety. Meditation involves relaxation techniques such as deep breathing but adds an element of mental focus, usually in the form of a thought or an image that is repeated or on which attention is focused. In this way, meditation may also serve to train mental imagery abilities and attention and concentration processes while also regulating physiological arousal. In learning meditation, an individual focuses on relaxing physically through deep breathing and gently releases competing thoughts, mental chatter, and worries while constantly returning to a "centering" image. This is best learned from either an experienced meditator or the use of tapes.

In developing meditation skills, regular practice is imperative, and patience and gentleness are critical, because it is extremely difficult to detach from "mental chatter." Finding a regular time and place for meditation practice often augments the process, but it is important that an athlete experience the setting for meditation as comfortable and peaceful. The duration of meditation practice is often specific to individual tolerance levels, but early practice sessions generally should be kept to shorter lengths (10–20 minutes for most beginners).

Other interventions permitting arousal and anxiety regulation may include prayer and hypnosis. Although one does not typically think of prayer as a sport performance skill, prayer may facilitate relaxation and focus attention. It may also provide an athlete with a calming reassurance that a divine force is operating on his or her behalf. Similarly, hypnotic states are defined by nearly total physical relaxation. However, not all individuals are easily hypnotizable, and trained hypnotists are not always available. Thus, unless an individual has access to a certified instructor, it is a less practical form of intervention and not an easily acquirable skill, although some individuals are able to learn to hypnotize themselves.

Finally, arousal regulation also involves "psyching up." Athletes may be trained to drastically increase arousal levels over a short time. This type of arousal regulation training has been less researched, but control may be exerted through physical channels (e.g., slapping oneself, shaking, jumping up and down) or through cognitive mechanisms (e.g., imagining anger-producing things, planning and anticipating challenges during competition, focusing on context cues such as the crowd, engaging in final rituals that have been repeatedly paired with high-arousal states in the past). Habitually underaroused athletes must explore

a variety of last-minute psyching strategies and must retain and practice those that seem to be most advantageous.

## 6.2. Part Practice

Behavioral practice skills and motor learning can be facilitated through part practice. In part practice, an athlete can focus greater attention and concentration resources on simpler components until they become automatized, thereby freeing up resources for execution of the whole task. This discussion considers three forms of part practice: segmentation, fractionation, and oversimplification.

In segmentation, a complex task is divided into a number of constituent dimensions, which are then trained separately. For instance, a gymnast training a routine on the balance beam may separately train the mount, actions performed on the beam, and the dismount. Often, repetition of task segments may involve graduated acquisition to more complex combinations of segments. In this way, task parts can be “chained” together, usually from the beginning pieces of a complex behavior to the end (forward chaining) or from the end pieces of a complex sequence to the beginning (backward chaining). The degree to which facility in task segments transfers to the whole task may be determined by whether the segmented practice involves integrated control of the “links” in the behavior chain. If an athlete does not learn necessary transitions between segments, the whole movement might not improve substantially even in the face of dramatic improvements in the segments.

In fractionation, an athlete trains separate pieces of a complex action that normally occur at the same time. Results regarding the effectiveness of this technique have been somewhat mixed, but it seems that for tasks in which component behaviors are highly integrated because they occur simultaneously, practicing various parts in isolation often fails to develop coordination among them. Research investigating fractionation does suggest that it may be useful for complex tasks where subcomponents must be randomly combined, that is, where there is not a single template or control/coordination process for the task. For instance, in a video game, multiple fractionated skills (e.g., making a figure run, jump, or fire a gun) may be called on, but in random combinations.

A final form of part practice is simplification. Simplification generally takes the form of training easy versions of complex motor behaviors as

precursors and building blocks. A juggler may first learn to juggle two scarves, then progress to three baseballs, and then finally to five or more torches or knives. In this example, the degree of difficulty (as well as the danger of potential mistakes) increases incrementally. Again, research on simplification has been somewhat mixed. Although it seems to be an effective teaching technique, there may be a point at which simplification of a complex motor behavior alters the essential nature of a task. At this point, it is unlikely that skill acquired at the simplified task would generalize to the more complex task. However, if the fundamental control processes of the complex task are not altered to a great degree, simplification may facilitate motor skill acquisition in some athletes.

## 6.3. Improving Attention and Concentration

Attention and concentration skills are malleable and can be learned and improved through repeated practice. Several exercises have been proposed, including scanning number tables and staring intently at a sport-related object. However, the merit of such training techniques lacks systematic substantiation. Nonetheless, one technique that has gained some credence is adversity training.

In adversity training, the specific alterations to practice sessions are less important than their general effect: to create a training atmosphere in which distractors, unconventional circumstances, situations, and/or environmental interferences exist. By training in the presence of such potential disruptions, athletes develop a more robust ability to focus their attention and resources on task-relevant cues and screen out potential impediments that arise randomly in the competition environment. By treating practice sessions of this sort “like the real thing,” athletes learn to screen out intrusive and novel stimuli as a matter of course so that if such nuisances appear in actual competition, they will be less disruptive. Although little systematic research has investigated adversity training, it is a manifestation of the well-established principle of state-dependent learning and anecdotal support is plentiful.

Another way in which divided attention skills may be developed is through dual-task practice. In sport, athletes are often required to perform multiple tasks simultaneously, and this requires switching skills to allocate attention between the different tasks.

Training multiple tasks simultaneously can help athletes to develop attention allocation strategies and increase attentional resources. For instance, a basketball player's dribbling skills may be improved by practicing dribbling two balls at once while looking ahead. In this way, dual-task practice facilitates automatization of component movements, thereby freeing an athlete's attention resources for allocation elsewhere.

However, it is important that dual-task practice be conducted with sport-specific tasks. Developing divided attention on an unrelated array of activities might not generalize to the idiosyncratic demands of a particular sport. Training attention skills through dual-task practice should also take into account an individual athlete's attention resources and skills. Individuals vary widely in their reservoir of attentional resources and in their skills at allocating them. But even for expert performers, continued overload of attentional resources through graduated dual-task or multitask practice may permit ongoing increases in attention skills.

Finally, a variety of other strategies have been proposed for improving attention and concentration skills. Preperformance routines, which generally consist of a series of motoric movements and/or accompanying thought and self-instruction, may serve to focus concentration. Centering exercises and mental imagery may also facilitate concentration on external task demands indirectly through augmenting internal focusing skills. Regulating arousal to optimal levels has salutary effects on attention or concentration. And although the ability to enter flow voluntarily seems to vary across individuals, certain strategies may help individuals to induce flow states. Setting miniature goals in practice sessions, such as executing a tennis serve with perfect form five times in a row, have been helpful at inducing flow in athletes. Transforming practice sessions into "games" may also help. For instance, soccer players practicing "juggling" the ball with their feet may elect to play a juggling game in which the player who can juggle the longest gets to skip a mandatory conditioning run.

#### **6.4. Self-Monitoring and Self-Instruction Training**

Self-monitoring may be facilitated through recording successful and unsuccessful behaviors in practice and competition. These records may then be reviewed either individually or with a teammate, coach, or trainer to examine patterns in an athlete's performance. With a

log of successes and failures (and the assistance of another pair of eyes), athletes may begin to see performance patterns that can help them to pinpoint specific areas that need closer attention. Videotaping and reviewing performance may also facilitate self-monitoring skills. For instance, an athlete may realize that a particular type of performance mistake is always preceded by a motoric or kinesthetic flaw that can then be expunged during subsequent practice sessions.

Self-instruction may be taught through the use of verbal scripts. Scripts serve the purpose of providing a format, or template, of helpful statements and important reminders that an athlete can implement before, during, and after competitions. Self-instruction scripts ideally provide the athlete with a ready storehouse of important reminders and encouragers that, with practice, will become automatic. When well practiced, self-instruction may provide instantaneous correction, redirection, and reassurance in the midst of competition.

Cue words also may assist in self-instruction. Salient cue words, when implemented regularly, will serve to convey immediate corrective feedback to athletes when performance errors occur or are in danger of occurring. Cue words should be personally selected by an athlete but should be relevant and bear some important meaning about the activity being performed. For instance, a cricket bowler may employ the phrase "follow through" to increase the accuracy of his or her pitches. Common cue words may also be shared with teammates or coaches, who may assist athletes in acquiring and regularly implementing instructive words during performance.

#### **6.5. Mental Imagery and Mental Rehearsal Training**

Mental imagery is often conducted in states of relaxation, although it may be practiced anywhere. In general, however, the absence of distractions in relaxed or meditative states permits more efficacious use of mental imagery. Guided imagery is best learned with the help of another knowledgeable coach, trainer, or sport psychologist. Usually conducted during a state of relaxation, a series of relevant sensations are described in detail, invoking mental representations of sights, sounds, smells, and so forth. The images may proceed from an external perspective (i.e., a "camera reel" of a performer in action) or from an internal perspective (i.e., from the performer's view). Again, the content and use of imagery should be dictated by the individual athlete's needs. For an athlete who has trouble "getting

in the zone” prior to a competition, bodily sensations of tension and anticipation may be invoked during guided imagery practice, along with a “tour” of the precompetition environment and preparations, ending with an image of the athlete fully aroused, attentive, and eager to execute the actions that are necessary for competition.

Mental rehearsal of athletic performance may be conducted in nearly limitless ways. Some sport psychologists have developed structured training programs to help athletes improve mental rehearsal. Researchers have recently proposed a model that may be useful in the training and practice of imagery. Athletes are encouraged to consider where (e.g., during practice, before a competition) and why (e.g., trying to build general confidence, trying to improve a specific skill) they employ imagery. Athletes are trained not only to select places and times to use imagery that is consistent with what they are trying to achieve but also to use different types of imagery for general or specific goals. Imagery may be either specific (e.g., imagining the successful performance of a particular skill) or general (e.g., imagining oneself competing with confidence). The bulk of research suggests that imagery practice enhances specific skills and, like physical practice, should be distributed across time.

Visualization may be trained by using photos or videos as guides. Because real-life visual stimuli tend to be detailed and vivid, using them as models may facilitate visualization in beginners. Training may proceed from looking at sport-specific pictures or films, concentrating and noticing every detail, to engaging in mental practice wherein the model is recalled with as much vividness and detail as possible. If visual models of an athlete’s own top performances are available, these may facilitate transition into personal visualization in practice. The content of visualization practice should match an athlete’s goals. At a generic level, an athlete might visualize himself or herself preparing for, engaged in, and winning a competition. But at an individually specific level, an athlete might train visualization skills around a particularly difficult maneuver or motor sequence in performance. If motivation or confidence is an issue, visualization of goal states, successful outcomes, and mistake recovery may be helpful.

## 6.6. Acquiring Emotion Regulation Skills

Some techniques for reducing anxiety were addressed previously in the discussion of arousal and anxiety

regulation. However, in addition to relaxation exercises, a variety of other strategies and interventions exist for regulating distressing emotions. These can often be developed using techniques found in most conventional psychotherapies. Although space constraints prohibit extensive coverage of these methods here, a variety of general self-help books provide thorough and accessible discussions of emotional regulation strategies.

One way in which emotional distress may be remediated involves examining and challenging the thoughts associated with negative emotions. This requires that an athlete first develop an awareness of the sort of thinking that accompanies negative emotions. For instance, in the introductory example, Andrea experiences a surge of anxiety. Such a rush of nervous feelings may very well be accompanied by the thought, “If I don’t perform well here, my career will be ruined.” Naturally, this sort of absolutistic thinking serves only to heighten feelings of anxiety.

Identifying the self-defeating, irrational, catastrophic, and/or distorted thoughts behind negative emotions can often be difficult and frustrating for athletes learning emotional regulation because such thoughts may be so automatic that they are hard to detect. The use of logs, in which athletes track emotionally distressing moments and the thoughts beneath them, often helps to illustrate the connection between thinking and emotion and illuminates patterns. When athletes develop such awareness, they know where to intervene when emotions turn problematic.

Once the thoughts driving negative emotional responses have been reliably identified, the next step in emotional regulation is to begin systematically challenging such thoughts. Athletes can learn to “act as their own defense attorneys” in response to self-derogating thoughts or self-anger and to question the likelihood of fictional finalisms or the validity of hidden assumptions. However, this process is often effortful and requires extensive practice because negativistic thinking patterns tend to be quite entrenched. It is important to note that identifying and challenging self-defeating thoughts is likely to be an unfamiliar experience to an athlete learning emotional regulation. For this reason, the athlete might wish to enlist the aid of a personal therapist, counselor, or trained sport psychologist.

Self-confidence may also be trained through techniques such as imagery and mental rehearsal of confident performance, monitoring and replacement of self-defeating thoughts, seeking out mastery experiences,

and validation and encouragement from others. However, self-confidence in athletics may also be rooted to deeper feelings of personal inadequacy. When this is the case, personal counseling or psychotherapy is likely to be useful. In fact, for athletes experiencing emotion regulation problems at a deeper level and/or extending outside of competition, personal counseling or psychotherapy is highly recommended not just to improve sport performance but also to improve overall happiness and quality of life in general.

### **6.7. Training Appropriate Goal Setting**

The consequences of inappropriate goal setting can be more deleterious to an athlete's motivation than training bereft of goals because the wrong type or set of goals can set up an athlete for failure. Research suggests various parameters that are most conducive to goal attainment, although goal setting should also be considered from an individualized perspective due to the wide variation in individual abilities.

In general, athletes should be encouraged to select a series of easy proximal goals early on (e.g., accomplishing small things that are close and immediate). Easy proximal goals offer the possibility of quick feedback and mastery experiences, which serve to enhance motivation. The arrangement of proximal goals may also form a chain toward some ultimate goal. For instance, a team may institute night-by-night goals during the regular season to reach the league championship. However, it is important to avoid an overly rapid or unrealistic series of proximal goals. Individuals or teams may be prone to set less realistic goals after successful achievement of early easy goals.

Athletes learning to set goals should also be encouraged to select specific goals. Specific goals permit achievement in precise areas relevant to success. For instance, rather than setting a goal to improve one's overall strength, a weightlifter may set a specific goal of improving his or her maximum clean and jerk by 2 kilograms and his or her maximal snatch by 1 kilogram. If the weightlifter also invokes the principle of proximal goal setting, he or she may resolve to achieve this at the end of 4 weeks of training sessions. Specific goals render large tasks (e.g., winning a competition) less imposing or intimidating by breaking down a desired end state into components.

Goal-setting skills are best developed by encouraging athletes to use mastery goals as well. Athletes

should be encouraged at first to "compete against themselves" by improving past performance or meeting specific competence levels independent of competition. Developing athletes who set competitive goals may rob themselves of a sense of personal control in pursuit of their desired ends because competitive goals tend to be at least partly contingent on other individuals. The achievement of mastery goals, in contrast, is generally more within the direct control of the athlete himself or herself. Mastery goals, in turn, promote the development of self-efficacy more readily than do competitive goals.

An athlete may also wish to begin with individually chosen goals. In cases where the athlete is uncertain of his or her abilities or appropriate first goals, it may be helpful to develop goals initially with a coach, trainer, or sport psychologist in a collaborative fashion. In such instances, goals that are personally meaningful to a given athlete may serve as the foundation for early development. The mastery experiences attached to maintaining individually salient goals may be more motivating than the sense of accomplishment felt at meeting institutional or external goals, although an athlete who is extrinsically oriented may also benefit from external goals.

Finally, it is important to note that these considerations apply primarily to the training of goal-setting skills in inexperienced athletes. Elite athletes may benefit just as much, or in some cases more, from distal and competitive goals. But the use of proximate, specific, mastery, and individual principles can be highly recommended to less advanced athletes wishing to develop goal-setting skills.

### **6.8. Enhancing Communication**

Hanin recommended three general strategies for enhancing communication skills in a team environment. First, if communication problems are a function of the specific subset of individuals involved, such as having an aggressive or belligerent athlete in the starting lineup, one may alter the composition of the unit. Communication patterns among players are likely to change accordingly. However, this is only a temporary solution and does not remediate the root of communication problems if it resides in one or more particular individuals.

If certain activities, tasks, or situations tend to produce communication difficulties, one may redefine, alter, or substitute tasks that do not evoke such difficulties. For instance, if one teammate always has trouble performing

in a certain drill and this leads to negative interactions with other players, an alternative drill that accomplishes the same thing may be instituted. Naturally in some situations, such as rehearsing a particular play that is indispensable to team performance, one cannot alter the situation or group task without compromising the integrity of practice. In such instances, communication problems may be dealt with by managing players' communication at a behavioral level.

At the level of communication behaviors, athletes may be taught to monitor and regulate the content, delivery, and even frequency of their messages. Through observation, instruction, and discussion, they may develop specific styles of communication for specific teammates. Negative feedback may be sandwiched between positive comments, accusing or derogatory content can be unlearned in favor of appropriate and task-relevant comments, and athletes can be taught active listening skills so that teammates feel that they are heard.

At both the individual and team levels, communication skills may be practiced through role-plays outside of practice. As new communication patterns are acquired, athletes should be encouraged to implement them during practice and to observe their own and others' reactions to these changes. The ability of an athlete or a team to transfer new communication skills to competition should then be monitored closely, with appropriate discussion of both positive and negative exchanges and necessary adjustments.

Whether communication deficits are a function of skill deficits or situations, facilitating communication skills in athletes requires close observation of problematic communication patterns, including their antecedents and effects. With such knowledge, coaches and/or other helpers may meet with individual team members, as well as the team itself, to discuss specific problems. It is also important to provide feedback on instances of effective communication as well so that appropriate interaction patterns can be reinforced.

## 7. CONCLUSIONS

Research on elite athletes reveals that rather than a general "athletic personality," psychological skills are more important determiners of athletic success. These skills are mostly acquired rather than innate, and they span the bodily/behavioral, cognitive, emotional, motivational, and interpersonal domains. Bodily/Behavioral control skills encompass arousal regulation and behavioral practice of motor skills. Within the cognitive realm,

mental rehearsal, attention and concentration, and self-monitoring and self-instruction skills are critical to athletic success. Emotional balance, controlling anxiety, and maintaining self-confidence are necessary. Motivational skills involve setting appropriate goals and capitalizing on social facilitation. Finally, interpersonal skills involve effective communication patterns and management of emotions within a team or group context.

These skills are trainable, and although research on skills training is still in its infancy, a variety of strategies and techniques exist to facilitate skill acquisition. Training of skills should proceed in a supervised environment and in a manner that is appropriate to an athlete's developmental level. Typically, developing skills requires investment and practice as well. Some techniques, particularly unfamiliar ones such as relaxation and meditation, may appear to be quite foreign and may frustrate beginners. But with repeated practice, they can be nurtured, with resulting improvement in performance.

Finally, although training of skills may be undertaken in a group or team format, the effective inculcation of psychological skills proceeds best at an individual level. Although some skill sets may be necessary for a class of athletes or team in general, individuals will often vary in the style or degree to which they implement a skill. In the end, efforts should be made to match skills training programs to individual athletes' strengths, weaknesses, desires, and capabilities. Attention to each athlete's uniqueness is ultimately the most important factor in enhancing his or her performance.

## See Also the Following Articles

Anxiety and Optimal Athletic Performance ■ Arousal in Sport ■ Attention and Concentration Training in Sport ■ Goal Setting and Achievement Motivation in Sport ■ Group Dynamics in Sport ■ Intrinsic and Extrinsic Motivation in Sport

## Further Reading

- Hanin, Y. L. (1997). Emotions and athletic performance: Individualized zones of optimal functioning model. *European Yearbook of Sport Psychology*, 1, 29–72.
- Hanin, Y. L. (2000). *Emotions in sport*. Champaign, IL: Human Kinetics.
- Mahoney, M. J. (1989). Sport psychology. In I. S. Cohen (Ed.), *The G. Stanley Hall lecture series* (Vol. 9, pp. 97–134). Washington, DC: American Psychological Association.
- Mahoney, M. J. (2002). *Relaxing, reaching peaks, and learning process*. Plainfield, IL: Kinder Path Press.

- Mahoney, M. J. (2003). *Constructive psychotherapy: A practical guide*. New York: Guilford.
- Moran, A. P. (1996). *The psychology of concentration in sport*. London: Taylor & Francis.
- Orlick, T. (1990). *In pursuit of excellence: How to win in sport and life through mental training*. Champaign, IL: Leisure Press.
- Singer, R. N., Hausenblas, H. A., & Janelle, C. M. (Eds.). (2001). *Handbook of sport psychology* (2nd ed.). New York: John Wiley.
- Van Raalte, J. L., & Brewer, B. W. (Eds.). (2002). *Exploring sport and exercise psychology* (2nd ed.). Washington, DC: American Psychological Association.



# Psychology and the Law, Overview

Roy S. Malpass

University of Texas at El Paso, El Paso, Texas, USA

Jane Goodman-Delahunty

University of New South Wales, Sydney, Australia

1. Introduction
  2. Important Attributes of the Field
  3. Articles in This Section
- Further Reading

## GLOSSARY

**adjudicative competence** Functional legal abilities that a criminal defendant must have in sufficient degree in order for the adjudicatory process to go forward. Also called competence to stand trial, competence to proceed, and fitness to proceed.

**best interests of the child doctrine** The philosophy, now incorporated into law in all 50 states, that in making decisions concerning custodial placement and parenting plans, the interests of the children should supersede those of the parents and/or others involved in the dispute.

**blank lineup (target absent lineup)** A lineup in which all the members are known to be innocent.

**cognitive interview** Forensic tool that comprises a series of memory retrieval techniques designed to increase the amount of information that can be obtained from a witness.

**direct questions** Yes/no or short-answer questions that probe for particular details about an experienced or witnessed event; questions can range from nonleading to suggestive, depending on how the question is phrased and the type of information included in the question.

**effective size** The number of plausible members in a lineup.

**encoding specificity principle** States that if cues that were present at encoding are also present at retrieval, recall of stored information is more likely.

**estimator variables** Factors in eyewitness situations whose influence can only be estimated and are not under the control of the justice system.

**ethical interviewing** A police interview technique, based upon three principles, designed to encourage suspects to tell their side of the story. The use of deceit is forbidden, and psychological pressure is not used.

**evaluation of comparative custodial suitability** A forensic psychological evaluation in which the primary focus is the relative strengths and deficiencies in the ability of parents to parent the specific children whose custodial placement is at issue.

**evidence-driven deliberation** Deliberation in which jurors engage in general discussion of evidence in the case prior to voting, in contrast to verdict-driven deliberation, in which voting typically precedes evidence discussion.

**forensic assessment instrument (FAI)** An assessment test, tool, or device designed to measure or to inform a clinical judgment about specific functional legal abilities required in a given legal context.

**forensic mental health assessment** The evaluation of defendants or litigants, usually conducted by mental health professionals such as psychologists, psychiatrists, or social workers, that is performed to assist the court in making better-informed decisions or to help the attorney present evidence relevant to his or her case.

**free-recall questions** Vague, general prompts requesting for a narrative description of a prior events; prompts provide minimal cues or information about what should be recalled.

**lineup bias** A lineup constructed in a manner that leads the suspect to stand out from the other lineup members.

**mental state at the time of an offense (MSO)** A general term inclusive of various legally recognized mental states

- (e.g., insanity, diminished capacity, automatism) that provide the bases for a defense to culpability for criminal behavior.
- mock witness** A person who is not a witness to a crime but who is asked to identify a perpetrator from a lineup based on another source of information, such as a verbal description.
- post-event (mis)information** Events that occur after a crime that influence eyewitness memories of the crime.
- Reid technique** A police interrogation technique, consisting of nine steps, designed to encourage reluctant suspects to talk during police interviews. The technique allows interrogators to use deceit and to build up psychological pressure.
- restorative justice** Legal proceedings that aim to make the victim and community whole following an offender's conduct, and to reintegrate the offender into the community, if appropriate.
- scientific jury selection** The use of scientific techniques by jury consultants to determine the individual demographic and attitudinal characteristics that are linked to views of a case. Techniques include community surveys, mock trials, focus groups, and in-court selection.
- script** A general knowledge structure that represents information from a class of similar events; a description of the typical, common features of repeated events.
- showup** The presentation of a single suspect to a witness in order to determine if that suspect is the perpetrator.
- source amnesia** The inability to recall the origin of the information that comes to mind when attempting to recall an event.
- story model** Theory of juror decision making in which jurors arrange evidence in the form of a narrative or story, and match the story to the closest verdict alternative.
- structured interviews** Scripted interview prompts and questions that are designed to obtain as much narrative information from children and be as minimally leading as possible.
- suggestibility** The act or process of impressing something (an idea, attitude, or desired action) on the mind of another.
- system variables** Factors that may influence eyewitness memory and that are under the control of the criminal justice system.
- therapeutic jurisprudence** Legal practices intended as therapeutic interventions for the benefit of the litigant.
- therapeutic role** Within the context of custody evaluations, perspective and actions that are consistent with one's obligation to assist those with whom one is professionally interacting.
- verbal overshadowing effect** When the act of verbally describing a perpetrator decreases the witness's ability to accurately identify a perpetrator from a lineup.
- voir dire** Questioning process during jury selection.

Psychology and law, or forensic psychology, is a branch of applied psychology, with three primary dimensions: research, practice, and policy implications. Major accomplishments in psychology and law are reviewed in this article, drawing on empirical research by

experimental and clinical psychologists with several distinct forensic populations, e.g., juveniles, offenders, witnesses, and jurors. Experimental psychological contributions on detection of deception, eyewitness memory, and jury decision making are outlined. Clinical research and practices regarding forensic mental health assessments are described, e.g., of legal competency in criminal, civil, and family law cases. Roles of psychologists in therapeutic versus forensic settings are distinguished, and standards for providing relevant and reliable expert evidence are described.

## 1. INTRODUCTION

If experimental psychology is to enter into its period of practical service, it cannot be a question of simply using the ready-made results for ends which were not in view during the experiments. What is needed is to adjust research to the practical problems themselves and thus, for instance, when education is in question, to start psychological experiments directly from educational problems. Applied Psychology will then become an independent experimental science which stands related to the ordinary experimental psychology as engineering to physics.

The time for such Applied Psychology is surely near, and work has been started from most various sides. Those fields of practical life which come first in question may be said to be education, medicine, art, economics and law.

This bit of advocacy was written by Hugo Muensterberg in 1907, and makes clear that both the vision and the empirical research program on which the field is founded were present a century ago. Research at the time, and much of the work since, has taken place in the areas of perception and memory, as well as abilities, intellectual performance and social relations, areas in the list of problems related by Muensterberg. All of these areas are treated in the present section of this encyclopedia, and while they would all be recognized by turn-of-the-19th-century psychologists, their elaboration since that time has been considerable. Legal psychology has grown in many ways, and has multiple foci.

## 2. IMPORTANT ATTRIBUTES OF THE FIELD

There are certain attributes of psychology that set it apart from other fields with relationships with law

(e.g., social work, psychiatry). These attributes form common themes across the articles in this section. The field of psychology and law displays an interesting diversity of approaches to its application, which reflects the full diversity of the underlying field. Its approach is at base scientific, and in the analysis of policy, practice, and evaluation, the goals are theoretical and empirical clarity. The orientation is, for the most part, one of engaging problems at a practical level, using available theory and new research to develop pragmatic responses tailored to the demands of situation and context—just as Muensterberg foresaw.

The field of psychology and law exhibits a distinctive scientific character. The field is committed to testing standards that are based in measurement theory and developed according to scientific standards that distinguish this professional enterprise from that of others, e.g., social workers, psychiatrists. Validated measures and tests exist for some topics (e.g., competence to stand trial and competence to consent to treatment), but are not yet developed for others. The cutting-edge developments in the field are precisely the formation of new tests and measures, new empirically validated techniques and procedures, and policies founded in empirical research.

The potential contributions of psychology to the various demands of law are by no means fully realized. As psychologists embark on analysis of many domains in the psychology and law interface, alternative approaches and explanatory hypotheses are framed and tested using multiple data sources—a part of the fields of history of philosophy of science and methodology.

Psychology has a special role of assisting courts in making retrospective assessments of what occurred and in identifying lies and distortions (i.e., in serving justice). The contexts in which these needs appear are vast. Witnesses to events of all kinds, from bank robberies to traffic accidents, are relied upon to provide accurate information for legal processes. Those alleging the occurrence of a crime, either as a witness or as a victim, are evaluated for accuracy and for incentives to distort. Report of memory for events is a large part of many fact-finding activities in law. Psychology's special role emerges from the fact that records of events and the accounts of them come from human memory and the perceptual and cognitive processes that both record and report them—both selectively. The report of a police officer, for example, comes from memory, and even when it is written contemporarily with the event, the selectivity typical

of human attention requires understanding and interpretation, especially if the officer's report conflicts with the reports of others. Psychology is uniquely positioned to address memory for events and other forms of experience, as well as intentional distortion and deception. No other field of study is so focused on these processes.

Memory for narrative integration of information, making judgments, and studying the effects of personal history, attitudes, and belief systems are all aspects of legal processes. Some take a central role and assume great importance. Jury decision making is a strong traditional area of psychology and law. The integration of new information into existing attitude and belief systems in the context of whatever constraints are placed on judgment by the court is an important area of understanding for scholars to pursue. Research in this area has the potential for important contributions to our understanding of why long-term trends of injustice (e.g., racial discrimination in all areas of opportunity and legal protection, gender inequality in ownership and employment) were able to continue and how and why change occurs.

Criminal suspects and litigants have incentives to fabricate, mislead, exaggerate, and confabulate. These incentives and the difficulties they present to fact finders extend through many areas, including forensic mental health assessment. As noted by Heilbrun and Landers in their article on assessment in this section,

Some individuals evaluated through FMHA respond to questions openly and honestly. Others exaggerate or fabricate symptoms and problems. Yet others minimize or deny symptoms that are genuinely experienced. Response style is a term used to describe how an individual reports his or her own symptoms, behaviors, and problems, and is used to classify an individual's responding as (1) reliable, (2) exaggerated/malingering, (3) defensive, or (4) uncooperative. Self-reported information from an individual who is not reliable will have significant inaccuracies. This principle emphasizes the value of third party information (records and interviews of collateral observers) in both detecting and correcting these inaccuracies.

The role of psychologists in a legal context is often structured very differently from that of psychologists in other contexts. For example, there is an important distinction between therapeutic and forensic roles. Ten differences elaborated by Heilbrun include (1) the purpose for which the evaluation is conducted, (2) the nature of the relationship between the evaluator and the individual being evaluated, (3) the need for a

formal notification of purpose, (4) who is being served, (5) the nature of the standard(s) applied in the evaluation, (6) the sources of data considered, (7) the response style of the individual being evaluated, (8) the need to clarify the reasoning and limits of knowledge, (9) the nature of the written report, and (10) the expectation of expert testimony.

Heilbrun and Lander remind us that it is only in the past 50 years that courts have been using psychologists as opposed to other mental health professionals to provide information on assessments. Recognition of the role of psychology is still lagging in some communities, and many courts still require a medical degree from the professionals whose opinions they seek. It is only in the last 25 years that expert testimony by experimental psychologists has been widely used in court in matters involving memory and eyewitness identification.

The role of expert witness is quite different from the role of researcher and teacher. An important example is ultimate opinion testimony. A researcher specializing in eyewitness identification will rarely speak with the witness(s) who made the identification, and so is sometimes asked how he or she can talk about the accuracy of the eyewitness identification without examining the witness. The answer is straightforward and contains an important constraint on the role of an expert in contrast to the fact finders in the legal system (e.g., juries, judges). The important factors influencing eyewitness identification are either facts elicited in depositions, hearings, or in trial or facts about the construction and administration of the eyewitness identification materials and process. These are all external to the witness him- or herself. It is a misconception (and maybe a stereotype) that experimental psychologists testifying as expert witnesses on eyewitness testimony make psychological assessments of the witness. The question also contains a trap and another misconception. Eyewitness experts do not offer conclusions about the accuracy of the eyewitness identification. On the contrary, they discuss the events and conditions that are known to influence the accuracy of identification. It remains the province of the jury (or other fact finder) to consider these factors in their own consideration and judgment about the probable accuracy of the identification. Offering an opinion about the accuracy of the identification is known as ultimate opinion testimony, in that it offers an ultimate opinion about an issue that is to be decided in the trial process. The role of the expert in this instance is to inform the jury so that in their evaluation of the evidence they can consider the expert's information—a kind of teaching

role for the expert. It is important to keep the decision making in the hands of those placed in that role by the legal process; this clearly structures the role of experts in court and related proceedings.

### 3. ARTICLES IN THIS SECTION

Nowadays, forensic psychologists are frequently consulted by courts or lawyers to provide information about a party's mental state or a party's decision-making ability or capacity, although some judges and other court personnel remain confused about the differences in training and expertise offered by a psychologist, a psychiatrist, a social worker, or other mental health professionals. The articles by Poythress, Heilbrun and Lander, and Martindale and Gould serve to clarify the nature and scope of empirically based contributions that psychologists can make by providing forensic mental health evaluations in civil and criminal cases.

The extent to which an individual is personally responsible for his or her conduct is critical in many situations involving both criminal and civil legal issues. For example, a defendant who engages in behavior that is illegal, but who is unaware of what he or she is doing, or who does not appreciate the significance or consequences of these actions may be entitled to a defense against criminal charges on this basis. Similarly, someone who is easily influenced because of a learning disability or some other psychological impairment can not be expected to take charge of his or her own affairs to the same degree as someone without any mental impairment, and once his or her disability is established, may require assistance in the form of a legal representative or guardian. Expert testimony or a report from a mental health professional such as a psychologist may assist the courts in understanding the cognitive limitations of any impairment, temporary or permanent, as it bears on a legal issue. Psychologists who are trained in forensic assessment and evaluation can help courts to resolve such legal issues.

Defenses to criminal charges premised on a mental disease or disorder are often unpopular in the community because of common erroneous assumptions about the frequency and consequences of such a plea. Research by Silver *et al.* on the public perception of the number of defendants who plead not guilty by reason of insanity shows the number is often exaggerated, and moreover, that the public wrongfully assumes that defendants who enter such a plea are released into the community. In fact, the insanity defense, which applies

only to the defendant's state of mind at the time the alleged offense was committed, is used in fewer than 1% of all criminal cases, and when proposed it is unsuccessful 74% of the time. As Perlin points out, the consequence of entering a plea of insanity is often a period of incarceration in a psychiatric hospital longer than the offender would spend in prison if convicted of the crime charged.

Other defenses proffered on grounds of diminished capacity or limited mental responsibility are similarly unpopular and controversial, as they, too, are often thought to permit culpable offenders to circumvent punishment for their transgressions. Expert assessment and expert testimony is common in these cases to explain how varied mental or neurological conditions may produce situations in which the conduct of the accused is unintended, e.g., because it was involuntary, perhaps a form of automatism, caused by events such as a head injury or a condition such as sleepwalking, hypoglycemia, epilepsy, consumption of alcohol or drugs, or dissociation. Certain mental conditions or disabilities may exonerate a defendant from criminal responsibility if the offender lacks the capacity to form the necessary criminal intent, e.g., defenses that the offender was suffering from battered woman syndrome, post-partum depression, sexual addiction, urban survival syndrome, false memory syndrome, post-traumatic stress syndrome, multiple personality disorder, or schizophrenia.

### 3.1. Legal Competency

The article on legal competency by Poythress delineates the types of decision-making abilities or competencies at discrete moments in the civil and criminal legal process about which courts most frequently seek advice from psychologists trained in forensic assessment and evaluation. The accompanying tables and charts succinctly illustrate the legal standards or issues in the cases and the related psychological competencies that bear on them. Three elements that must be addressed by a forensic psychologist in each case involving an assessment of fitness or competency are (1) the presence of a mental condition that causes an impairment, (2) specifications of the functional impairment, and (3) the distinct legal abilities key to the case that are affected.

The article begins by orienting practitioners to the context of the task by outlining a framework within the law that psychologists are well advised to understand in order to offer more useful information to the courts. A helpful conceptual tool for practitioners and

psycholegal researchers is a table listing the five aspects of legal competencies identified by Grisso, as it sets forth those aspects of the task most clearly within the province of the forensic psychologist, and those aspects most clearly within the province of the fact finder (judge or jury). In particular, it is important to note that the legal standards or tests for fitness vary from one circumstance to the next. Accordingly, the psychologist must appropriately tailor the information critical to the fact finders in applying the legal standard appropriate to a given case. Many of the competence tests and standards pertinent to civil cases have received less research attention than those pertinent in criminal cases, i.e., criteria relevant to determinations of fitness to make decisions about guardianship, treatment, research participation, and testamentary capacity.

A point of significance is the limited utility of a diagnosis in many legal proceedings. Although the goal or end point of many clinical mental health inquiries is a diagnosis, legally, the diagnosis itself is of little value to a court charged with determining whether someone meets the criteria for a defense of insanity or otherwise lacks fitness to proceed to trial. The fact finder seeks a functional description or understanding of the nature and scope of an individual's impairment from the consulting expert. A diagnosis by a mental health professional of a mentally disabling condition or disorder does not guarantee that the applicable legal test will be met, but the diagnosis may be of some relevance. Severe mental disturbance or lack of a diagnosis neither guarantees nor is fatal to a claim of incompetence or insanity.

The terms insanity and incompetence are legal, not psychological, concepts. The legal test for insanity varies from one jurisdiction or community to another, although most tests incorporate the offender's cognitive awareness of the consequences of the conduct at issue. Tests that address some of the "irresistible impulse" standards have in some cases devolved into a question as to whether the accused would engage in the same conduct were there a policeman present. In some states or countries, the standard of "guilty but mentally ill" has been adopted as a variant of "not guilty by reason of insanity." In other communities, legal policy reforms have established that voluntary conduct that produces diminished capacities, such as drinking alcohol or taking drugs, obviates resort to this defense. Finkel points out that policies along these lines conform with the results of studies of juror decision making in insanity cases, which revealed that jurors often take into account the offender's capacity

to make responsible choices and whether the offender was negligent or reckless in bringing about his or her mental disability.

In his article, Poythress adroitly sets controversial topics such as the insanity defense in context by pointing out the infrequency of its use and its low success rate. He also emphasizes the importance of and demand for many other fitness determinations by psychologists, such as fitness to stand trial, or when the death penalty applies, fitness to be executed. Although it is difficult to obtain precise figures of the number of evaluations prepared annually on issues of fitness or competence, Costanzo shows that approximately 5% of all criminal defendants are assessed for competence to stand trial. Thus, Zapf and Roesch provide estimates that between 25,000 and 39,000 psychological evaluations for competence to stand trial are performed annually in the United States, making this one of the most frequently requested psychological services. Of the group of offenders evaluated for this purpose, Melton *et al.* indicate that a relatively small proportion, 12%, is found incompetent.

This article provides an excellent overview of methodological issues facing professionals who are bound to offer evidence-based conclusions to the court. The article includes practical advice to practitioners, such as avoiding controversy as an expert by refusing to prepare a report that addresses more than one type of fitness. Similarly, psychologists are advised to avoid making any statement about the "ultimate legal issue." As Poythress points out, to do so mixes moral and professional standards.

Poythress also illustrates the significance of a clear understanding of the relevant comparative group in interpreting the test outcomes of standardized test instruments. Over the past decade, more test instruments designed to address legal competence and fitness have been developed and studied, thus the dearth of available forensic tests is decreasing. One of the drawbacks in developing evidence-based instruments is finding appropriate control groups and obtaining permission to conduct randomized trials on relevant forensic populations to produce comparative data and appropriate test norms.

### **3.2. Forensic Mental Health Assessment in Legal Contexts**

The article by Heilbrun and Lander on forensic mental health in legal contexts complements the work of

Poythress by providing a more hands-on, practically oriented discussion. Heilbrun and Lander include step-by-step guidance on how to conduct a forensic mental health assessment, including advice on how to structure a report intended for use by the court. Heilbrun and Lander guide the forensic practitioner through 29 steps for conducting a forensic mental health assessment, starting with the initial consultation and ending with expert testimony. This article leads practitioners through substantive considerations at each of three major steps in conducting forensic assessment, namely, preparation, data collection, and data interpretation. For each step, the authors itemize the standards or guidelines that provide parameters for sound practice. Thus, this article is of value in outlining the various pertinent authorities: legal, ethical, scientific, and practical or clinical.

One of the vital themes developed for sound forensic practice is the importance of gathering information from multiple sources to permit more extensive scrutiny of the reliability of the information gathered in the context of litigation. Heilbrun and Lander state the following:

A single source of information is usually less accurate than multiple sources considered together. This is particularly applicable in forensic mental health assessment, as some of the information sources can be particularly distorted. A defendant might be inclined to exaggerate or minimize symptoms or capacities, depending on the possible implications of each in litigation.

Others exaggerate or fabricate symptoms and problems. Yet others minimize or deny symptoms that are genuinely experienced. Response style is a term used to describe how an individual reports his or her own symptoms, behaviors, and problems, and is used to classify an individual's responding as (1) reliable, (2) exaggerated/malingering, (3) defensive, or (4) uncooperative. Self-reported information from an individual who is not reliable will have significant inaccuracies. This principle emphasizes the value of third party information (records and interviews of collateral observers) in both detecting and correcting these inaccuracies.

Other researchers, such as Rogers, have developed forensic instruments to assist practitioners in discerning distorted responses from litigants who seek to appear better or worse than they are in reality.

A second major theme of forensic mental health assessment in legal contexts, developed by Poythress, Heilbrun, and Lander, is acknowledgement of the

distinction between psychologists who are retained to perform a therapeutic role, e.g., by treating offenders, and those who are retained to perform an evaluative forensic role. The full nature of the distinction is appreciated when one considers the ten differences between these roles elaborated by Heilbrun, outlined previously in this article.

Psychologists are often consulted to assess parental competence in the context of a marital dissolution, where decisions about the placement of children in the custody of one, both, or neither of the parents must be made by a court. One parent may be mentally ill, or there may be allegations of physical or emotional abuse or neglect by a parent; in such cases, the courts must determine whether that parent is unfit to raise the children. More frequently, courts must grapple with the question of which type of custodial arrangement will promote the child's best interests.

Psychologists who are well versed in the research on psychological effects on children of divorce and custodial arrangements can play a significant role in assisting courts with determinations about parental fitness. Family life has traditionally held a sacrosanct position within most legal systems. Over time, social and cultural factors regarding who has worked outside of the home and who has been a primary caregiver for a child have shifted the status and rights of children, women, mothers, and fathers. As Krauss and Sales point out, definitions of the family have undergone a dramatic change, particularly regarding the nature of nuclear and extended families. Hetherington *et al.* show that about one-half of all Australian and North American children born in the 1990s will be raised by a solo parent, and many children have no siblings. The rise of "no fault" divorce laws has contributed to these changes and has removed much of the stigma previously associated with marital dissolution. However, in certain cultures, remaining married to a spouse who takes on a new partner provides more benefits and social advantages to the ousted wife and children than does the option of divorce.

Along with changes to the definition of family, the concept of the psychological parent has developed. Accordingly, the rights of grandparents, step-parents, gay parents, and other domestic partnerships have influenced trends in custodial determinations. The extent to which children are consulted regarding their preferences, and the age at which this occurs, varies from one community to another.

The contributions of and participation of psychologists as researchers and practitioners in issues relating

to marriage, divorce, and child custody have been welcomed by legal practitioners. Bausermann demonstrates how research has contributed to our understanding of circumstances that promote healthy adjustment in children, and can assist in assessing custodial options such as sole versus joint custody. Eight studies that compared children's adjustment in intact families versus joint custodial arrangements showed no adjustment differences in the children. Data from 33 studies indicated that (1) no causal relationship has emerged between the type of custodial arrangements and children's adjustment; (2) when the parents are in extreme conflict, children experience adjustment difficulties no matter what the custodial arrangement; (3) more parental conflict ensues between ex-partners when one parent retains sole custody than when custody is joint legal and physical; and (4) the ongoing involvement of the father with the child is predicted by the relationship between the parents, not by the relationship between the father and the child.

### 3.3. Child Custody

The article on child custody by Martinson and Gould traces the four major doctrines relied upon by courts in making custody decisions: (1) the "tender years" doctrine for children under the age of 7, favoring maternal custody; (2) the "psychological parent" rule, favoring the primary caregiver; (3) the "least detrimental alternative," favoring the stability and ongoing relationships in the child's life; and (4) the "best interests of the child." Although the latter is the prevailing standard, it has repeatedly been criticized as too vague, e.g., regarding the time frame to take into consideration when it is applied.

This article also outlines the key features of ethical practice standards, such as the American Psychological Association Guidelines on Custody Evaluations. These guidelines assume the legal criterion is the best interests of the child. In line with the distinction between functional and diagnostic criteria noted earlier, the guidelines emphasize that evaluators must avoid focusing on the psychopathology of parents except to the extent that it affects parenting skills and fitness. The guidelines echo the need to use multiple methods of data gathering, such as observations, interviews, contacting third parties or collaterals courses, obtaining school reports and records, as well as formal psychological assessments, in order to offer a more reliable recommendation to the courts.

Because the stakes are high in custody disputes, practitioners are advised to take steps to uncover efforts by the parties undergoing evaluation to exaggerate positive information or minimize negative information. The principle of substantiating one's opinion with evidence to back it up is critical to avoid recourse to clinical opinions that are more susceptible to attack. Many of the recommendations for successful forensic consultation in child custody cases can be applied to other areas of forensic practice.

Commentators have pointed out that one consequence of the implementation of no fault divorce laws is the greater likelihood of resorting to mediation to resolve issues of custody. In many jurisdictions, parties litigating in family court must participate in mandatory mediation or family conferences prior to attending any court hearing, and in other jurisdictions, referral to mediation is routine. More custody cases are resolved via mediation than by trial. Research on uses of mediation in family cases is more extensive than in other areas of the law. Sales and Beck argue that mediation may produce more losses for mothers and gains for fathers and can have both positive and negative effects. Success of mediation has more to do with the readiness of the parties for the procedure. The best candidates for mediation are those who are cooperative in resolving conflicts. There is no evidence that the psychological functioning of parents enmeshed in conflict can be enhanced by mediation, or that mediation can change the relationship patterns of parents who are angry, volatile, disengaged, or withdrawn. Researchers also caution against generalizing from measures of satisfaction with mediation.

The approaches to evaluation and assessment outlined by Poythress, Heilbrun, Lander and Martindale and Gould establish core principles that can be applied to all forms of forensic assessment. Greenberg and Brodsky discuss their application to civil damages and assessment of the nature and scope of injuries following an accident, and Foote and Goodman-Delahunty discuss their application following a statutory violation, such as employment discrimination.

Just as children have been given greater and more careful attention in many legal processes than in the past, more recently their testimony has achieved greater acceptance. They have been seen more frequently as victims of crimes and therefore as witnesses. Many nations have experienced an increased focus on children and their welfare in a wide range of legal contexts. As divorce rates have increased, the question of how to provide for the children has come more into focus. In

that context alone, there has been an increased concern for the abilities of children to know their own minds, to give accurate testimony, to withstand the pressures of family members, and to participate in the legal process. Although many of these concerns focused on protecting children from becoming victims of their situation, many children do indeed become victims, often within the family. Various kinds of abuse are observed, from neglect, to beating, to sexual abuse. Abuse that in the past would not have come to light in public or in legal proceedings came to the attention of professionals working for public agencies charged with overseeing (among other things) child welfare. As these events are reported, questions inevitably arise regarding the abilities of children to remember the important events, report their memory, differentiate between what they experience directly from what they have been told, and withstand the pressures of testimony and cross-examination in the presence of strangers and the imposing atmosphere of the courtroom. There are also concerns about the appropriate means of interviewing children to get the most reliable information with the least cost to the child.

### **3.4. Children's Testimony**

Quas and Beck divide the research questions that have been asked in this area into three major categories: characteristics of the child, characteristics of the to-be-remembered event, and characteristics of the interview. Age is a rough way of indexing the cognitive and social changes that occur with psychological development. Many changes occur as children grow and age: their general knowledge, sense of what is important, and ability to interpret and remember events all change in important ways. Other factors (e.g., intelligence, temperament, emotional reactivity) are also important influences on the child's memory and ability to provide reliable testimony. The authors point out that it is difficult to apply knowledge of these factors to make judgments about the validity of a given child's report. However, such information can assist in deciding how best to interview the child, and how to protect the child during the process. The recent move by the Home Office in the United Kingdom to identify and moderate treatment of vulnerable witnesses is of particular interest in this connection.

Event characteristics are of interest here just as they are in other areas of eyewitness memory. Quas and Beck review the major findings from research with children. Emotional events, for example, are remembered better than neutral events, although peripheral

aspects of an emotional event may not be well remembered. But again, applying this knowledge is not always straightforward. What is central to a child may not be what is central to an adult, and as with adults, children vary with regard to what aspects of an event they consider central or peripheral. Like adults, children learn generalized accounts of common, repeated events. But more so than adults, they fail to differentiate between the common characteristics of an event and the unique ones, often failing to report unique events and common, expected events that did not occur in the instance of interest.

The characteristics of both interviews that lead to erroneous reports and those that lead to more accurate and complete reports have been the subject of a very large portion of research on children's testimony. To some degree this was driven by the nature of the court cases that were very much in the public eye in both North America and Europe. The conclusions reached for children are very similar to those developed by Vrij for interviewing adults. Free recall questions are fairly accurate (although they may be incomplete), and suggestive questions lead to errors, in the direction of adopting the suggested information.

### 3.5. Interviewing and Interrogation

Obtaining information from persons associated with a criminal event (and certain other events such as food poisoning, accidents, and workplace events) is a process that is both important and fraught with possibilities for error. Some errors are intentional—e.g., withholding self-incriminating information—while others are mired in the complexities of retrieving precisely accurate memories. The difference between these cases is the difference between interrogations and interviews. This and the related topic of detecting deceitful testimony are treated in the article by Vrij on interviewing and interrogation.

Posing suggestive questions to victims and other witnesses risks the integrity of the resulting information. When witnesses are asked leading questions containing one possible answer to the question, e.g., "Was the car red?", when the answer is "Yes" it is not clear whether it is because the car was in fact red or because the witness has no memory of this fact but answers "Yes" because of a judgment that the questioner probably knows the right answer and is looking for confirmation. Alternatively, the suggestion may actually modify the witness's memory. On subsequent occasions, the witness may confidently and spontaneously

answer "Red" and experience the attribute of "red" associated with that object and event, firmly believing it was red.

One alternative to asking leading questions is to ask questions containing multiple choices, but these are still limiting and leading, albeit in a broader sense. The best practice is the use of open questions, often a number of open questions, sequentially. For example, "Tell me everything you can remember about the car," followed by "What was the color of the car?" if color is not mentioned. In general, it is better to begin with the more open questions, probing with more open questions, and ending with specific questions (leading, in the end) to elicit specific responses. However, if the witness cannot come up with "Red" in response to "What color was the car?", asking "Was it red?" yields interpretable information only if the answer is "No." These topics are discussed in more detail by Vrij.

For more than 20 years, techniques based on context reinstatement have been known to enhance the recall of information, even after significant time delays. The most commonly used of these techniques is the cognitive interview (CI), as developed by Fisher. CI techniques are based on the premise that the witness has the information wanted by the interviewer, that the information is organized by the witness according to meaningful criteria applied by the witness at the time of encoding or organizing the information for retention, and that the retrieval of this information must be organized by the witness (not the interviewer) in ways that fit the organizational criteria imposed on it by the witness in the first place. This set of assumptions requires that much of the social interaction—as well as the cognitive interaction—about the recall process be in the hands of the witness. Witnesses are therefore encouraged to take an active part in the interview, to recall and report everything, and to return to previously discussed topics if additional details have been recalled. Witnesses are also encouraged to report the events under investigation from the perspective of other actors present during the events. Research indicates that context reinstatement techniques greatly increase the gross amount of memory reports. The ratio of correct to incorrect items recalled does not change, however, so while the amount of material produced is greatly increased, its efficiency at differentiating correct from false reports is unchanged. Nevertheless, having much more to work with is a great advantage to investigators.

The Reid technique is the single most popular approach to the interrogation of uncooperative suspects. It is

used to obtain confessions from people whom investigators believe are guilty. Of course, the detection of guilt or innocence is well known to be a judgment prone to error, and the confession is meant to establish guilt, so the technique is to some degree circular. Its adherents defend it as being nearly perfect, producing confessions from the guilty but not from the innocent. However, there is an increasing catalog of cases in which Reid-like techniques are associated with false confessions and wrongful conviction. The technique has achieved a negative reputation in psychological research circles because of certain components of the technique, such as altering the suspect's decision-making environment by lying about the nature and strength of the evidence against him, by painting the picture of his situation as hopeless, and creating a (false) understanding that confession is the only way he can obtain any leniency (something usually outside the investigators control). Vrij's article reviews the Reid technique and provides an extended analysis of its difficulties.

Vrij offers as an alternative a technique used increasingly in the United Kingdom as a response to a series of notorious British miscarriages of justice involving false confessions. The technique involves the extensive use of detailed knowledge of the crime by the investigating officers and the skillful use of questioning techniques.

False confessions have received a great deal of attention in the popular press in the last few years, and many nations have their own catalog of prominent cases. Vrij distinguishes three types of false confessions: voluntary, coerced-compliant, and coerced-internalized. Voluntary false confessions are given with no special effort from investigators. The famous case of Henry Lee Lucas, who is estimated to have confessed to more than 600 murders is offered as an example by Gudjonson. Coerced-compliant confessions are those that result from the pressures and incentives of an interrogation. Coerced-internalized false confessions occur when people become convinced that they actually did commit the crime even though they have no memory of the event. Analysis of these forms of confession is provided, and means of detecting false confessions are discussed.

Law enforcement personnel often display great confidence in their ability to detect deceit in suspects they are interviewing. The bad news is that in controlled studies, people are able to detect truthfulness a bit above the level of chance expectation but detect lies somewhat below the level of chance expectation. Studies using law enforcement personnel, as compared with college students, find similar accuracy rates. Some

professionals (e.g., U.S. Secret Service officers and others with a special interest in deception) perform better than chance levels. Vrij reviews a number of common pitfalls in the deceit detection process, discusses cultural differences in social interaction norms that interfere in the process, and debunks some popular beliefs about the behavior of liars. Techniques for improving deceit detection are reviewed, e.g., attention to the content of a witness's speech leads to more effective assessments of deceit than does attention to non-verbal behavior.

Another form of error, although rarely intentional, is that made by eyewitnesses when they are asked if the perpetrator is among the persons shown to them in a "lineup." The prevalence rate of such errors is not altogether clear. However, the problem has persistently attracted attention for more than 100 years and is not easy to dismiss. The most recent incarnation of attention to this problem is found in the work of the Innocence Project, which as of April 2004 has documented 143 exonerations of wrongfully convicted persons. Eyewitness identification errors are the single most important contributor to these wrongful convictions. The malleability of memory is an important contributor. New information can be incorporated into old memories, and the source of memories is often forgotten. It is often not possible to tease out the origins of what appear to be intact and accurate memories. In addition, it is not uncommon for aspects of a typical event to be incorporated into memory for a specific event in which certain of the typical elements were not present.

### **3.6. Eyewitness Testimony**

The misinformation effect, discussed by Levine and Loftus in their article on eyewitness testimony, is an important example. They show how the wording of a question can lead to the intrusion of non-existent elements into reports of memory. Similarly, the strength of the verb used to characterize and event (e.g., "How fast were the cars going when they smashed/collided/bumped/contacted . . . ?") affects the answer. Not surprisingly, the effects are stronger when the actual memory is weaker. And the passage of time only fosters integration of the new or modified elements into an altered but apparently unitary memory.

The authors review the relatively new and somewhat surprising findings on the subject of creation of new memories of personal events that in fact never

happened. Suggestions made by trusted others are important ways in which non-existent events can be created in memory. But imagining an event oneself can also result in the creation of memories, as the fact that the event was only imagined is itself a memory that may be forgotten.

The spurious relationship between confidence in the accuracy of a memory and its actual accuracy is briefly discussed. Belief in the validity of a confidence accuracy relationship is strong on the part of many people, some of whom will serve as jurors or fact finders in legal contexts. Even though confidently given accounts of events may be convincing, the confident manner of reporting is in itself no indication of the validity of the recollection. Only a very small relationship is found between confidence and accuracy. And to make matters worse, confidence is easily manipulated independently of accuracy, making it an even less useful factor.

The role of emotion in memory is an interesting area of study. Emotional events appear to be well remembered, but it appears that this is primarily for central events, or those that have important meaning for the observer. Peripheral events appear to be more poorly remembered. Studies of activation of brain areas and their relation to memories of emotional events can help to clarify our interpretation of the memory–emotion relationship.

The confrontation of a criminal by victims and other witnesses captures interest and imagination; when the accusing finger points at the perpetrator of a crime, it is the moment of justice. However, this is no simple imposition of justice at a personal level. Everything about the identification process has meaning for the outcome, and eyewitness identification is notoriously fallible. Because the identification and the preparations for it have a central role in a significant proportion of criminal cases, concern for identification procedures has attracted the attention of law enforcement agencies, non-governmental organizations concerned with justice, and the research community.

### 3.7. Eyewitness Identification

Tredoux, Meissner, Malpass, and Zimmerman introduce the topic of eyewitness identification and set forth some of the complexities of this area of applied research. The authors distinguish three stages in the eyewitness process: perceiving an event, storing and assimilating the information obtained, and remembering and acting upon the event. They review and discuss

the ways in which factors related to these stages affect the result of the identification process.

Any identification is dependent on the amount and quality of the information obtained from the original event. This is affected by the amount of time the witness had to view the event and the perpetrator, distance from the event, and its visibility, including lighting and conditions that obscure vision. Apart from external conditions, aspects of the witness are also important at the initial stage: stress, fear, and the presence of a weapon all influence the information-gathering process.

Certain enduring characteristics of witnesses are related to their performance. Their age (at least at the extremes) has some influence. Alcohol and other drugs also have effects. The “race,” sex, and occupation of witnesses have also been studied. While there is no general advantage of males or females in identification performance, there are some differences in memory for events that appear to be matters of attention rather than memory per se. While it is a popular belief that law enforcement officers will be better in facial identification, the small research literature on this topic does not support this idea. However, experience as a officer does relate to the quality of the descriptions they provide, both of persons and events. However, they also tend to misperceive innocent events as being criminal in nature. The race of the witness is unrelated to facial or event memory of reporting performance.

Attributes of the perpetrator also contribute to accuracy of identification, in interesting ways. First, disguises are effective, especially those that cover the most informative areas of the face. There are also distinctive patterns of facial appearance. Typical faces are more difficult to distinguish from each other, while non-typical faces are more distinctive and easier to remember. Some faces can be characterized as having specific characteristics (such as “criminality”); observers agree on this, even though it is unrelated to actual criminality. The race of the perpetrator alone does not contribute to accuracy of identification, but the combination of the race of the perpetrator and the witness does. “Cross-race” identifications are more frequently in error than “own-race” identifications.

An important part of the investigative process is getting information from witnesses, about both the event and the appearance of persons taking part in the event. Verbal descriptions of perpetrators are notoriously sparse, but there are techniques that assist first responders in obtaining more useful descriptions and information; some of these are described. Consequences

of asking witnesses to describe the faces of perpetrators include short-term reduction in the ability to recognize the face described.

The technology of pursuing the identification of suspects by eyewitnesses includes such procedures as searching mug shot collections, constructing composite portraits, and constructing and administering lineup and other identification procedures. These topics are all discussed in this article. More detailed attention is given to the construction, evaluation, and administration of lineups.

It is important for law enforcement organizations to keep collections of facial images so that fair lineups can be constructed for criminal investigation. The authors point out that these collections and the means of perusing them are behind the available technology. Examining mug shots is a facial recognition process. Facial recall, however, externalizes a facial image in memory to a representation that can be shared. Verbal descriptions are a form of face recall; another is the construction of a composite portrait. Of course, if the portrait is constructed by another person through the witness's verbal descriptions and then approval by recognition, it is not exactly facial recall. But when the witness uses composite construction computer programs, the work of the eyewitness is more constructive. These are moot points, however. There is no evidence that composite construction is a reliable means of producing high-quality likenesses from memory. There are other dangers in their use: first, many investigators believe in their usefulness, and second, many times a witness is given a copy of their composite to keep with them. Evidence suggests that repeated viewing of a composite changes memory for the original face.

Lineups should be constructed to be fair in two respects: the suspect should not stand out from the other members of the lineup, and the fillers—the five other members of a six-person lineup—should be meaningful alternatives to the suspect. Quantitative methods for measuring the fairness of a lineup have been developed and can readily be applied by law enforcement personnel who create them. The authors describe these techniques and show their importance for effective use of identification evidence in an investigation and prosecution.

Many law enforcement agencies have policies governing the admonitions that are read to eyewitnesses during the identification process. The authors argue that one of the more important problems in eyewitness identification is the tendency for eyewitnesses to believe that they must choose someone from the lineup

they are shown. Admonitions of various kinds have been developed in an attempt to correct this tendency, and to orient the eyewitness to the important nature of the identification decision.

Alternative forms of lineup procedures have been proposed and researched in the last decade. The authors describe and evaluate a series of these, and describe what is known about the use of lineups involving other modes than visual sensory ones. Finally, the authors consider the beliefs and knowledge held by prospective jurors about eyewitness identification. This is information that feeds into the evaluation of the use of expert testimony in the evaluation of eyewitness identification. The balance between informing the jury and influencing the jury is an important one, and so it is important to know whether juries can benefit from expert assistance. Research indicates that lay jurors lack the requisite knowledge to appropriately evaluate eyewitness identification evidence.

### **3.8. Jury Decision Making**

Opportunities for laypersons to participate in legal decision making exist in many jurisdictions, most commonly provided through some form of jury service, although there is considerable variability in jury systems and the scope of the jury role and the types of cases in which the public participate around the world. Albertson, Farley, and Hans report on developments in the jury decision making area. Jury behavior is one feature of the legal system that has been the subject of extensive psychological research for several decades, starting with the University of Chicago Jury Project. The research database on jury studies is so vast that gaining an overview is challenging. A recent meta-analysis of jury research published in an 18-year period (1977–1994) revealed that studies of criminal cases far outstripped those about civil jury decision making (89 vs 11%). Thus study also provided some insight into the most frequently studied topics of jury behavior: the impact of litigation strategies, the influence of different types of witnesses, and the effects of juror or litigant characteristics. Distinctive subtopics addressed most often by researchers emerged: criminal defense strategies, capital punishment and sentencing, characteristics of individual jurors, judicial instructions to jurors, general litigation strategies, witness performance, evaluation of evidence by jurors, and characteristics of offenders or parties.

Jury researchers have employed a diverse array of methods and approaches to gain insight into factors

that influence jury performance. These include archival studies of jury verdicts in actual cases, case analyses, community attitude surveys (e.g., via telephone), observations of trials and interviews of actual jurors, experimental simulations, and experimental or quasi-experimental designs using real juries.

One of the limiting factors in studying jury decision making is that the deliberations that take place among jurors at the conclusion of the trial are typically confidential, and the observation or recording of jury deliberations in many jurisdictions is prohibited. This policy has impeded researchers from gathering much information about the scope of jury-initiated misconduct in actual trials. Researchers who rely on post-trial interviews can not accurately gauge the extent to which juror recall is accurate and reliable. At times, researchers have employed “shadow jurors” who attend courtroom trials and then are interviewed about their observations to infer what the actual jurors might think of the evidence. Occasionally, extra alternate jurors have been empanelled and their deliberations have been videotaped and analyzed to gain insight into how the information they heard and witnessed is used in reaching a verdict.

Research involving real jurors in actual cases, including archival studies of jury verdicts, is relatively rare, accounting for approximately 10% of the jury studies conducted over an 18-year period. Because this method is not very informative about the decision-making process of the jury, many researchers have adopted a mock trial paradigm in which simulated juries receive information akin to that presented in an actual case. The materials may range in realism from brief written descriptions or case summaries to extensive videotaped trials in which attorneys present opening and closing arguments and question witnesses via direct and cross-examination, and jury instructions on the law are provided to the mock jurors. A benefit of this approach is the more extensive empirical control gained by the researcher, permitting causal inferences to be drawn. When mock jurors are used, the extent to which they are similar to actual or potential jurors is an issue. Although much of this research has been conducted with psychology undergraduates, the use of jury-eligible participants has steadily increased, enhancing the generalizability of these results.

The unit of analysis studies by jury researchers has varied. In some cases, individual jurors are the target of the studies, while in others, group deliberations are included, and the unit of analysis is the jury. The article by Albertson, Farley, and Hans carefully outlines the

major findings to emerge from both types of studies, i.e., those that have focused on individual jurors and those that have addressed the jury as a group. Studies of the deliberation process have increased in the past 30 years, and they have examined the influence of features such as the size of the jury, the way in which differences of opinion between majority and minority groups are resolved, the influence of majority versus unanimity rules, and so forth.

The significance of distinctions between external validity and ecological validity have been highlighted by Bornstein and Vidmar, specifically in discussing whether research conditions in simulation studies include sufficient features of real trials to render them adequately representative of actual trials, permitting researchers to draw on research findings to make policy recommendations regarding reforms to enhance jury performance. Questions of normative criteria against which jury performance should be assessed have also been raised, particularly since research has shown that many judges are susceptible to the same biases or weaknesses as are many jurors. The authors suggested that profitable future avenues of research might include the influence of juror and jury anonymity, options for jurors to become more active during the trial, and comparative jury studies.

Before his death in 1916, at age 53, Hugo Muensterberg conducted research on several of the topics discussed in this section, namely, assessment, detection of deception, bystander and witness accuracy, and jury persuasion. At that time, concern was widespread that he was pushing the state of scientific knowledge beyond its limits. Each of the forensic topics explored by Muensterberg has been the subject of extensive inquiry and refinement by psychologists in the ensuing century. As noted previously, concerns that forensic psychologists must work within the limits of scientific knowledge, in particular by scrutinizing the external and ecological validity of their findings, have remained constant. Many forensic topics that Muensterberg presaged through his interest in environmental influences on individual differences have since been added to the repertoire of psychology and law studies, for instance, interventions to treat violent and sex offenders and the identification of individual and environmental factors that lead to offending behavior. Contemporary events have stimulated advances on topics such as the psychology of law enforcement and of terrorism. Newly emerging topics, such as restorative justice and therapeutic jurisprudence, with more emphasis on contextual factors and community involvement derive

from closer interdisciplinary collaboration between scholars trained in law and psychology than was typical in the field in the past. Research, practice, and policy development borne of collaborative endeavours by those trained in psychology and law may meet with less resistance than did Muensterberg's early work in this field.

### See Also the Following Articles

Child Custody ■ Child Testimony ■ Eyewitness Identification ■ Eyewitness Testimony ■ Forensic Mental Health Assessment ■ Interrogation and Interviewing ■ Jury Decision Making ■ Legal Competency

### Further Reading

- Bausermann, R. (2002). Child adjustment in joint custody versus sole custody arrangements: A meta-analytic review. *Journal of Family Psychology, 16*, 91–102.
- Bornstein, B. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior, 23*, 75–91.
- Costanzo, M. (2004). *Psychology applied to law*. Belmont, CA: Wadsworth/Thompson.
- Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law, 7*, 622–727.
- Finkel, N. J. (1995). *Commonsense justice: Jurors' notions of the law*. Cambridge, MA: Harvard University Press.
- Fisher, R. P. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy and Law, 4*, 732–764.
- Foote, W. E., & Goodman-Delahunty, J. (2004). *Evaluating sexual harassment: Psychological, social, and legal considerations in forensic examinations*. Washington, DC: American Psychological Association.
- Garven, S., Wood, J. M., Malpass, R. S., & Shaw, J. S., III (1998). More than suggestion: The effect of interviewing techniques from the McMartin Preschool case. *Journal of Applied Psychology, 83*, 347–359.
- Grisso, T. (2002). *Evaluating competencies* (2nd ed.). New York: Kluwer Academic/Plenum Press.
- Gudjonsson, G. (1999). The making of a serial false confessor: the confessions of Henry Lee Lucas. *The Journal of Forensic Psychiatry, 10*, 416–426.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer Academic/Plenum Press.
- Heilbrun, K., Marczyk, G., & DeMatteo, D. (2002). *Forensic mental health assessment: A casebook*. New York: Oxford University Press.
- Hetherington, E. M., Bridges, M., & Insabella, G. M. (1998). What matters? What does not/Five perspectives on the associations between marital transitions and children's adjustment. *American Psychologist, 53*, 167–184.
- Home Office (2002). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children*. London: Home Office Communication Directorate.
- Krauss, D. A., & Sales, B. D. (2000). Legal standards, expertise, and experts in the resolution of contested child custody cases. *Psychology, Public Policy, and Law, 6*, 843–879.
- Landy, F. J. (1992). Hugo Muensterberg: Victim or visionary? *Journal of Applied Psychology, 77*, 787–802.
- Melton, G., Pettila, J., Poythress, N., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and attorneys* (2nd ed.) New York: Guilford Press.
- Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., Roth, L., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York: Oxford University Press.
- Muensterberg, H. (1907). *On the witness stand*. New York: Doubleday, Page & Company.
- Neitzel, M. T., McCarthy, D., & Kern, M. J. (1999). Juries: The current state of the empirical literature. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 23–54). New York: Kluwer/Plenum.
- Nicholson, A. (1999). Forensic assessment. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 121–173). New York: Kluwer/Plenum.
- Perlin, M. L. (2000). *The hidden prejudice: Mental disability on trial*. Washington, DC: American Psychological Association.
- Poythress, N. G., Bonnie, R. B., Monahan, J., Otto, R. K., & Hoge, S. K. (2002). *Adjudicative competence: The MacArthur studies*. New York: Kluwer/Plenum.
- Rogers, R. (1997). *The clinical assessment of malingering and deception*. New York: Guilford.
- Silver, E., Cirincione, C., & Steadman, H. (1994). Demythologizing inaccurate perceptions of the insanity defense. *Law and Human Behavior, 18*, 63–70.
- Vidmar, N. (2004). Experimental simulations and tort reform: Avoidance, error and overreaching in Sunstein et al.'s *Punitive Damages* (2002). *Emory Law Journal*, in press.
- Zapf, P., & Roesch, R. (1997). Assessing fitness to stand trial: A comparison of institution-based evaluations and a brief screening interview. *Canadian Journal of Community Mental Health, 16*, 53–66.



# Psychometric Tests

Peter F. Merenda

University of Rhode Island, Kingston, Rhode Island, USA

1. Psychometric Tests Defined
  2. Early History of Psychometrics
  3. Psychometric Theory and Principles
  4. Requirement of Technical Manual for All Psychometric Tests
  5. Consequences of Using Tests That Do Not Comply with Testing Standards and Ethical Principles
  6. Conclusions
- Further Reading

## GLOSSARY

**measurement principles and scales** Psychological scaling methods for scoring tests.

**objective scoring** Procedures that yield the same exact score regardless of who or what conducts the scoring of tests.

**projective techniques** Measurement instruments and devices in which the stimulus items are unstructured and/or ambiguous.

**psychological testing** A procedure that involves the use of tests to measure psychological constructs of individual persons.

**psychometrics** The formal study and practice of test construction principles and evaluation of numerical test results and the interpretation of protocols.

**psychophysics** The study and determination of sensory thresholds.

**subjective scoring** Procedures in the scoring of tests that involve intuition, bias, and other subjective judgments that yield different test scores depending on who or what does the scoring.

**technical manual** Booklet accompanying tests that provides technical information relating to sound psychometric properties of the tests.

**testing standards** Standards prepared periodically by professional associations with expertise in psychological testing and made available to test constructors and users by the American Psychological Association.

## 1. PSYCHOMETRIC TESTS DEFINED

The term “psychometric” is defined in *Webster’s* dictionary as “the measurement of the duration, force, interrelations, or other aspects of mental processes, as by psychological tests.” However, not all psychological tests are psychometric tests. Cronbach, an early prominent psychometrician and author of books and journal articles in the field of psychological testing, identified psychometric tests as those that obtain numerical estimates of human behavior and performance, in accordance with Thorndike’s dictum that “if a thing exists, it exists in some amount.” Cronbach went on to distinguish psychometric testing from impressionistic testing. Within Cronbach’s framework and classification of tests in the former category are tests of general intelligence, ability achievement, performance, personality, and the like. In the latter category are the projective devices and techniques as well as rating scales and interview measures based on subjective judgments. An even earlier distinction between these classifications of tests was formally made by the American Psychological Association (APA) in its first publication of *Ethical Standards of Psychologists* and citing incidents of reported ethical problems of psychologists. In Section 5.4,

“Publishing and Using Psychological Tests and Diagnostic Aids,” it was inferred that the former are psychometric tests and the latter are not. (Incidentally, it was in this section that the three-tier (i.e., A, B, C) hierarchy in classifying tests and diagnostic aids was established and first published.)

### 1.1. Characteristics of Psychometric Tests

Paramount among the features of psychometric tests is objectivity of scores obtained through standardized testing procedures. This means that whatever or whoever produces the scores, they are always the same; subjective judgment never enters into the process of scoring. The numerical scores of psychometric tests yield continuous or discrete frequency distributions that may legitimately be analyzed statistically. Furthermore, for a test, a scale, or an inventory to qualify as being a psychometric one, there must be evidence that the instrument possesses sound psychometric properties. Finally, the instrument should be shown to measure psychological constructs, particularly those consistent with well-known psychological theory.

### 1.2. Characteristics of Nonpsychometric Tests

A distinguishing feature, particularly of projective techniques, is the unstructured tasks that are presented to respondents and the ambiguous stimulus items. The principal contrasting feature between psychometric and nonpsychometric tests is the objectivity versus subjectivity of scoring. In the vernacular of psychometricians and measurement specialists, the scoring of nonpsychometric tests involves, in the main, what are referred to as “subjective–intuitive procedures.” However, in spite of these differences, instruments in this category are required by national testing standards to be in compliance with those published by national professional societies. Table 1 provides examples of both psychometric and nonpsychometric tests.

## 2. EARLY HISTORY OF PSYCHOMETRICS

In the social sciences, particularly education, psychology, and sociology, psychometrics as an academic field of study is of relatively recent origin. It is barely 50

TABLE 1

Examples of Psychometric and Nonpsychometric Tests	
Psychometric tests	Nonpsychometric tests
California Personality Inventory (CPI)	Children's Apperception Test (CAT)
Differential Aptitude Tests (DAT)	Machover Draw-A-Person Test (D-A-P)
Scholastic Assessment Test (SAT)	Rorshach Inkblots (RI) <sup>a</sup>
State-Trait Anxiety Inventory (STAI)	Rosenzweig Picture Frustration Study (RPFS)
Wechsler Intelligence Scale For Children (WISC)	Thematic Apperception Test (TAT)

<sup>a</sup>To be distinguished from the Holtzman Inkblot Technique, for which a genuine attempt has been made by the author, Wayne Holtzman, to meet standards of psychometric tests.

years old. The first formal programs to train psychometricians in graduate schools in the United States were established shortly after the end of World War II. They were the programs at the University of Chicago (headed by Palmer O. Johnson) and at Harvard University (headed by Phillip J. Rulon). However, the underlying roots can be traced to the 19th century, even prior to the establishment of psychology as a discipline by Wundt in 1879.

### 2.1. Derivation From Psychophysics

Historically, psychometrics can be viewed as an outgrowth of psychophysics during the latter part of the 19th century. In Germany, Fechner and his colleague Weber, who in 1834 had postulated the famous Weber's Law of just noticeable differences, were the founders of psychophysics. Both were early physiologists and were credited with founding psychophysiology. It was these original works by Fechner and Weber that led to the postulation of the measurement model in the United States by Stevens in 1946. But because Fechner's works laid the early foundation to the basic principles of measurement, Fechner is often referred to as the first psychometrician.

### 2.2. The Measurement Model

At Harvard University, Stevens was the psychologist who provided the social sciences with a taxonomy of measurement scales that form the base for the scoring

of psychometric tests. Stevens had been greatly influenced by the earlier work of Fechner and Weber with physical scales. Besides Stevens, other psychologists in the United States, such as Thurstone and Likert, had already begun to extend Fechner's work to measure psychological constructs. Later, fundamental advances to furthering work in psychometrics beyond psychophysical theory and methods were the measurement models based on true score theory and item response theory.

### 3. PSYCHOMETRIC THEORY AND PRINCIPLES

Commensurate with and preceding the existence of formal academic graduate courses and programs in psychometrics, as well as the publication and distribution of national testing standards and ethical principles governing the rise of tests operationally and in research, a number of textbooks explaining psychometric theory began to be published. Prominent among these early texts were those of Guilford in 1936 and 1954, Gulliksen in 1950, Lord and Novick in 1968, and Nunnally in 1967. All of these books are still in print and available to students, professors, and researchers. The Lord and Novick book was supplemented by that of Wainer and Braun in 1988. In 1978, Nunnally revised the first edition of *Psychometric Theory*, and several years after his untimely death, a third edition by Nunnally and Bernstein appeared on the scene in 1994.

#### 3.1. Test and Testing Standards

The first test and testing standards were published by APA in 1954, that is, 1 year after its publication of the *Ethical Standards of Psychologists*. These testing standards were the *Technical Recommendations for Psychological Tests and Diagnostic Techniques*. They were published as a supplement to the *Psychological Bulletin*. Subsequently, the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME) collaborated with APA in producing the *Standards for Educational and Psychological Tests* in 1966, 1974, 1985, and 1999. The documents are published and distributed by APA. They set forth the principles of sound psychometric properties to be met by all psychometric tests as well as all nonpsychometric tests.

#### 3.2. Ethical Principles Governing Use of Tests and Testing Practices

Following the first publication of the *Ethical Standards of Psychologists* in 1953, APA has periodically revised and updated these principles. In each case, a section has been included that relates to the proper use of tests and testing practices. On June 1, 2003, an expanded *Ethical Principles of Psychologists and Code of Ethics* took effect.

#### 3.3. Principal Psychometric Properties of Tests

The current testing *Standards* publication is the most comprehensive single documentation of tests and testing practices ever published and circulated by APA. There are 15 sections of standards that cover 194 pages (including the glossary and index). Foremost among the standards with which all psychometric (and other measurement) instruments must comply are Standards 1 to 6:

1. Validity
2. Reliability and errors of measurement
3. Test development and revision
4. Scales, norms, and score comparability
5. Test administration, scoring, and reporting
6. Supporting documentation for tests

Adhering to the principles espoused by the *Standards* for testing and mandated by the *Ethical Principles* leads to the assurance that measurement instruments and the scores they yield possess the required psychometric properties. Only in this way can it be established that the assessments measure the psychological constructs that are purported to be measured by the instruments. Equally important is evidence that adherence leads to a determination that the scores produced are internally consistent and stable over time (i.e., reliability) and that the purposes for which the tests were developed are being fulfilled. Compliance also ensures that the properties of the tests demonstrate evidence of and support for the meaningfulness and usefulness of the interpretation of the scores yielded by the tests and are consistent with psychological theory (i.e., validity).

### 4. REQUIREMENT OF TECHNICAL MANUAL FOR ALL PSYCHOMETRIC TESTS

In *Section 6* of the *Standards*, authors and test publishers are advised to provide test users with empirical

documentation that the tests in question possess the required sound psychometric properties. Standard 6.1 states, "Test documents (e.g., test manuals, technical manuals, user's guides, and supplemental materials) should be made available to prospective test users and other qualified persons at the time a test is published or released for use." Until recently, many psychometricians and measurement specialists who were requested to review and critique tests for the *Mental Measurement Yearbooks*, including the current author, would decline to review tests that were not accompanied by technical manuals and other supporting documentation. The 2001 edition of the yearbook contains an appendix titled "Tests Lacking Sufficient Technical Documentation for Review" listing 94 such tests that had been submitted for review to the Buros Institute of Mental Measurements.

### 5. CONSEQUENCES OF USING TESTS THAT DO NOT COMPLY WITH TESTING STANDARDS AND ETHICAL PRINCIPLES

Prior to the publication by APA of the first ethical codes of psychologists in 1953 and the first testing standards in 1954, there were no national regulations in the United States governing the use of tests, whether psychometric or otherwise. However, the current author began to testify in the federal courts as an expert witness (a psychometrician) to support the sound psychometric properties of tests in 1952. Over the subsequent 50 years, he continued to testify in favor of the proper use of psychometric tests and in opposition to the misuse of tests. Much of the misuse has involved violations of the testing standards and the ethical codes of psychologists. These violations have included failure to accompany tests with technical manuals and further documentation, improper application of tests, improper interpretation of test scores, and "stealing" of materials and systems copyrighted by legitimate test authors.

The consequences of violations of the *Standards and Ethical Principles* may fall into moral or legal categories. Moral violations may be those on the part of psychologists who are APA members and/or American Psychological Society members. Violations in this category usually result in dismissal of membership in either or both professional associations as well as loss of license for psychologists in private practice. (The same is true for members of other professional associations

and societies.) The consequences in the legal category are much more severe and costly, especially in court decisions in class-action suits. As an example of the latter, a summary of a class-action suit in which the current author was the principal expert witness (as a psychometrician) for the plaintiffs is presented next.

A 1991 case, *Soroka, et al. v. Dayton-Hudson Corporation*, was settled by an out-of-court agreement between the parties. The consent decree was ordered by the trial court judge and was signed by both parties on September 23, 1993. The general terms of the agreement provided for payment of \$1.3 million to an estimated 2500 job applicants who were administered the test at issue between 1988 and 1991, payment of an additional \$60,000 for four plaintiffs legal fees, and payment of an additional \$35,000 for services to the plaintiffs' legal staff in administering the claims procedure and the settlement funds.

The case was filed by applicants for the position of store security officer at Target Stores in California. Target Stores is owned and operated by Dayton-Hudson, with headquarters in Minneapolis, Minnesota. The applicants claimed that they were required to take the 704-item Rodgers Condensed California Personality Inventory-Minnesota Multiphasic Personality Inventory (CPI-MMPI) and alleged that the test was not psychometrically sound, that it violated California's fair employment laws, and that it constituted an unlawful invasion of privacy protected by the constitution of the State of California.

Four expert witnesses for the plaintiffs (three psychometricians and one industrial-organizational psychologist) were successful in convincing the court of the merits of the complaints and charges filed against Dayton-Hudson. It was the experts' respective testimonies, based primarily on alleged violations by the defendant, that led Dayton-Hudson to seek the out-of-court agreement. On October 25, 1991, the court granted the plaintiffs a preliminary injunction on further testing in concluding that the plaintiffs had in fact convinced the court of the merits of the claims that had been filed. The defendant had considered appealing the ruling to the California Supreme Court but then decided against following through on the appeal.

### 6. CONCLUSIONS

Psychometrics has enjoyed a history about as long as that of psychology itself. It has been viewed as an outgrowth of the earlier psychophysics. Many

psychometricians and measurement specialists recognize Fechner, the early German psychophysicist, as the Father of Psychometrics.

Psychological tests are classified into two major categories: psychometric and nonpsychometric. The principal distinctions between the two rest mainly with (a) the nature of the scoring of test items and (b) the nature of the items themselves. To qualify as a psychometric test, the scoring procedures must be purely objective, whereas in the scoring procedures for nonpsychometric tests, subjective judgment on the part of the scorers enters into the process of scoring the items. Relating to the items themselves, those for psychometric tests are structured, whereas those for nonpsychometric tests are unstructured and generally ambiguous. However, both types of psychological tests are required by national regulations to comply with the *Ethical Standards of Psychologists* and with the *Ethical Principles of Psychologists and Code of Conduct*. The publication and dissemination of these testing standards and ethical principles and codes for test constructors and test users dates back more than 50 years. Violation of these regulatory standards and principles may result in dire consequences to the violators so as to protect the test takers.

### See Also the Following Articles

Assessment and Evaluation, Overview ■ Intelligence Assessment ■ Personality Assessment ■ Psychological Assessment, Standards and Guidelines for ■ Psychophysiological Assessment

### Further Reading

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: Authors.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1974). *Standards for educational and psychological tests*. Washington, DC: Authors.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: Authors.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological measurement*. Washington, DC: Authors.
- American Psychological Association. (1953). *Ethical standards of psychologists*. Washington, DC: Author.
- American Psychological Association. (1954). *Technical recommendations for psychological tests and diagnostic techniques*. Washington, DC: Author.
- American Psychological Association. (1992). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.
- Buros Institute of Mental Measurement. (2001). *Mental Measurement Yearbook* (14th ed.). Lincoln: University of Nebraska Press.
- Camara, W. J., & Merenda, P. F. (2000). Using personality tests in employment screening: Issues raised in *Soroka v. Dayton-Hudson Corporation*. *Psychology, Public Policy, and Law*, 6, 1164–1186.
- Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper & Row.
- Fechner, G. (1860). *Elements de Psycholigsik*. Leipzig, Germany: Bretikopf & Hartel.
- Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: McGraw-Hill.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 44–53.
- Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs*, 61(No. 4).
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Merenda, P. F. (1990). Present and future issues in psychological testing in the United States. *Psychological Assessment*, 6, 3–31.
- Merenda, P. F. (1993). An anticipated landmark decision involving issues related to psychological assessment of applicants for positions in business and industry. *Psychological Reports*, 72, 102.
- Merenda, P. F. (1994). The settlement of the “Target” case and its aftermath. *Psychological Reports*, 75, 1485–1486.
- Merenda, P. F. (1995). Substantive issues in the *Soroka v. Dayton-Hudson* case. *Psychological Reports*, 77, 595–606.
- Merenda, P. F. (2003). Measurements in the future: Beyond the 20th century. *Psychological Reports*, 92, 209–217.
- Nunnally, J. (1967). *Psychometric theory*. New York: McGraw-Hill.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J. C., & Bernstein, L. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Paedagogiske Institute.

Sereci, S. G., Wainer, H., & Braun, H. (1998). *Psychometrics*.  
Princeton, NJ: Educational Testing Service.

Stevens, S. S. (1946). On the theory of scales of measurement.  
*Science*, 103, 667–680.

Thurstone, L. L. (1927). A law of comparative judgment.  
*Psychological Review*, 34, 273–286.

Wainer, H., & Braun, H. J. (Eds.). (1988). *Test validity*.  
Mahwah, NJ: Lawrence Erlbaum.



# Psychoneuroimmunology

Lise Solberg Nes and Suzanne C. Segerstrom

University of Kentucky, Lexington, Kentucky, USA

1. About the Immune System
  2. Pathways between the Brain and the Immune System
  3. Stress, Hormones, and Immunity
  4. Psychological and Behavioral Interventions
  5. PNI and Interventions for Chronic Disease
  6. Conclusion
- Further Reading

## GLOSSARY

**cytokines** Proteins created by cells that influence the behavior of other cells; they often act as messenger molecules that enable communication between cells, particularly immune cells.

**endocrinology** The study of internal secretions of the body, particularly hormonal secretions, and activities involved in such secretions.

**granulocytes** White blood cells containing granules that themselves contain substances that can be used to combat invaders (e.g., oxygen radicals).

**hormone** A substance secreted by the body that affects distant target cells.

**immune system** A system of organs and cells that protects the body from intruding organisms and, in some cases, malfunctioning cells.

**interventions** Attempts to modify behavior, cognition, or emotions so as to improve psychological and physical well-being.

**leukocytes** White blood cells, including granulocytes, monocytes, and lymphocytes.

**lymphocytes** White blood cells that have specific abilities to kill particular invaders (e.g., a given cell recognizes and responds to only one stimulus) through cytotoxic action or antibody production; lymphocytes include B and T cells and natural killer (NK) cells, which are a type of nonspecific lymphocyte.

**monocytes** Mononuclear cells that can migrate into tissue (where they are called macrophages), induce inflammation, and phagocytose (i.e., engulf or eat) invaders.

**neuroendocrine** Relationships between the nervous and endocrine systems as well as cells that release hormones into the blood in response to a neural stimulus.

**psychoneuroimmunology** The study of interactions among the mind, the nervous system, and the immune system.

**psychosocial** Having to do with an individual's psychological or social state.

Psychoneuroimmunology (PNI) is the study of interactions among the mind, the nervous system, and the immune system. The mind, or psyche, involves thoughts, emotions, experiences, and ideas; the nervous system involves the brain, the spinal cord, and nerves throughout the body; and the immune system consists of organs and cells that defend the body against invaders. The mind and the immune system communicate through the peripheral nervous system, hormones, and cytokines. This communication allows the immune system to be responsive to psychosocial factors and allows the immune system to signal the brain (e.g., about infection). The field of PNI is interdisciplinary and

connects knowledge from endocrinology, immunology, psychology, neurology, and other areas. There is increasing interest in the applied aspects of PNI, specifically whether psychosocial interventions can affect diseases such as human immunodeficiency virus (HIV) and cancer.

## **1. ABOUT THE IMMUNE SYSTEM**

The main function of the immune system is to protect the body from infections. When foreign substances or pathogens (e.g., bacteria, viruses, parasites, fungi) attack the body, the immune system detects and destroys these invaders. In addition, the immune system can repair and heal tissue damage from injury as well as destroy and eliminate otherwise normal cells that have been altered by infections or, in some cases, malignancies.

The immune system consists of white blood cells that develop in the bone marrow and the thymus; congregate in the spleen, the lymph nodes, the lungs, the gastrointestinal (GI) tract, and the skin; and circulate through the blood and lymph. These cells include granulocytes and macrophages, which act nonspecifically in processes such as inflammation. They also include lymphocytes, such as B, T, and natural killer (NK) cells, which act specifically on targeted invaders to kill infected cells (i.e., cytotoxicity) and produce proteins that attack invaders (i.e., antibodies). A critical element of the immune response is the production of molecules that amplify and direct the immune response (i.e., cytokines). Different cytokines promote inflammation, cytotoxicity, antibody production, or some combination thereof.

## **2. PATHWAYS BETWEEN THE BRAIN AND THE IMMUNE SYSTEM**

Historically, the immune system was thought to be an independent system. However, during the 1960s and 1970s, evidence that the immune system could be influenced by psychological phenomena began to accumulate. In 1964, Solomon and Moos proposed that the development of rheumatoid arthritis was related to personality factors, advancing the possibility of psychoneuroimmunological phenomena. In 1975, Ader and Cohen demonstrated that conditioning could affect the immune response by pairing taste with an immunosuppressive drug. Animals that were

conditioned in this way showed evidence of immunosuppression when the taste was later presented alone. Evidence of anatomical pathways between the brain and the immune system began with the discovery of sympathetic nervous system innervation of immune organs and has accumulated rapidly to include numerous neurological, hormonal, and cytokine pathways.

### **2.1. Autonomic Nervous System**

Bidirectional neurological pathways exist between the central nervous system (CNS) and the immune system. The most direct pathway from the CNS to the immune system is through the autonomic nervous system (ANS) as sympathetic nerve fibers innervate lymphoid organs such as the thymus, the bone marrow, and the spleen. The influence on the immune system depends on the specific cells and receptors involved and can be inhibitory or stimulatory. Recent evidence suggests that the immune system can also signal the brain through the ANS, specifically the subdiaphragmatic vagus nerve.

### **2.2. Endocrine System**

The endocrine system consists of a variety of glands that secrete hormones. Hormones are chemical substances that are released into the bloodstream, stimulate receptors in target cells, and alter the behavior of those cells. Ample evidence exists that neuroendocrine substances influence immune cells. For example, immune cells react to many hormones, including cortisol, substance P, vasoactive intestinal peptide, prolactin, and growth hormone. Many of the cells involved in the immune response carry surface receptors for a number of hormones, allowing them to respond to signals from the brain as well as from other organ systems.

### **2.3. Immune-to-Brain Communication**

Cytokines, particularly proinflammatory cytokines, can affect both neuroendocrine and brain function in an endocrine fashion. When an immune response occurs, cytokines signal a state of disease to the CNS. Proinflammatory cytokines raise body temperature (fever), and this retards the growth of some bacteria but does so at a cost: Fever increases the host's metabolic rate by 13% for each 1 degree Celsius. As a result, behavior changes to preserve metabolic resources for fighting the infection. "Sickness behavior," which

accompanies fever, includes feeling tired and weak, loss of interest in surroundings, loss of appetite, and loss of concentration. These behavioral changes are consistent with centrally mediated change in priorities away from activity and pursuit of goals to conservation of energy.

### 3. STRESS, HORMONES, AND IMMUNITY

The idea that stressor exposure might influence the immune response was an early topic of interest in psychoneuroimmunology (PNI), and evidence clearly indicates that psychological and physical stressor exposure can affect the immune system. Both the type of stress and the person involved may affect the influence of stress on the immune system. Stress can be physical or psychological, controllable or uncontrollable, and acute or chronic. Furthermore, different people may handle stressor exposure in different ways.

When a situation is perceived as threatening, the sympathetic nervous system responds by releasing epinephrine and norepinephrine. Consequences of this response include increased heart rate, increased blood pressure, and peripheral vasoconstriction. Stressor exposure also influences the endocrine response, changing levels of hormones such as cortisol, a steroid hormone produced by the hypothalamic–pituitary–adrenal (HPA) axis. These responses can, in turn, affect the immune system.

There are more than 300 studies that have linked stress and stressors to immune parameters. In a meta-analysis of these studies, both acute stressors (lasting minutes) and chronic stressors (lasting days to years) were reliably associated with changes in the immune system. The effects of stressors and stress hormones depend critically on characteristics of the stressor, particularly time course. In general, the changes associated with acute stressors could be adaptive as part of the “fight or flight” response; acute stressors (lasting minutes) are associated with large increases of non-specific defenders (e.g., neutrophils, NK cells) in the bloodstream, release of antibody into saliva, and increased production of cytokines that stimulate non-specific immunity. Mechanistic investigations have shown that many of these changes are due to sympathetic signaling through  $\beta_2$ -adrenergic receptors on lymphocytes.

However, longer term stressors had less potentially beneficial effects. The shortest “chronic” stressors, such as academic examinations, were associated with decreases in the function of cytotoxic lymphocytes (e.g., NK cells, T cells), but antibody-producing functions were generally unchanged. Chronic stressors that were long-lasting, such as having a spouse with dementia, decreased both cytotoxic and antibody-producing functions of immune cells.

### 4. PSYCHOLOGICAL AND BEHAVIORAL INTERVENTIONS

PNI is mainly a basic science. However, the number of studies investigating the impact of psychosocial interventions on the immune system is increasing. In particular, the hypothesis and evidence that psychosocial stress can affect the immune system has generated the secondary hypothesis that psychosocial interventions to alleviate stress could have the opposite effect on the immune system. Such interventions commonly involve stress management, relaxation, and hypnosis. A recent meta-analysis found that there were few effects of interventions on the immune system but also that most stress management and relaxation interventions failed to enroll stressed participants. In those that did, there was some evidence that the interventions could alleviate stressor-related immune change.

#### 4.1. Stress Management

In stress management interventions, participants are taught to recognize situations they experience as stressful and to identify the reactions they have in these situations. The participants are then trained to avoid unnecessary stressor exposure and to cope better with stressors. Stress management interventions have mostly been targeted to patients with chronic medical illnesses such as cancer and HIV. Such interventions usually include information and education about the illness, access to psychological support, help in changing cognition about the disease, and training to improve coping skills. Studies involving stress management interventions have yielded mixed results, but some such interventions have been found to reduce negative emotions and have been associated with positive changes in immune function, including increases in CD4+ T lymphocyte numbers, NK cell numbers, and NK cell activity.

## 4.2. Disclosure and Health

Some data indicate that undisclosed traumatic cognitions or emotions might lead to serious physiological consequences. In intervention studies on this topic, participants are asked to write or talk about stressful or traumatic situations that they have experienced. The cognitive processing and emotional expression during this activity are thought to lead to better regulation of physiological systems. Research in this area has found associations between disclosure and reduced negative affect, reduced activity in the sympathetic nervous system, and reduced health center visits. Immunological consequences have included better response to vaccination and better immune control of latent viruses.

## 4.3. Relaxation

A relaxed state can be induced in several ways, including progressive muscle relaxation, relaxation response, biofeedback, imagery, hypnosis, and meditation. In general, all of these forms of relaxation have been related to decreased negative mood, decreased hormone levels, and decreased activity in the sympathetic nervous system. However, the results of relaxation techniques are inconsistent with regard to immune responses. Relaxation exercises are associated with increased levels of antibody in saliva and may increase levels of CD4 helper T cells in the blood as well as NK cell activity.

## 4.4. Hypnosis

Hypnosis involves the induction of an altered state of consciousness, including focused attention and high suggestibility. In research studies, the application of hypnosis to influence the immune system is often through hypnotic suggestion to affect an allergic response. The participant first has an allergen injected under the skin. A hypnotic state is then induced, and the hypnotized participant is instructed to either enhance or suppress his or her response to the allergen. In most hypnosis interventions, the direction of effect is specified. Following such interventions, hypnosis has successfully suppressed, but not enhanced, the immune response. Studies have also shown that hypnotic suggestions can lead to changes in the ANS such as alterations in heart rate and blood pressure. In medical contexts, hypnosis has been used to manage chronic pain and to enhance adjustment to physical conditions and illnesses. There are great differences in people's susceptibility to hypnosis, and highly

hypnotizable individuals have the strongest immune changes.

## 5. PNI AND INTERVENTIONS FOR CHRONIC DISEASE

### 5.1. Interventions in HIV

The clinical course of HIV infection varies greatly, and although some people can live with HIV infection for years without developing acquired immune deficiency syndrome (AIDS), others progress more quickly after being infected. A number of factors, such as viral strain, genetic immune factors, presence of other infections, nutrition, and drug abuse, are thought to influence progression. In addition, a host of psychosocial stressors related to the diagnosis and stigma of the illness may contribute to differences. A person infected with HIV may face social, medical, legal, professional, and mental problems on diagnosis, and coping with these stressors may make a difference in how the illness progresses.

HIV infection develops into AIDS when the immune system is too weak to fight infections. HIV infection has been associated with physical and psychological stressor exposure and distress, and because stressor exposure and distress have been associated with suppression of immune functions, the progression of HIV infection may be influenced by interventions that can affect responses to stress. In intervention studies, HIV-infected patients in cognitive-behavioral stress management programs had decreased distress levels and higher NK cell and CD4+ levels than did HIV-infected controls. HIV-infected patients in cognitive-behavioral stress management programs also had decreases in depression and anxiety, and both cognitive-behavioral training and relaxation training appear to have positive impact on immune function among patients. However, findings from other studies have yielded contradictory results, and although relationships among behavior, psychosocial stressors, immune status, and HIV progression apparently exist, these relationships are not completely understood.

### 5.2. Interventions in Cancer

PNI may have implications for cancer. One possibility is that psychosocial factors influence the onset and especially the advancement of cancer through psychological impact on the immune system. This is the most

traditional view, and psychological influences on the immune system are often assumed to be the connection between psychosocial factors and cancer. However, controversy surrounds this topic because few studies have been able to detect clear associations between psychological effects on the immune function and cancer outcomes. Some of this confusion may arise from the heterogeneity among various types of cancer that have different molecular mechanisms and different immunological vulnerabilities.

PNI may also be of importance to oncology in that psychological influences on the immune system may affect the risk of infectious disease. Infectious disease is a major source of concern in oncology because the immune suppression often seen in patients due to cancer treatments and psychological distress may render patients' immune systems incapable of fighting off intruding viruses and bacteria, leaving patients more susceptible to infectious disease. However, there is not yet much research in this area.

Associations have been found between support group interventions and longevity in people with cancer. However, other studies have found no association. Some interventions have demonstrated an immune benefit for people with cancer. In cognitive-behavioral stress management interventions, breast cancer and malignant melanoma patients received health education, stress management training, problem-solving skills training, and social support. Participants had decreased depression and anxiety as well as increased immune parameters. However, such effects remain to be replicated, and the consequences of these changes for patient health remain unclear.

## 6. CONCLUSION

Clear connections exist between the nervous and neuroendocrine systems and the immune system, and compelling evidence points to the influence of psychosocial factors on immune function as well as vice versa. Interventions to reduce potentially deleterious effects of stress and depression on the immune system are in their infancy, but there are encouraging examples of how applied PNI may contribute to better immune function and, potentially, to better health.

## See Also the Following Articles

Health and Culture ■ Health Psychology, Cross-Cultural ■ Hypnosis ■ Stress

## Further Reading

- Ader, R., & Cohen, N. (1975). Behaviorally conditioned immunosuppression. *Psychosomatic Medicine*, 37, 333-340.
- Ader, R., Felton, L., & Cohen, N. (2001). *Psychoneuroimmunology* (3rd ed.). San Diego: Academic Press.
- Fawzy, F. I., Fawzy, N. W., & Hyun, C. S. (1994). Short-term psychiatric intervention for patients with malignant melanoma: Effects on psychological state, coping, and the immune system. In C. E. Lewis, C. O'Sullivan, & J. Barraclough (Eds.), *The psychoimmunology of cancer: Mind and body in the fight for survival* (pp. 292-319). New York: Oxford University Press.
- Kiecolt-Glaser, J. K., & Glaser, R. (1992). Psychoneuroimmunology: Can psychological interventions modulate immunity? *Journal of Consulting and Clinical Psychology*, 60, 569-575.
- Maier, S. F., Watkins, L. R., & Fleshner, M. (1994). Psychoneuroimmunology: The interface between behavior, brain, and immunity. *American Psychologist*, 49, 1004-1017.
- Miller, G. E., & Cohen, S. (2001). Psychological interventions and the immune system: A meta-analytic review and critique. *Health Psychology*, 20, 47-63.
- Pennebaker, J. W. (1995). *Emotion, disclosure, and health*. Washington, DC: American Psychological Association.
- Rabin, B. S. (1999). *Stress, immune function, and health*. New York: Wiley-Liss.
- Schedlowski, M., & Tewes, U. (1999). *Psychoneuroimmunology: An interdisciplinary introduction*. New York: Kluwer Academic/Plenum.
- Seegerstrom, S. C., & Miller, G. E. (in press). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychological Bulletin*.
- Solomon, G. H., & Moos, R. H. (1964). Emotions, immunity, and disease: A speculative theoretical integration. *Archives of General Psychiatry*, 11, 657-674.
- Spiegel, D., Bloom, J. R., Kraemer, H. C., & Gottheil, E. (1989). Effect of psychosocial treatment on survival of patients with metastatic breast cancer. *Lancet*, 2, 888-891.



# Psychophysiological Assessment

Jaime Vila

University of Granada, Granada, Spain

1. Introduction
  2. Historical Antecedents
  3. Peripheral Psychophysiological Measures
  4. Central Psychophysiological Measures
  5. The Psychological Context
  6. Assessment of Basic Psychological Processes
  7. Clinical Applications
- Further Reading

## GLOSSARY

- cardiovascular activity** Functional activity of the heart and vessels controlled by the autonomic nervous system.
- electrodermal activity** Electrical activity of the skin due to activation of the sweat glands by the sympathetic nervous system.
- electromyography** Electrical activity of the striated muscle controlled by the somatosensory nervous system.
- event-related potential** Specific change in the electroencephalogram (EEG) due to the presentation of a stimulus or in anticipation of a response; it is recorded after averaging the EEG signals produced by a number of repeated events.
- functional magnetic resonance imaging** Neuroimaging technique based on the magnetic properties of the spinning protons; the greater the magnetic signal in a brain area, the greater the blood flow and neural activity in that area.
- neuropsychological disorder** Psychological disorder due to brain lesion.
- psychopathological disorder** Psychological disorder not due to brain lesion or other organic disorder.
- psychophysiological disorder** Physiological disorder due to psychological factors.

**psychophysiology** The scientific study of psychological processes through noninvasive recording of physiological responses.

Psychophysiology is the scientific discipline devoted to the study of the relationships between the physiological and psychological aspects of behavior. Although such relationships have been studied from different perspectives, the psychophysiological perspective emphasizes the use of noninvasive physiological measures—both peripheral and central—to assess the psychological processes underlying human behavior.

## 1. INTRODUCTION

The scientific study of human behavior becomes enriched when it is done in close relationship with the biological body that sustains behavior. To study the activity of the organisms from this perspective means to examine what people do, think, and feel while simultaneously taking into account the physiological and psychological mechanisms that control human behavior.

The relationships between physiology and psychology have preoccupied philosophers and scientists throughout history. The different debates about the soul–body, spirit–matter, and mind–brain relationships all are variations on a single theme: the connection between the behavior—people’s acts, thoughts, and feelings—and its sustaining biological body.

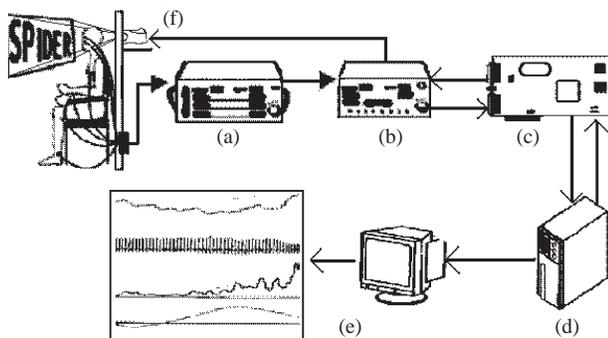
## 2. HISTORICAL ANTECEDENTS

Nearly all physiological recording methods are based on the bioelectrical nature of the living tissue activity. Since the discovery of electricity by the end of the 18th century and throughout the electronic and computer revolution of the 20th century, it is not surprising that the history of the psychophysiological approach to assessment evolved in parallel with technological evolution.

The first psychophysiological instruments were based on the galvanometer, that is, a moving coil suspended in the magnetic field of a permanent magnet that rotates when an electric current flows through the coil. A galvanometer-type device used with a writing system mounted on the moving coil is the basis of the traditional polygraphs.

The electronic and computer revolution during the 20th century influenced psychophysiological instruments in three basic aspects: (a) increased precision and reduced size and weight of the amplifiers; (b) improvements in the traditional recording systems using digital computers to represent, store, and analyze the psychophysiological signals, and (c) allowing the discovery of new psychophysiological variables only available after complex computerized systems have been used to detect and extract the biological signals.

Psychophysiological equipment must take into account the physiological recording and the psychological manipulation. Figure 1 illustrates the basic elements of current equipment. On the left, the



**FIGURE 1** Diagram representing the basic elements of a psychophysiological laboratory: (a) the physiological recording system; (b) the connection box; (c) the analog-to-digital and digital-to-analog card; (d) the computer central processing unit where the card is inserted; (e) the computer monitor where the physiological recording is being displayed; and (f) the slide projector, controlled by the computer, to present the psychological stimuli.

participant has the sensors placed on the surface of the body while the psychological task is being performed. On the right, in the experimenter's room, are the physiological recording system, the connection box, the analog-to-digital and digital-to-analog card, the computer central processing unit where the card is inserted, the computer monitor where the physiological recording is being displayed, and the slide projector, controlled by the computer, to present the psychological stimuli.

## 3. PERIPHERAL PSYCHOPHYSIOLOGICAL MEASURES

Psychophysiological techniques are classified into peripheral and central techniques. Peripheral techniques include those variables directly controlled by the somatosensory nervous system and the autonomic nervous system (e.g., surface electromyography, eye movements, respiration, cardiovascular activity, electrodermal activity). Central techniques include those variables directly controlled by the central nervous system (e.g., event-related potentials [ERPs], functional magnetic resonance imaging [fMRI], positron emission tomography [PET]). Table I summarizes the various psychophysiological techniques following this classification. This section describes one psychophysiological technique belonging to the peripheral somatosensory system (surface electromyography) and two techniques belonging to the peripheral autonomic nervous system (cardiovascular activity and electrodermal activity).

### 3.1. Surface Electromyography

The biological basis of electromyography is the electrical activity of the striated muscle, that is, that type of muscle responsible for motor activity and made up of a large number of parallel fibers. Such activity is triggered by the activation of the motoneuron that innervates the fiber. This functional physiological entity, made up of a single motoneuron and the individual muscle fibers innervated by the motoneuron, is called the motor unit. Recording of the electrical activity of the striated muscles from the surface of the skin is referred to as surface electromyography (EMG). This EMG signal represents the ensemble of action potentials produced by the muscle at a given moment. The

**TABLE I**  
**Classification of Psychophysiological Techniques**

<i>Neurophysiological mechanism</i>	<i>Psychophysiological techniques</i>
Somatic nervous system	Surface electromyography Direct electromyography Integrated electromyography Eye movements Respiration
Autonomic nervous system	Cardiovascular activity Electrocardiogram Impedance cardiography Photoplethysmography Blood pressure recording Electrodermal activity Pupillary activity Gastrointestinal activity Others
Central nervous system	Electromagnetic techniques Electroencephalogram Event-related potentials Brain electrical activity mapping Magnetoencephalography Metabolic techniques Positron emission tomography Functional magnetic resonance imaging

aggregate signal is characterized by a frequency range between 20 and 1000 hertz and an amplitude range between 2 and 2000 microvolts, displaying a direct relationship between frequency and amplitude; the greater the frequency, the greater the amplitude. In psychophysiological assessment, it is common to transform the direct surface EMG into integrated EMG. The integration procedure implies, first, rectification of the signal (i.e., elimination of the negative mirror part of the signal) and, second, smoothing of the remaining signal by a low-pass filter. This produces a contour following integration, that is, a varying voltage proportional to the envelope of the direct EMG signal.

### 3.2. Cardiovascular Activity

The cardiovascular system consists of the heart and a network of vessels for distribution, exchange, and return of the blood throughout the body. The

cardiovascular system reacts quickly to physical and psychological demands, with such activity being controlled by neural and humoral pathways. The neural pathway to the heart involves the two branches of the autonomic nervous system: the sympathetic and parasympathetic branches. When the sympathetic nerves are activated, the heart rate accelerates and heart contractions become stronger. When the parasympathetic nerves are activated, the heart decelerates. The neural pathway to the vessels involve mainly the sympathetic nerves, and its activation produces either vasoconstriction (in skin and viscera) or vasodilation (in motor muscles). The humoral pathway involves the neural sympathetic activation of the endocrine system, resulting in the secretion by the adrenal glands of various hormones into the bloodstream (e.g., adrenaline, cortisol). The impact of psychological processes on cardiovascular activity is assessed through noninvasive measurements of the various components of cardiovascular activity; heart rate (the number of ventricular contractions per minute), blood pressure (the force of the circulating blood in the arteries, and blood flow (the amount of circulating blood in a particular area). In psychophysiological studies, heart rate is the main index to assess cardiac function. Heart rate is measured in terms of beat-by-beat heart rate values derived from the electrocardiogram (R-R interval).

### 3.3. Electrodermal Activity

Known since the end of 19th century, electrodermal activity is one of the peripheral measures most frequently used in psychophysiological assessment, with its biological basis being the activity of the eccrine sweat glands. Sweat glands are distributed along the surface of the body with a greater density in the inner parts of the hands and feet. The activities of the glands are controlled by the sympathetic branch of the autonomic nervous system, and when the glands are activated, there is a hydration of the skin accompanied by an electrical signal. The electrical signal can be detected on the surface (called skin potential), and there is an increase in conductivity to the passage of an external current (called skin conductance). This activity responds to thermal and psychological stimuli. The psychological significance of the electrodermal activity is predominant when the temperature is below 25 degrees Celsius. The recording of electrodermal activity can be done using two general procedures. First, the endogenous procedure records the electrical potential produced in the skin without passing any external

current. Second, the exogenous procedure records the change in conductivity to the passage of an external current. Depending on the type of external current (DC or AC) and whether the current maintains a constant voltage or constant intensity, the recording procedure is called skin conductance (DC and constant voltage), skin resistance (DC and constant intensity), skin admittance (AC and constant voltage), or skin impedance (AC and constant intensity).

#### 4. CENTRAL PSYCHOPHYSIOLOGICAL MEASURES

The techniques to measure central nervous system activity are classified into two categories: electromagnetic and metabolic techniques. The electromagnetic techniques are based on the electrical activity of the brain tissue, that is, the neurons' action potentials. The metabolic techniques are based on the differential distribution of blood elements (mainly oxygen and hydrogen) in the brain.

##### 4.1. Electromagnetic Techniques

Electroencephalography (EEG) was the first electromagnetic technique developed. Like all psychophysiological techniques, it requires careful attention to electrode location, choice of electrode reference, grounding, proper filtering and artifact control. A good EEG recording is characterized by changes in voltage as a function of time with two inversely related parameters: frequency and amplitude. The lowest frequency band (0–4 hertz, called the delta range) has the greatest amplitude (20–200 microvolts), and the highest frequency band (36–44 hertz, called the gamma range) has the lowest amplitude (2–20 microvolts). The intermediate frequency bands, called the theta (5–7 hertz), alpha (8–12 hertz), and beta (13–30 hertz) ranges, have a progressively decreasing amplitude. An important technological development based on the EEG occurred during the 1960s with the advent of digital computers. Digital computers allowed the event-related brain activity to be extracted from the background EEG by applying the averaging procedure, that is, repeating a particular event and averaging the digital samples of the EEG that are time-locked to that event. The resulting signal, the ERP, contains a number of positive and negative peaks identified by the letters P and N, respectively, each with a numeric

subscript (1/2/3/4 or 100/200/300/400) indicating the order or time in milliseconds of the peak, respectively. Examples of ERPs include the N<sub>100</sub>, P<sub>300</sub>, and N<sub>400</sub>. Once extracted, ERPs are interpreted as manifestations of the specific psychological processes activated by the event.

##### 4.2. Metabolic Techniques

The metabolic techniques are similar to traditional radiological techniques (e.g., X ray, computerized tomography). The neuroimage is elaborated from the recording of radiological signals emitted by moving blood particles in the brain (oxygen and hydrogen) necessary for neuronal activity. The first metabolic neuroimaging techniques developed were the radiotracer techniques. PET is one of them. The radiotracer commonly used in PET is a pharmacological compound labeled with a positron-emitting radioisotope ( $O_{15}$ ) of short life (approximately 4–10 minutes). Once administered to the participant either intravenously or by inhalation, the radiotracer is distributed through the blood to all parts of the body, including the brain. The number of emitted positrons detected from outside the head reflects the amount of oxygen demand in the various brain areas.

The most promising metabolic technique, as an alternative to radiotracers, is fMRI. The physical foundations of fMRI are the magnetic properties of the spinning protons within the nucleus of certain atoms. These subatomic particles behave like small magnetic needles. When the body is placed within the magnetic field of a potent external magnet, all hydrogen and oxygen protons line up along the axis of the external magnet. The orientated spinning protons rotate, or precess, about the axis of the external magnetic field with a rotating frequency that is within the range of radio frequency waves. Using a transmit–receive coil, the MRI technique then applies a momentary external radio frequency wave perpendicular to the magnet axis and identical to the precessional frequency of the spinning protons to produce the resonance phenomenon, that is, a momentary electromagnetic signal detected by the receive coil and used to create the image.

#### 5. THE PSYCHOLOGICAL CONTEXT

The physiological recording becomes psychophysiological only when the environmental context and the activity of the participant during the recording are well controlled. The task of the participant during the

recording is the critical aspect providing psychological meaning to the physiology. Control of the context implies paying attention to three items: the place, the participant, and the task.

### 5.1. The Place

The place refers to the physical and social environment of the recording. There are numerous physical and social environmental variables that may affect the physiological recording. The main physical variables are noise, light, temperature, and humidity. The main social variables are the presence of people during the recording (including the experimenter). To control these variables, specific rooms in which such variables could be eliminated or maintained constant should be used. The optimal place is the psychophysiological laboratory, with a participant's room independent of the experimenter's room (Fig. 1).

### 5.2. The Participant

The participant always introduces a set of variables that affect the physiological recording. Some of these variables are characteristics of the participant such as age, gender, menstrual cycle (in women), whether or not the participant smokes, and whether or not the participant has any physical disease or psychological condition. Other variables affect the participant indirectly such as physical activity done just before the recording, hour of the day, and food, medication, or drinks taken before the recording. All of these variables should be included within the information protocol obtained at the beginning of the assessment session. When possible, they should be controlled either by exclusion of participants with characteristics that saliently affect the recording or by elimination or matching of the conditions affecting the recording.

### 5.3. The Task

The tasks that should be used depend on the aim of the assessment. Each research area of psychology has developed its own assessment paradigms and tasks. Therefore, assessment must be linked to the theoretical contexts in which the different tasks have been developed. Tasks developed to assess attention are different from tasks developed to assess emotion, memory, language, thinking, motor control, or learning. However, all psychophysiological tasks share a common structure: a period of recording with no stimulation (resting period)

used to stabilize the physiological measures and a period of recording with stimulation (the specific task).

## 6. ASSESSMENT OF BASIC PSYCHOLOGICAL PROCESSES

Interest in psychophysiological assessment has existed since the early years of scientific psychology. Examples of psychophysiological applications for the study of basic processes can be found in the work of James (emotion), Pavlov (learning), and Cannon (motivation). This section presents some examples of current basic applications.

### 6.1. Electromyography

Electromyography has been used to study arousal, facial expressions, somatic correlates of thought and mental imagery, motor reflexes, motor skill learning, and so forth. An example of current application of EMG to the assessment of emotional and attentional processes is the startle probe paradigm. In humans, the first and most reliable component of startle is eyeblink. Eyeblink is measured using integrated EMG of the orbicularis oculi (i.e., the muscle around the eye). Startle is typically elicited by a brief sudden noise. Emotional modulation of startle is observed when the reflex is elicited while the participant views affective pictures presented for 6 seconds. Compared with neutral pictures, unpleasant pictures increase the reflex and pleasant pictures inhibit it. Attentional modulation of startle is observed when another stimulus precedes the eliciting stimulus by less than 500 milliseconds. Under these conditions, the reflex is inhibited (i.e., prepulse inhibition) independently of the affective content. Normal and pathological emotional and attentional processes can be assessed using this paradigm.

### 6.2. Heart Rate

Cardiovascular activity has been used in numerous research contexts, including unconditioned reflexes (i.e., orienting and defense), habituation, sensitization, Pavlovian conditioning, instrumental learning, motivation, emotion, memory, and mental workload. Heart rate has been the cardiovascular measure most frequently used in these contexts.

The psychological meaning of heart rate has been a subject of debate in the past. Two major interpretations

can be identified: the cognitive and motivational interpretations. The cognitive interpretation assumes that cardiac changes in response to external or internal stimuli reflect attentional and perceptual mechanisms aimed at facilitating (heart rate deceleration) or inhibiting (heart rate acceleration) sensory processing. On the other hand, the motivational interpretation assumes that cardiac changes reflect metabolic mechanisms aimed at providing energy to adaptive behaviors. If the appropriate behavior is to be passive, the cardiac response will be a heart rate deceleration. If the appropriate behavior is to be active either physically or psychologically, the cardiac response will be a heart rate acceleration.

### 6.3. Skin Conductance

Skin conductance is one of the most sensitive measures of mental activity. It has been used extensively to study attention, habituation, and conditioning. A skin conductance response to stimuli of moderate or low intensity is interpreted as an attentional response indicative that a novel stimulus has been detected and processed. Habituation of the response (when the same stimulus is repeated) and dishabituation (when the repeated stimulus changes in some parameter) is also indicative that the system is correctly processing the environmental information. In the context of conditioning studies, the skin conductance response is interpreted as indicative that a neutral stimulus has acquired the signal value of the unconditioned stimulus, allowing assessment of the degree of learning or unlearning.

Skin conductance has also proved to be sensitive to unconscious presentation of threat stimuli. Using masked stimuli (visual pictures presented for 30 milliseconds followed by a mask), Öhman and colleagues showed that fear-relevant stimuli (pictures of snakes, spiders, and fear–angry faces), as compared with irrelevant stimuli (pictures of flowers, mushrooms, and neutral–happy faces), can be processed without awareness.

### 6.4. Central Electromagnetic Measures

Among the central electromagnetic measures, the ERPs are the most frequently used to assess cognitive functions. The  $N_{100}$ ,  $P_{300}$ ,  $N_{400}$ , and mismatch negativity all are ERPs that are studied extensively in relation to attentional and linguistic processing. Each ERP

requires, in addition to the EEG recording, the use of specific tasks.

One of the tasks to obtain the  $N_{100}$  and the  $P_{300}$  is the oddball task. It consists of a large series of presentations of two stimuli with different degrees of probability, for example, a tone of 1000 hertz presented 160 times (80% probability) and a tone of 1500 hertz presented 40 times (20% probability) in random order. The task of the participant is to pay attention only to the tone with lower probability. Under these conditions, the  $N_{100}$  reflects the nonspecific attention paid to both stimuli, whereas the  $P_{300}$  reflects the specific attention paid only to the low-probability stimulus.

### 6.5. Central Metabolic Measures

PET and fMRI are the latest techniques used in psychophysiological assessment. Identification of the brain areas involved in the psychological functions activated by the task requires application of the subtraction method. It consists of obtaining images at different stages of the task. For example, in a semantic processing task in which a participant is asked to think of the meaning of a word, the task is divided into several stages: to hear the word, to hear the word and repeat it, and to hear the word and generate a verb representing a use of the word (e.g., hammer: to hit). By subtracting the images from the different stages, the specific brain areas involved in the act of thinking can be identified.

## 7. CLINICAL APPLICATIONS

Psychophysiological assessment has been used in many applied areas, including ergonomics, sport, lie detection, and clinical psychology. This section presents some examples of applications within clinical psychology.

### 7.1. Psychopathological Disorders

Psychopathological disorders are characterized by symptoms of psychological dysfunction with no diagnosable brain lesion or physical disease. Psychophysiological assessment of psychopathology has focused on identification of psychophysiological markers, that is, deviation in one physiological variable that is reliably associated with a specific disorder. For example, research on schizophrenia has suggested several markers, including reduced  $P_{300}$ , reduced skin conductance orienting response, reduced prepulse

inhibition of startle, and abnormal eye movements. Similar suggestions have been made in relation to other disorders such as psychopathy, depression, and anxiety, although more research is needed before useful clinical protocols can be developed for these disorders.

## 7.2. Psychophysiological Disorders

Psychophysiological disorders are characterized by symptoms of physiological dysfunction in specific organs such as the heart, lungs, stomach, muscles, and neural tissues. Such dysfunction is caused or aggravated by psychological factors such as emotions, stress, personality, and lifestyle.

Therefore, psychophysiological assessment is a key procedure for reliable diagnosis of these disorders. Psychophysiological disorders in which an etiological role played by psychological factors has been demonstrated include cardiovascular disorders (e.g., essential hypertension, cardiac arrhythmias, migraine), respiratory disorders (e.g., asthma, hyperventilation), gastrointestinal disorders (e.g., ulcer, irritable colon, colitis, chronic constipation), neuromuscular disorders (e.g., tics, tremblers, tension headache), dermatological disorders (e.g., eczema, acne, psoriasis), sexual disorders (e.g., dysfunction of sexual desire, excitation, and orgasm), and neurological disorders (e.g., sleep disturbances, epilepsy).

## 7.3. Neuropsychological Disorders

Neuropsychological disorders are characterized by symptoms of psychological dysfunction caused by a diagnosable brain lesion. The psychological dysfunction is usually a cognitive one such as altered perception (agnosias), memory (amnesias), language (aphasias), or motor coordination (apraxias). The neural lesion is usually a stroke, a traumatic brain lesion, a tumor, or a degenerative neural disease.

Although neuropsychological assessment does not normally use psychophysiological methods, recent attempts to combine both methods are promising. For example, ERPs are being used to assess cognitive deficits in Alzheimer's disease, epilepsy, alcoholism, and mental retardation. Similar applications of functional

neuroimaging techniques are helping to better understand the specific brain dysfunctions produced by different lesions.

## Acknowledgments

The preparation of this article was partially supported by Research Grant BSO2001-3211 from the Spanish Ministry of Science and Technology and by the Research Group HUM-388 of the Junta de Andalucía.

## See Also the Following Articles

Clinical Assessment ■ Neuropsychological Assessment in Schools ■ Psychological Assessment, Standards and Guidelines for ■ Psychological Rehabilitation Therapies

## Further Reading

- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.). (2000). *Handbook of psychophysiology*. Cambridge, UK: Cambridge University Press.
- Cuthbert, B. N., Lang, P. J., Strauss, C., Drobles, D., Patrick, C. J., & Bradley, M. M. (2003). The psychophysiology of anxiety disorder: Fear memory imagery. *Psychophysiology*, *40*, 407–422.
- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (Eds.). (2002). *Cognitive neuroscience* (2nd ed.). New York: Norton.
- Graham, F. K. (1979). Distinguishing among orienting, defense, and startle reflexes. In H. D. Kimmel, E. H. Van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 137–167). Hillsdale, NJ: Lawrence Erlbaum.
- Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50*, 372–385.
- Obrist, P. A. (1981). *Cardiovascular psychophysiology: A perspective*. New York: Plenum.
- Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 533–575). Cambridge, UK: Cambridge University Press.
- Öhman, A., & Soares, J. J. F. (1994). Unconscious anxiety: Phobic responses to masked stimuli. *Journal of Abnormal Psychology*, *103*, 231–240.
- Posner, M. I. (1999). *Education and the human brain*. Granada, Spain: University of Granada Press.



# Psychotherapy in Older Adults

Bob G. Knight and Steven David

University of Southern California, Los Angeles, California, USA

1. Presenting Problems
  2. Assessment, Diagnosis, and Prevalence
  3. Interventions
  4. Diversity
  5. Conclusions
- Further Reading

## GLOSSARY

**cohort based** A focus on generational groups rather than age groups; examples include the World War II generation, the baby-boomer generation, and Generation X.

**contextual** The idea that changes in therapy often relate to the social–environmental context within which older adults reside rather than their developmental stage; this is especially apparent in settings such as hospitals and nursing homes.

**contextual, cohort-based, maturity, specific challenge (CCMSC) model** Model that proposes that older adults are more mature than younger adults in certain important ways and yet are concurrently facing some of life's most difficult challenges; the model also asserts that the current social context and past sociocultural circumstances experienced as members of earlier born cohorts must be considered when conceptualizing psychotherapy with older adults.

**loss–deficit model** A negative conceptual model of aging that portrays late life development as the accumulation of deficits and a series of losses as the normative experience of older adults; the result is characterized by negative mental health outcomes.

**maturity** The developmental maturation that occurs during normal aging; although this maturation includes changes such as cognitive slowing and use of simpler language, it also includes greater emotional complexity and a greater source of expertise from which to draw.

**specific challenges** The idea that the high prevalence of chronic medical problems and neurological disorders among older adults will result in the frequent presentation of these problems in assessment and therapy; grief for deceased loved ones and caregiving for other older adults (e.g., spouses, parents) are also specific challenges faced by older clients.

This article integrates discussion of psychotherapy with current research and clinical observations in the field of gerontology. It briefly summarizes past and current research on aging in several areas of practical interest to psychotherapists seeking to apply various psychotherapy treatment techniques in working with older adult clients. The framework for this discussion is the contextual, cohort-based, maturity, specific challenge (CCMSC) model, which grew from the application of gerontology to the practice of psychotherapy with older adults. This model is intended to be nonspecific with regard to system and modality of therapy and, thus, affords a wide utility. The article discusses common presenting problems, issues of assessment and diagnosis, and prevalence rates relevant to psychotherapy with older adults. It reviews the literature concerning effective psychotherapeutic interventions across systems and

modalities in an effort to report the state of the art. It also addresses issues of diversity and the importance of cultural competence among psychotherapists who treat older adults.

## **1. PRESENTING PROBLEMS**

Older adults seeking therapy often are dealing with problems that would threaten psychological well-being at any age in the life span. These problems include dealing with chronic illness and disability, grieving the loss of loved ones, and caregiving for loved ones with cognitive or physical disability. Although these problems are not unique to later life, they are more likely during this period. Additional problems seen earlier in the life span, such as interpersonal relationship difficulties and personal failures, appear later in the life span as well. Finally, many people who have struggled with depression and anxiety in earlier life continue to struggle with these problems in later life.

Working with distressed older adults often means working with those who are struggling to adjust to the problems associated with being chronically ill or physically disabled. Thus, therapists must acquaint themselves with chronic illnesses and their psychological impact, pain control, adherence to medical treatment, strategies of rehabilitation, and assessment of behavioral signs of medication reactions. The increased presence of chronic illness and disability with each decade of life makes it necessary for therapists working with older adults to develop the ability to discuss physical problems and to discuss when a problem may have physical causes.

Similarly, working with older adults often involves grief work. Although the death of loved ones can be experienced at any point in the life span, this experience is more common in later life. Older adults presenting with depression often have experienced the death of loved ones during recent months or years. In addition, many older adults do not experience the death of loved ones as normal or expected. This is the case even when their loved ones have been ill for some time. These losses are still experienced as surprising and tragic, and they may be experienced more deeply due to the length of their relationships. As a result, psychotherapy with older adults may often require grief work.

Finally, older adults often find themselves in the role of caregivers for physically or cognitively impaired spouses or parents. The course of these physical and

cognitive disorders is frequently progressive and debilitating. As such, older adult caregivers are forced to adapt to increasingly difficult caregiving tasks and, consequently, may begin to appraise caregiving as a burden. Thus, therapists working with older adult caregivers must be prepared to treat the stress resulting from this burden.

## **2. ASSESSMENT, DIAGNOSIS, AND PREVALENCE**

Prevalence studies of depression in older adults who are chronically ill or physically disabled have reported rates up to 59%. Further examination of the relationship between depression and physical illness finds that the disability often accompanying physical illness, rather than the illness itself, is a risk factor for depression in older adults. Conversely, depression can be seen as a serious physical and psychological health problem because it has been found to be associated with increases in disability, poorer rehabilitation from physical health problems, and greater risk of mortality. Physicians frequently do not detect symptoms of depression in their older adult clients, and symptoms of depression in older adults often go untreated. This problem is exacerbated by the fact that older adults often present symptoms of depression to their primary care physicians rather than attempting to seek out mental health professionals.

Depression can also be a problem for older adults suffering from Alzheimer's disease or other forms of dementia. Thus, the issue of screening clients with depression for dementia is an important one when working with older adults. Conversely, because depression and dementia may present similar symptoms, clients suspected of having dementia should be screened for depression. In addition, clinicians should be aware of possible cognitive impairment and should include a brief screening instrument, such as the Folstein Mini-Mental State Exam, as a matter of routine when assessing older adult clients. Signs of cognitive impairment on initial screening indicate the need for more extensive cognitive assessment and may necessitate adjustment of treatment goals and intervention style.

Although depression is prevalent in older adults who experience chronic illness with disability, overall research contradicts the stereotype of the older adult as frequently depressed. In reality, depression is less

prevalent in older adults than in younger adults. However, it is important to note that although prevalence of major depressive disorder appears to be low among older adults, studies of self-reported depressive symptoms show increases among older adults. As such, care should be taken to assess both clinical and subclinical depression in older adults. Similarly, multiple symptoms of anxiety may exist among older adults even in cases where a constellation of symptoms meeting diagnostic criteria for anxiety disorders is not present. Regarding older adults who do indeed meet diagnostic criteria, the combined prevalence of all anxiety disorders in adults over 65 years of age is approximately 5.5%.

Although prevalence rates of alcoholism for older adults are lower than those for younger adults, there remain many serious concerns involving alcohol and older adults. In addition to the myriad of psychosocial ramifications of alcohol addiction, other problems include increased risk of falling, potential liver damage, adverse medication interactions, and decreased cognitive functioning. Although some older adult alcoholics have been addicted since younger adulthood, others develop problems with alcohol in older adulthood in response to inadequate social support, depression, and loneliness. This occurs at higher rates for older adults than for younger adults.

Finally, insomnia should be noted as a common problem among older adults. Studies have found that 10 to 25% of adults over 65 years of age report experiencing chronic sleep difficulties. Sleep disturbances may influence daily functioning and, thus, should be included as part of standard assessment practice with older adults.

### 3. INTERVENTIONS

#### 3.1. Cognitive–Behavioral Therapy

Empirical evidence regarding the effectiveness of psychological interventions with older adults has been collected over the past two to three decades. In a 1994 meta-analysis by Scogin and McElreath of mainly cognitive–behavioral psychological interventions in the treatment of depression in older adults, an aggregate effect size was shown to be roughly equal to that found in another 1994 meta-analysis by Schneider for the efficacy of antidepressant medications.

The factors affecting outcomes with older adults were explored in a 2001 meta-analysis by Pinquart and

Sörensen. These include psychotherapy as superior to psychosocial interventions, more than nine sessions of therapy, and specialized training of the therapist. Older adults with clinical depression improved more than those considered to be at risk for depression. It was found that individually based interventions for both depression and caregiver distress were generally more efficacious than group-based interventions. A 2003 study by Wetherell and colleagues of treatment of generalized anxiety disorder found that brief cognitive–behavioral therapy (CBT) had large beneficial effects, and a discussion group treatment had medium beneficial effects, among older adults with generalized anxiety disorder. These results support the effectiveness with older adults of professional therapy by well-trained providers.

In an effort to establish a set of standards for evaluation of therapy effectiveness, the American Psychological Association's Division 12 has developed guidelines to identify empirically supported treatments for specific psychological problems. In 1998, Gatz and colleagues used the then current guidelines in their evaluation of the literature on psychological treatments with older adults. They reported that efficacious treatments for older adults included CBT for sleep disorders and cognitive and behavioral treatments for clinical depression. Gatz and colleagues also found strong support for the effectiveness of behavioral interventions and environmental modifications for persons with dementia.

With regard to subclinical problems in later life, Gatz and colleagues also found that CBT had above-average effects on self-ratings of subjective well-being. In support of these findings, Pinquart and Sörensen's meta-analysis of psychotherapeutic interventions with older adults found that CBT had above-average effects on depression and other self-ratings of subjective well-being. In addition, Gatz and colleagues reported that for nonsyndromal problems of aging, cognitive training is often efficacious in slowing cognitive decline. Therefore, there is considerable evidence for the effectiveness of CBT with older adults.

#### 3.2. Other Individual Therapies

The use of a short-term manualized form of psychodynamic psychotherapy (PDP) has also been studied among older adults. In a 1982 study comparing CBT with PDP among depressed older adults, Gallagher and Thompson found that PDP and CBT were equally effective at posttest. However, this effect did not hold up

for PDP at follow-up. Findings from Schneider's study support the use of a PDP variation, interpersonal counseling (IPC), with older adults. Gatz and colleagues' review of empirically validated treatments examined several studies that included brief psychodynamic therapy as a treatment condition for depression among older adults and found it to be equally as efficacious as CBT. As such, brief psychodynamic therapy can be considered an often efficacious treatment for depression in late life. In addition, Gatz and colleagues concluded that the technique of structured life review or reminiscence is often efficacious for treating depressive symptoms among older adults. However, it is noted that several studies suggest that when life review is compared with other methods of treatment, the alternative methods are more efficacious.

### **3.3. Family and Caregiving Interventions**

Some of the most salient issues in psychotherapy with older adults appear within the family context. Older adults often depend on family members for both emotional and instrumental support. In the event that these family relationships become strained, disruption of support can ensue, resulting in distress for older adults. Exploration of older adults' family histories and interpersonal relationships can often inform the therapeutic process in such cases. In 2000, Qualls argued that although few empirical studies examining the efficacy of family therapy for older adults have been conducted, a solid rationale for its use with this population exists. Prominent in this rationale is the assertion that the lives of many older adults are essentially embedded within the functional context of family. Similarly, the contextual, cohort-based, maturity, specific challenge (CCMSC) model can be applied to the adaptation of family therapy for older adults. This adaptation includes the examination of specific challenges facing older relatives and the family system, the integration of cross-cohort views when working across a multiple-generation family structure, and the influence of age-related social context on older adult family members.

Sørensen and colleagues' meta-analysis of interventions with caregivers of older adult family members with physical or dementing illnesses found that these interventions were generally effective in reducing burden and depression, increasing subjective well-being, and increasing caregiver mastery. It is important to

note that although these results were significant, the effect sizes were, on average, small to moderate. These appear to be lower than the previously mentioned effect sizes for the treatment of depression among older adults. In that study, psychoeducational and psychotherapeutic interventions consistently produced short-term effects on outcomes. The majority of these effects remained intact after an average of 7 months postintervention. It was also found that intervention effects were smaller for dementia caregivers than for caregivers of family members with other illnesses.

## **4. DIVERSITY**

### **4.1. Ethnicity**

The CCMSC model asserts that the majority of diversity issues in psychotherapy with older adults can be considered as cohort and contextual differences among clients. These differences are apparent when ethnicity, gender, and sexual identity are considered. Cohort differences are particularly salient in relation to ethnicity. Ethnic minorities belonging to earlier born cohorts have had to contend with especially difficult experiences of racism and economic hardship. In addition, social context may vary by ethnicity. In some ethnic groups, the use of large intergenerational family structures and an increased emphasis on familism is prominent. It is important for clinicians to become sensitive to potential ethnicity-related contextual or cohort differences. Some of these differences may include verbal and nonverbal behavior and communication styles that differ according to the level of acculturation. Thus, rather than making assumptions about clients, it is incumbent on clinicians to respectfully encourage clients to openly define for themselves the salience of their ethnicity to their issues in therapy.

### **4.2. Gender**

In addition to contextual and cohort differences related to ethnicity, gender-related differences should be considered. The majority of older adults are women, and women tend to be overrepresented among therapy clients. Earlier born cohorts of women have seen the metamorphosis of women's roles in society over time, yet many may find that their personal experiences differ from those of later born cohorts. Many of these women may approach therapy with perceptions that are unique to their cohorts. In addition, because many

women find themselves surviving their husbands in later life, they may suddenly be faced with role transitions that are presenting themselves for the first time. Therefore, clinicians must be sensitive to the unique challenges faced by women in later life.

### 4.3. Sexual Identity

Finally, differences related to sexual minority status should be considered. Cohort differences among lesbian, gay, bisexual, and transgender (LGBT) older adults reveal differences between earlier born cohorts coming of age prior to the 1969 Stonewall riots (and the gay rights movement that followed) and those cohorts whose members matured later. However, some of these LGBT older adults may have been involved in earlier smaller movements that supported rights for sexual minorities. Thus, although many of these earlier born cohorts may have experienced a large portion of their lives under the weight of secrecy and shame, others may have experienced a more publicly expressed identity. Nevertheless, a range of LGBT older adults have endured the tremendous stress of both personally and socially experienced stigmatization. As with ethnicity, it is important for clinicians to refrain from developing assumptions regarding LGBT older adult clients. Instead, clinicians should allow these clients to define for themselves the salience of their sexual minority status to issues in therapy.

## 5. CONCLUSIONS

The CCMSC model proposes that older adults are more mature than younger adults in certain important ways and yet are concurrently facing some of life's most difficult challenges, including adjusting to chronic illness and disability in addition to frequently grieving the loss of others. The model also asserts that the current social context and past sociocultural circumstances experienced as members of earlier born cohorts must be considered when conceptualizing psychotherapy with older adults. Older adults seeking therapy often are dealing with problems that include coping with chronic illness and disability, grieving the loss of loved ones, and caregiving for loved ones with cognitive or physical disability. Although these problems are not unique to later life, they are more frequent during this period. Prevalence studies indicate the presence of depression and anxiety disorders among older adults, albeit at lower rates than those among younger adults.

Nevertheless, it is important to note that although rates of clinical depression decline in later life, rates of self-reported symptoms of depression increase. Thus, clinicians should attend to the possible presence of subclinical depression among older adults. In addition, issues of alcoholism and sleep disorders appear during older adulthood.

Research on empirically validated treatments has shown that efficacious interventions among older adults include (a) CBT, IPT, brief psychodynamic therapy, and life review for depression; (b) CBT for increases in subjective well-being; (c) CBT for sleep disorders; (d) cognitive treatment to slow nonsyndromal cognitive decline; and (e) psychoeducational and psychotherapeutic interventions for caregiver distress.

Finally, in keeping with the CCMSC model, contextual and cohort-based issues should be considered regarding the diverse experiences of ethnicity, gender, and sexual identity among older adults. Clinicians must be sensitive to this diversity by setting aside assumptions and encouraging the expression of diversity throughout clients' therapeutic process.

## Acknowledgment

The second author was supported in part by a National Institute on Aging multidisciplinary training grant.

## See Also the Following Articles

Age-Related Issues among Minority Populations ■ Aging and Competency ■ Anxiety Disorders in Late Life ■ Depression in Late Life ■ Personality and Emotion in Late Life

## Further Reading

- Gallagher, D. E., & Thompson, L. W. (1982). Treatment of major depressive disorder in older adult outpatients with brief psychotherapies. *Psychotherapy: Theory, Research, and Practice*, 19, 482-490.
- Gatz, M., Fiske, A., Fox, L. S., Kaskie, B., Kasl-Godley, J. E., McCallum, T. J., & Wetherell, J. L. (1998). Empirically validated psychological treatments for older adults. *Journal of Mental Health and Aging*, 4, 9-46.
- Knight, B. G. (2004). *Psychotherapy with older adults* (3rd ed.). Thousand Oaks, CA: Sage.
- Pinquart, M., & Sörensen, S. (2001). How effective are psychotherapeutic and other psychosocial interventions with older adults? A meta-analysis. *Journal of Mental Health and Aging*, 7, 207-243.

- Qualls, S. H. (2000). Therapy with aging families: Rationale, opportunities, and challenges. *Aging & Mental Health, 4*, 191–199.
- Schneider, L. S. (1994). Meta-analysis from a clinician's perspective. In L. S. Schneider, & C. F. Reynolds, III (Eds.), *Diagnosis and treatment of depression in late life: Results of the NIH Consensus Development Conference* (pp. 361–373). Washington, DC: American Psychiatric Association.
- Scogin, F., & McElreath, L. (1994). Efficacy of psychosocial treatments for geriatric depression: A quantitative review. *Journal of Consulting and Clinical Psychology, 62*, 69–73.
- Sörensen, S., Pinquart, M., & Duberstein, P. (2002). How effective are interventions with caregivers? An updated meta-analysis. *The Gerontologist, 42*, 356–372.
- Wetherell, J. L., Gatz, M., & Craske, M. G. (2003). Treatment of generalized anxiety disorder in older adults. *Journal of Consulting and Clinical Psychology, 71*, 31–40.



# Racial and Ethnic Minorities, Counseling of

Edward A. Delgado-Romero

Indiana University, Bloomington, Indiana,  
USA

Debra Mollen

Texas Women's University, Texas, USA

Charles R. Ridley

Indiana University, Bloomington, Indiana,  
USA

1. Introduction
  2. Motivation for Counseling Racial and Ethnic Minorities
  3. Acculturation and Counseling
  4. Unintentional Racism/Aversive Racism
  5. Racial Identity
  6. Multicultural Counseling Competence
- Further Reading

*race* A group of humans identified as distinctive by physical characteristics such as skin color.

*racial identity* The view one has about the role of race in dictating social group membership.

*racism* Fear, hatred, and oppression solely based on the perceived racial or ethnic characteristics; may be intentional or unintentional.

## GLOSSARY

*acculturation* A process of attitudinal and behavior change experienced by individuals who have come in contact with a different and dominant culture.

*competence* The ability to predetermine and then to attain the desired outcomes of one's profession or vocation by coordinating and integrating relevant skills and microskills.

*counseling* The purposeful development and structuring of a professional relationship of which the intention is to facilitate meaningful changes in the lives of its recipients. The focus is typically on normal and developmental aspects of functioning as opposed to psychopathology.

*culture* The ideals, beliefs, skills, customs, values, and social behavior of a particular group of people.

*ethnicity* A group of humans identified as distinctive by cultural practices and/or beliefs.

*multicultural* Relative to two or more groups or cultures.

Counseling racial and ethnic minorities entails the provision of counseling services (including individual and group therapy, consultation, and research) to members of racial and ethnic minorities in the United States. In response to societal and demographic changes, as well as a growing ethical awareness, the need to address completely the unique needs of racial and ethnic minorities has been a powerful influence in counseling psychology. Of particular importance are the relationship of acculturation to counseling, the influence of unintentional or aversive racism in counseling, racial identity in regard to counseling, and especially the overarching issue of what constitutes multicultural counseling competence.

## 1. INTRODUCTION

Historically, the field of applied psychology has been negligent in its delivery of services to racial and ethnic

minorities. The assumptions that have guided practice throughout the history of psychology have relied on White (English-speaking, European-descended, heterosexual) males as the standard for normalcy. Persons from other backgrounds (i.e., race, ethnicity, socioeconomic status, and sexual orientation) were conceptualized, regarded, and treated as deviants; systematically denied access to quality services; and often discounted and ignored by psychology professionals. As the United States entered the Civil Rights era in the 1950s, the injustice of ignoring race and culture in counseling psychology became apparent. In 1962, Wrenn coined the phrase “cultural encapsulation” to describe counselors who defined a client’s reality according to their own cultural assumptions and stereotypes, minimized cultural differences in favor of universal explanations, disregarded their own cultural biases, conceptualized other cultures in terms of pathology, and used a technique-oriented model of the counseling process.

Gradually, partially in response to the Civil Rights movement, applied psychology grew more reflective as a specialty. Practitioners, researchers, and scholars began to recognize the extensive harm caused by the profession’s racial and ethnic biases and prejudices. Professionals began to observe and identify biases in their own approach to minority clients. It was clear that applied psychology was far from being a value-neutral profession.

In what has commonly been referred to as psychology’s “fourth force,” multiculturalism has assumed a position of significant impact in the past two decades, prompting psychologists to revisit and repair their damaging beliefs and practices as well as to make a concerted effort to ensure that counseling services are accessible and amenable to racial and ethnic minority clients. Concurrent with the new awareness has come a growing commitment to training members of racial and ethnic minority populations as counselors.

Progress in the field has been hampered by unclear definitions of the constructs of race and ethnicity and vague descriptions of who constitutes a minority. Race and ethnicity have been criticized as being unscientific and socially constructed categories that are inadequately defined and operationalized. Often, issues of race and ethnicity are confounded by other factors, such as socioeconomic status. In addition, the term multicultural has been criticized as too general and serving to obscure unresolved racial issues.

## **2. MOTIVATION FOR COUNSELING RACIAL AND ETHNIC MINORITIES**

In addition to the historical and societal forces identified previously, two main motivations for the competent counseling of racial and ethnic minorities have been identified: demographic change and ethical concerns. In terms of demographics, the United States has seen a steady increase in some minority populations. The U.S. Census has documented the changing demographic composition of the United States. Census predictions suggest that by 2050 Whites will no longer represent a majority of the U.S. population; in some states, such as California, this change is already a demographic reality. The increase of racial and ethnic minorities, however, does not automatically yield economic success, access to education, or political empowerment. On the contrary, although in some areas people of color are no longer a statistical minority, true economic, educational, and political equality for them remains an ideal to which many aspire rather than a reality that is practiced.

Both psychology’s and counseling’s ethical guidelines have reflected the evolution toward a more culturally sensitive applied profession, reflecting the theme of social justice. The most recent version of each set of guidelines takes culture—defined broadly to include age, gender, race, ethnicity, national origin, religion, sexual orientation, disability, language, and socioeconomic status—into account. The guidelines impart the expectations that counselors and psychologists be adequately prepared to treat clients from diverse backgrounds, that they do not practice unfair discrimination, and that they consider culture in the practice of applied psychology. However, professional ethics codes and standards of conduct—believed by many to be an improvement over the earliest versions, in which consideration of culture was absent—still meet with criticism by many scholars. Critiques tend to focus on the underpinnings of the codes, which are individualistic, based on principle ethics, and therefore problematic when viewed through a collectivistic lens; on the vagueness of the standards and practices, which fail to give practitioners sufficient direction; and on the absolutist assumption, which fails to give culture satisfactory emphasis. Considerable work must be done before the profession can ensure that racial and ethnic minority clients receive ethically appropriate treatment from culturally sensitive, competent practitioners.

### **3. ACCULTURATION AND COUNSELING**

Acculturation can be defined as a process of attitudinal and behavioral change experienced by individuals who live in multicultural societies or who have come in contact with a different culture due to colonization, invasion, political change, globalization, and the increased mobility of society due to technological advances. Acculturation refers to the balance between changing attitudes and behaviors as a result of contact with a dominant group and retention of existing cultural values, beliefs, and traditions. In the United States, institutions such as the educational system and the media are a part of the acculturation process. Another component of the acculturation process may include the acceptance of the need for counseling. The degree of acculturation to a dominant culture may also play a role in preference for the ethnicity of one's counselor or the client's willingness to discuss racial, ethnic, or cultural issues. For example, a client who has low levels of acculturation may not view counseling as an option and instead prefer to rely on family or other culturally sanctioned sources of support. On the other hand, a highly acculturated individual may be very receptive to counseling and may even prefer to work with a White counselor.

### **4. UNINTENTIONAL RACISM/ AVERSIVE RACISM**

In a cultural zeitgeist of increased awareness about the damaging effects of racial and ethnic prejudice and discrimination, overt acts of racism have abated to some extent. It would be erroneous to conclude, however, that decreased incidents of blatant racism indicate an absence of racial and ethnic prejudice. Instead, as public and professional awareness about racism increases, more covert examples of racism are evident. These have been termed aversive and unintentional racism.

Aversive racism occurs when individuals who identify themselves and others as egalitarian and unprejudiced subtly discriminate against racial and ethnic minorities in their beliefs and feelings. This may occur when such prejudice can be rationalized or when information the individual has received is ambiguous. In 1995, Ridley identified eight methods of unintentional but pervasive racist defenses that occur

in counseling. Color blindness is the tendency for counselors to disregard the race of their clients, assuming that all clients are similar. They may minimize or deny the harm inherent in racism, prejudice, and discrimination. Color consciousness, on the other hand, occurs when counselors place too much emphasis on race, ignoring pathology or attributing all presenting problems to culture. Cultural transference can occur when racial and ethnic minority clients project their feelings about the majority culture onto their counselor, whereas cultural countertransference indicates the counselor's projection of emotional reactions toward a minority group onto a particular minority client. Cultural ambivalence can occur when counselors use the therapy arena to act out their own feelings about race and ethnicity, either by dominating and controlling or by seeking approval and acceptance from the client. Pseudo-transference refers to the tendency for counselors to discount their clients' perceptions of them as racist and prejudiced; the clients' reactions to their counselors may be casually dismissed as transference when actually rooted in reality. Overidentification occurs when minority counselors relate significantly with their minority clients and may unwittingly focus on racial issues to the neglect of other important presenting problems. Finally, identification with the oppressor refers to the tendency for some racial and ethnic minority counselors to deny their experiences as minorities and overidentify instead with the majority culture. Ridley offered counselors a variety of ways to combat unintentional racism.

### **5. RACIAL IDENTITY**

The ways individuals perceive themselves as racial beings form the processes by which people acquire their racial identity, and thereby influence both the counseling process and the selection of counselors. Researchers have proposed several different models that outline the distinctive processes that formulate racial identity.

Phinney based her 1993 model of ethnic identity development primarily on her work with adolescents. She proposed a three-stage model to capture the process by which adolescents formulate their ethnic identity. The first stage, unexamined ethnic identity, represents a relatively static condition in which individuals are not in the process of exploring themselves as

ethnic individuals. In the second stage, ethnic identity search/moratorium, individuals are involved in actively exploring themselves ethnically and trying to make meaning of their ethnicity. In the third stage, achieved ethnic identity, individuals have a clear understanding of themselves in terms of their ethnic identity.

Updating Cross's earlier work on Black identity, Cross and Vandiver in 2001 proposed a four-stage model to elucidate racial identity formulation, referred to as the expanded nigrescence model. In stage 1, preencounter, individuals have not yet begun to explore themselves as racial beings and may even denigrate their own culture. Eventually, individuals encounter an experience that prompts them to begin to explore themselves as racial people; this experience is referred to as an encounter experience. This can include an incident involving racism that is experienced personally or having a powerful response to a cultural event, historical incident, or other racial experience. In stage 2, immersion–emersion, individuals experience the most pronounced period of transition. As they continue to concretize their racial identity, they may embrace all things affiliated with their racial or ethnic group and they may eschew those things affiliated with the dominant culture. In stage 3, internalization, individuals have worked through a significant portion of the transition period and may embrace their racial identities as one facet of themselves while simultaneously considering multiple salient roles. In stage 4, internalization–commitment, individuals sustain their interest in their racial identity and may choose to devote themselves to promoting the needs and interests of their particular ethnic group. Cross and Vandiver note that there are cluster identities within each stage, and these cluster identities acknowledge that within stages individuals will experience identity development differently.

Therefore, according to the stage of their racial identity development, racial minority clients may have different views about the purpose of counseling, the preferred race of their therapist, and what they want to work on in counseling. For example, a client in the preencounter stage may deny the salience of race and endorse “color-blind” views of the United States. A client in the immersion–emersion stage may need to explore actively the meaning of race and struggle to make sense of an encounter experience. However, it is not only the client who undergoes racial identity development; the counselor also has a racial identity and undergoes development. The interaction of the two identities can be a powerful dynamic in therapy.

Cross and Vandiver's model may also be generalized beyond Black racial identity and serve as an exemplar of the racial or ethnic identity process.

Janet Helms has utilized her models both on White racial identity and on people of color racial identity to create an interpersonal process model that may be particularly useful for counselors seeking a greater understanding of the dynamics that occur in cross-racial dyads. Helms posited that it is the expressed racial identity of the individuals involved in a particular dialogue that is most critical. In parallel interactions, individuals share similar attitudes and assumptions and generally try to maintain accord. In regressive interactions, the person afforded the most power socially functions from a less evolved status than the person with less power. In progressive interactions, the individual with more social power responds from a relatively more evolved status than the person with less social power. Finally, in crossed interactions, participants interact from places that are diametrically opposite in terms of responding to racial matters.

## **6. MULTICULTURAL COUNSELING COMPETENCE**

During the past two decades, significant attention has been focused on the notion of multicultural counseling competence. The profession of applied psychology has particularly endorsed the need for practitioners who are multiculturally competent.

One such approach originally proposed by Sue and colleagues in 1992 consisted of creating guidelines for multiculturally competent practitioners. This entailed specific directives based on a tripartite model of knowledge, skills, and attitudes. The focus is on the acquisition and endorsement of those facets of competence endorsed by the authors and supported by the profession at large. In this view, multicultural counseling competence consists of professionals who are actively open to examining and adjusting their own assumptions and worldviews, who work toward nonjudgmentally understanding their clients' viewpoints, and who practice appropriate and sensitive strategies with clients from diverse racial and ethnic backgrounds. Many instruments have been created in response to the tripartite considerations of knowledge, skills, and attitudes.

Another approach by Ridley and Mollen recognizes the importance of the aforementioned competencies model and endorses a model that embraces an

integrative deep structure as well—a structure that involves such facets of practice as purposeful application, motivation, selection, sequencing, and timing. This approach posits that multicultural counseling competence consists not only of the building blocks of therapy called microskills and competencies but also of the purposeful, committed intention of the counselors to create an environment suitable for meaningful and relevant client change in consideration of racial, ethnic, and cultural variables.

### See Also the Following Articles

Acculturation ■ Conformity across Cultures ■ Cross-Cultural Psychology, Overview ■ Health Psychology, Cross-Cultural

### Further Reading

- American Psychological Association. (2003). Guidelines on multicultural education, training, research, practice and organizational change for psychologists. *American Psychologist*, 58, 377–402.
- Cross, W. E., & Vandiver, B. J. (2001). Nigrescence theory and measurement: Introducing the Cross Racial Identity Scale (CRIS). In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (2nd ed., pp. 371–393). Thousand Oaks, CA: Sage.
- Helms, J. E. (1984). Toward a theoretical explanation of the effects of race on counseling: A black and white model. *The Counseling Psychologist*, 12, 153–165.
- Helms, J. E. (1995). An update of Helm's white and people of color racial identity models. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 181–191). Thousand Oaks, CA: Sage.
- Phinney, J. (1993). A three-stage model of ethnic identity development. In M. Bernal, & G. Knight (Eds.), *Ethnic identity: Formation and transmission among Hispanics and other minorities* (pp. 61–79). Albany: State University of New York Press.
- Ponterotto, J. G., Casas, J. M., Suzuki, L. A., & Alexander, C. M. (Eds.). (2001). *Handbook of multicultural counseling* (2nd ed.). Thousand Oaks, CA: Sage.
- Ridley, C. R. (1995). *Overcoming unintentional racism in counseling and therapy: A practitioner's guide to intentional intervention* (Multicultural aspects of counseling series, Vol. 5). Thousand Oaks, CA: Sage.
- Sue, D. W., Arredondo, P., & McDavis, R. J. (1992). Multicultural counseling competencies and standards: A call to the profession. *Journal of Counseling and Development*, 70, 477–483.
- Sue, D. W., & Sue, D. (1999). *Counseling the culturally different* (3rd ed.). New York: Wiley.



# Rape Prevention

Barbara Krahé

University of Potsdam, Potsdam, Germany

1. Rape: Definition and Prevalence
  2. Prevention Measures Targeting Individuals
  3. Prevention Measures at the Societal Level
  4. Conclusion
- Further Reading

## GLOSSARY

**rape** Oral, anal, or vaginal penetration through threat or use of force or intentional incapacitation.

**rape myth acceptance** Tendency to accept as true statements about rape and rape victims that imply victim blame and are not supported by data (e.g., “Every woman has the secret desire to be raped”).

**rape prevention beliefs** Perceptions of personal risk and of the effectiveness of different preventive strategies.

**rape shield laws** Legal restrictions limiting the admission of evidence about complainant sexual history.

**rape trauma syndrome** A distinct pattern of psychological and physical symptoms characteristic of victims of sexual assault.

**real rape stereotype** Idea that the genuine rape is a surprise attack in a dark alleyway by an assailant who is a stranger to the victim and uses force during the assault.

Rape is a pervasive problem in societies around the world and causes severe suffering and trauma to victims. The vast majority of rapes are committed by men against women, but male-on-male rape and female sexual aggression against men have also been recognized

during recent years, as reflected in changes in legal definitions as well as in research attention. Therefore, interventions designed to reduce the rate of rape are of paramount significance. Most interventions are directed at individuals, targeting either men as potential perpetrators or women as potential victims. Systematic evaluations of these programs show that their success is limited. In addition, measures are implemented at the societal level to enforce more rigorous legal prosecution of sexual offenses and to challenge traditional gender stereotypes that are conducive to rape.

## 1. RAPE: DEFINITION AND PREVALENCE

In a worldwide survey of sexual violence published in 2002, the World Health Organization defined rape as physically forced or otherwise coerced penetration—even if slight—of the vulva or anus, using a penis, another body part, or an object. Although this definition provides a common basis for documenting the prevalence of rape across a range of countries, legal definitions of the offense show some variations in the key elements in the definition of rape. In the United States, for example, the Federal Bureau of Investigation (FBI) adopts a definition of rape that is restricted to forcible vaginal intercourse with a woman without her consent by force or threat of force. In contrast, most federal states have adopted legislation in which rape is defined more broadly in terms of oral, anal, or vaginal

penetration through threat or use of force or intentional incapacitation. There is no longer a restriction of legal definitions to extramarital acts of sexual aggression, thereby instituting “marital rape” as a criminal offense. Similar legal definitions can be found in other Western countries. In Britain, rape is legally defined as sexual intercourse (i.e., vaginal or anal penetration with the penis) without the other person’s consent and where the defendant knows that the other person does not consent or is reckless as to whether that person is consenting or not. In Germany, the current legal definition includes the use or threat of violence or the exploitation of the victim’s incapacitated state to force another person into sexual activities. Forced sexual activities qualify as rape if the act involves penetration of the victim’s body, and this is no longer restricted to penile–vaginal intercourse.

Official crime statistics record only those incidents of rape that are reported to the police, and there is a general consensus that they severely underestimate the true scale of rape. Nonetheless, these sources show that rape is a widespread problem. For 2001, the Uniform Crime Reports reveal that 62.2 per 100,000 women in the United States were victims of forcible rape and that a rape was committed every 5.8 minutes. In terms of lifetime prevalence, it is estimated that 15% of women in the United States become the victims of rape at some point in their lives. Figures in other Western countries are lower but still substantial, for example, a rate of 13.6 per 100,000 female citizens in Germany in 2000 (including both use/threat of force and exploitation of incapacitation).

Evidence based on community surveys is able to detect at least a proportion of cases not reported to the police and, therefore, typically yields higher prevalence rates. In 1987, Koss and colleagues published a groundbreaking survey of college students from 32 higher education institutions in the United States that found that 15.4% of all women reported they had been raped since 14 years of age. In 1999, Spitzberg integrated evidence from 120 studies with a total of more than 100,000 women and arrived at a similar figure of 12.9%. Evidence compiled in the World Health Organization report of 2002 from a range of samples in Africa, Latin American, Asia, and Eastern Europe reflect prevalence rates between 0.3% (The Philippines) and 8.0% (Brazil), but these figures include only incidents from 16 years of age onward.

The majority of rapes are committed by perpetrators known to the victims either as acquaintances or as dating partners. Despite the persistence of the “real

rape stereotype” picturing rape as a surprise attack in a dark alleyway by an assailant who uses force during the assault, sexual assaults by strangers are the exception rather than the rule. This fact has important implications for rape prevention efforts. Such efforts can be broadly classified into two groups: (a) prevention efforts targeting individuals as potential perpetrators or victims and (b) measures at the societal level designed to discourage sexual violence.

## 2. PREVENTION MEASURES TARGETING INDIVIDUALS

The majority of rape prevention efforts are directed at individuals, aiming to make men less likely to engage in sexual violence and to prepare women to anticipate and avoid sexual assault. The specific aims and contents of such measures vary depending on the target audience. Three main types of programs have been implemented and evaluated in systematic research: (a) programs directed at men only, (b) programs directed at women only, and (c) programs directed at mixed-sex audiences.

### 2.1. Programs Directed at All-Male Audiences

The purpose of these programs is to address men’s cognitive and affective predispositions toward sexual aggression. Several studies have demonstrated that greater knowledge about rape, particularly about the rape trauma syndrome, is associated with men’s lower likelihood of committing sexually aggressive acts. Measures in this category are directed at changing men’s rape-supportive attitudes, such as rape myth acceptance, and promoting information about the adverse effects of sexual assault on the victims. They typically consist of one or two sessions lasting up to 2 hours each. In terms of success, meta-analytic evidence shows that human sexuality courses, workshops, video interventions, and other formats have short-term success in reducing rape myth acceptance. However, to prevent men from acting out rape myths in sexually aggressive behavior, such intervention effects need to persist over a longer period. In addition to preventing first-time sexual assault through supplementing rape-supportive attitudes by greater victim empathy, programs need to focus on sexually aggressive men to prevent them from reoffending. This is particularly

important given the high proportion of multiple offenders among sexually aggressive men. One approach is to train men to discriminate between their arousal by consensual sexual stimuli and that by nonconsensual sexual stimuli and to teach them to activate inhibitory self-persuasion with regard to nonconsensual sexual cues. However, programs directed specifically at men with histories of sexual aggression show that this group responds less well to rape prevention efforts, even in the short term, than do unselected male audiences.

## **2.2. Programs Directed at All-Female Audiences**

This category includes measures directed at women as the most likely victims of sexual assault. Measures are designed to reduce the risk of sexual victimization by alerting women to potentially dangerous situations and behaviors and by enabling them to resist sexual assaults more successfully. Despite the fact that the ultimate responsibility for avoiding rape lies with the perpetrators, there is now a broad consensus that women should be empowered to take a more active self-protective role. This includes both provision of information about the risk factors of sexual victimization and specific advice and training on how to fight off an attack.

High levels of sexual activity, alcohol consumption in the context of sexual encounters, and the ambiguous communication of sexual intentions have been linked to an increased risk of sexual victimization across a range of studies. Women must be informed about the risks involved in these behaviors to give them the opportunity to adjust themselves to the threat of rape. Programs teaching self-defense skills and strengthening women's assertiveness are aimed at reducing women's feelings of vulnerability and increasing their self-confidence when faced with a potentially dangerous situation. However, systematic evaluations of rape prevention programs suggest that their success is limited. In particular, there is evidence showing that rape prevention programs may lead to a small reduction in the incidence of first-time assaults on women but are ineffective in reducing the risk of revictimization in women who had experienced prior sexual assault. In addition, some programs were found to increase participants' knowledge about rape but failed to affect the incidence of sexual victimization compared with a control group.

Whether or not women are willing to take behavioral precautions against the risk of sexual assault depends, to a large extent, on their "rape prevention beliefs," that is,

their perceptions of personal risk and of the effectiveness of various preventive strategies. The rape prevention beliefs of most women are restricted to the stranger rape scenario and fail to acknowledge the risk of acquaintance rape. Moreover, rape prevention beliefs are primarily concerned with responding to a sexual attack once it has occurred so as to minimize harm rather than with avoiding an assault in the first place. Women generally perceive a low personal risk of acquaintance rape and, therefore, are less likely to see a need to adopt protective behaviors against this form of threat.

An important component of rape prevention efforts directed at women is to promote women's reflection about potential protective responses in an assault situation. There is conclusive evidence suggesting that fighting back is more likely to succeed in avoiding rape than to lead to greater injury. Therefore, most researchers advise that active self-defense is a prudent response to a sexual attack. However, it is important to bear in mind that the positive evaluation of active resistance is based on probability estimates for samples of assault victims and does not guarantee that resistance is the best strategy in any single case. Therefore, authors are careful to stress that the decision for or against physically fighting back in an attack must be based, in the final instance, on an assessment of the particular circumstances of the attack.

However, it has been noted that empowering women to fight back when they are sexually assaulted involves more than providing self-defense skills. It requires a fundamental challenge of traditional gender stereotypes that portray active physical resistance as unfeminine and, therefore, as inappropriate for women. This applies in particular to rape resistance in assaults by acquaintances or dates where women may be reluctant to acknowledge men's behavior as dangerous and take protective action for fear of appearing to be rude or even paranoid. Prevention programs concerned specifically with date or acquaintance rape are scarce, and the few studies available do not reveal a substantive impact on subsequent victimization rates compared with nonintervention controls.

## **2.3. Programs Directed at Mixed-Sex Audiences**

In terms of their contents, programs directed at mixed-sexed target groups largely address the same aspects as single-sex programs, with a strong emphasis on providing information about the traumatic effects of rape

and on changing men's rape-supportive attitudes. Programs are also similar in terms of their short duration, rarely lasting more than 2 to 3 hours. If such programs are at all successful, their impact is short-lived and generally disappears within a few weeks. Although there may be practical advantages of combining male and female participants in a rape prevention program, this is outweighed by significant disadvantages. One is that men, particularly those with previous histories of sexual aggression, may be unwilling to approach their rape-supportive attitudes and coercive behaviors in an open way in the presence of women. In addition, the message—directed at female participants—that rape is prevalent and that the probability of prosecution for rapists is low may send the message to men that they can expect to get away with rape. Therefore, rape prevention experts recommend that rape prevention programs be directed at men and women separately.

#### **2.4. Problems of Evaluation**

Studies implementing and evaluating rape prevention programs are fraught with a number of conceptual and methodological problems that limit their significance. First, participation is voluntary, leading to self-selected samples of participants. In the case of men, this may lead to samples whose members are more open-minded and willing to learn and not representing men with strong rape-supportive attitudes and/or previous sexual aggression. In the case of women, self-selection may lead to the systematic underrepresentation of women who have previous histories of sexual victimization and who might find it too painful to approach the issue in a group setting. As a result, little is known about the success of rape prevention efforts in those target groups that are particularly in need of intervention. A second methodological problem concerns the operation of social desirability biases in the assessment of program success. Reports of greater victim empathy or lower support for rape myths may reflect respondents' tendencies to present themselves in a socially desirable way as much as they may indicate genuine change.

At the conceptual level, a major problem is the selection of criteria by which the success of an intervention is assessed. In the majority of programs, post-intervention measures of success are restricted to changes in cognition and affect such as reductions in rape-supportive attitudes and the promotion of victim empathy. As noted previously, significant but small

changes are typically found on these variables immediately following the intervention. Far more important from an applied perspective is (a) whether cognitive and affective changes are of a sufficient magnitude and persistence over time to be able to exert an influence on behavior and (b) whether a program is successful in reducing the rates of sexual aggression. The evidence available suggests that the answer to the first question is no. Program effects on attitudes and victim empathy typically dissipate within 2 to 4 weeks after the intervention. The answer to the second question is hampered by the fact that there are few longitudinal studies that relate program participation to subsequent reports of sexual aggression in male participants or sexual victimization in female participants. However, the limited evidence that is available suggests that rape prevention programs so far have been largely unsuccessful in reducing the incidence of rape.

### **3. PREVENTION MEASURES AT THE SOCIETAL LEVEL**

As a necessary complement to programs targeting individuals, measures in this category aim to change the sociocultural context in which sexual violence occurs. Their purpose is to increase the costs of rape for the perpetrators and to reduce opportunities for sexual assault in a society. Essential components of this approach are challenging gender stereotypes condoning aggressive behavior in men, reducing opportunities for sexual assault through policies designed to increase women's safety, and reforming the legal treatment of rape to raise the likelihood of conviction for rape offenders and improve the treatment of victims. Rape reflects a power differential between men and women that is part-and-parcel of a wider imbalance of power in society in Western culture and in most other parts of the world. Therefore, addressing this imbalance by promoting equal rights and equal participation of men and women in society will affect the perception of men's entitlement to women to gratify their sexual needs. In addition, gender stereotypes and sexual scripts placing men in the role of initiators and women in the role of gatekeepers of sexual interaction must be challenged and replaced by a more egalitarian view of how sexual intimacy should be negotiated.

Drawing on evidence that links sexual aggression to exposure to violent pornography, some have called for stricter control of accessibility of such material.

Moreover, education campaigns are needed to inform the public about the negative aspects of violent pornography with regard to promoting aggressive behavior.

### **3.1. Reducing Opportunities for Sexual Assault**

This approach includes policy decisions designed to increase women's safety at the local community level. They include practical measures such as providing free late-night transport for women, designating safe spaces for women in public automobile parking lots, installing security cameras in public places, and ensuring that public pathways are well lit at night. These measures reduce opportunities for stranger sexual assault by making it harder for potential assailants to reach their targets without the risk of being caught and arrested. They can also serve to reduce women's fear of rape and to give women a greater sense of security. However, such measures are more likely to lead to a shift of sexual assaults to less well-protected locations than to a genuine reduction in sexual violence. In addition, their effectiveness is taken at face value and has not been subjected to systematic evaluation.

Different measures are required to reduce the incidence of date and acquaintance rape. Given a conclusive database that links both the perpetration of sexual assault and the risk of sexual victimization to alcohol consumption, legal regulations prohibiting the sale of alcohol to adolescents must be enforced. Because alcohol abuse on university campuses is widespread, particularly in fraternity cultures, special campaigns directed at curbing student drinking can also be instrumental in reducing sexual violence in this group.

### **3.2. Rape Prevention through Legal Reform**

Among the most important objectives of rape prevention efforts at the sociostructural level is to increase the number of rapes reported to the police as well as the number of court cases that lead to convictions. There is no doubt that rape is a vastly underreported crime, as evidenced not least in the fact that victim surveys produce much higher incidence rates than do official crime reports. It is estimated that no more than 10 to 15% of all rapes are reported to the police. Victims' reluctance to report rapes has been attributed in large part to insensitive treatment of victims by the police.

Many victims who report their rapes perceive police response as a "second assault," particularly if their cases depart from the real rape stereotype of a stranger assault. To encourage victims of rape to come forward, confidence must be established in the treatment that victims expect to receive from police, medical professionals, and legal institutions. To address this aim, police authorities have implemented procedures designed to ensure more sympathetic treatment of victims such as assigning female officers to the job of recording victims' statements. They have also developed training programs attacking derogatory attitudes toward rape victims and providing information about the traumatizing impact of rape. Despite these efforts, the reporting rate remains low compared with evidence based on large-scale crime surveys of the general population. Even if a rape is reported to the police, the likelihood of the case going to court is low and the probability of a conviction is even lower. Fewer than 10% of reported rapes end with convictions of the defendants. One reason for the low conviction rate in rape trials, particularly for date and acquaintance rape, lies in the tendency by defense lawyers to undermine victims' credibility by referring to the women's past sexual behavior or moral conduct. Rape shield legislation has been introduced to restrict the admissibility of past sexual history evidence. However, the law allows for exceptions to the rape shield provisions that open up the possibility of victim blame and reduce the likelihood of defendants being found guilty.

## **4. CONCLUSION**

Rape prevention is a challenge that requires a multilevel approach, integrating measures directed at individuals as potential perpetrators and victims of sexual assault with macro-level approaches designed to change the sociocultural climate that is conducive to rape. The evidence reviewed in this article makes it clear that rape prevention efforts have had only limited success so far. Ultimately, a reduction in the prevalence of sexual aggression is the criterion by which the effectiveness of prevention efforts must be measured. To date, however, there is no substantive database that demonstrates the success of rape prevention programs in reducing incidence rates of sexual assault. For the time being, most prevention measures are judged on the basis of whether or not they are successful in changing the cognitive and affective antecedents of sexually aggressive behavior rather than in changing the behavior

itself. The restricted time span of most programs targeting individuals is undoubtedly a major limiting condition of their success. Given the deep-rooted attitudes and stereotypes underpinning sexual violence, it seems unrealistic to expect sustainable effects of an intervention lasting just a few hours. Moreover, the large-scale underreporting of sexual victimization is a major problem. So long as the true scale of the problem of sexual assault can neither be accurately measured nor reliably estimated, the ultimate success of rape reduction measures will remain hard to assess.

### See Also the Following Articles

Aggression and Culture ■ Sexual Behavior ■ Sexual Behavior and Culture ■ Sexual Harassment

### Further Reading

Abbey, A. A., McAuslan, P., Zawicki, T., Clinton, A. M., & Buck, P. O. (2001). Attitudinal, experiential, and situational predictors of sexual assault perpetration. *Journal of Interpersonal Violence, 16*, 784–807.

Federal Bureau of Investigation. (2001). *Uniform Crime Reports*. [www.fbi.gov](http://www.fbi.gov).

Flores, S. A., & Hartlaub, M. G. (1998). Reducing rape-myth acceptance in male college students: A meta-analysis of intervention studies. *Journal of College Student Development, 39*, 438–448.

Koss, M. P., Gidycz, C. A., & Wisniewski, N. (1987). The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of higher education students. *Journal of Consulting and Clinical Psychology, 55*, 162–170.

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva, Switzerland: World Health Organization. [www.who.int](http://www.who.int).

Rozee, P. D., & Koss, M. P. (2001). Rape: A century of resistance. *Psychology of Women Quarterly, 25*, 295–311.

Schwartz, M. D., & DeKeseredy, W. S. (1997). *Sexual assault on the college campus: The role of male peer support*. Thousand Oaks, CA: Sage.

Spitzberg, B. H. (1999). An analysis of empirical estimates of sexual aggression victimization and perpetration. *Violence and Victims, 14*, 241–260.

Yeater, E. A., & O'Donohue, W. T. (1999). Sexual assault prevention programs. *Clinical Psychology Review, 19*, 739–771.



# Reading Interventions

Laurice M. Joseph

The Ohio State University, Columbus, Ohio, USA

1. Introduction
2. Characteristics of Individuals with Reading Problems
3. Phonic Interventions
4. Whole Word Interventions
5. Repeated Readings
6. Vocabulary Development and Reading Comprehension Strategies
7. Principles of Teaching and Learning Reading
8. Targeting and Continuous Improvement of Reading Skills through Processes of Linking Assessment to Intervention Further Reading

## GLOSSARY

**alphabetic principle** Knowing that there are 26 letters in the standard English alphabet and that each letter is represented by a sound; knowing the symbols and corresponding sounds of one's own alphabetic collection (e.g. Cyrillic, Greek, and English).

**morphemes** The meaning units of words—roots, prefixes, and suffixes. With an understanding of morphological rules, a student may be able to find the root of an unfamiliar word and/or redeploy the term with alternative suffixes or prefixes.

**orthography** Conventionalized rules of letter patterns and spelling strategies (e.g., "I before E . . .").

**phonemic awareness** The alertness to and operation of discrete sounds in spoken language, including rhyme, rhyme production, sound blending, sound deletion and substitution, and sound segmentation.

**phonological recoding** The combining of phonological with orthographic knowledge about words. When a student is

presented with a printed word, he or she may undergo a process of making letter–sound associations and matching the printed word to the spoken form of the word.

**phonology** Understanding the variety of possible sounds found in a language.

Reading interventions are necessary for children who struggle to acquire reading skills. There are several reading interventions for various types of reading problems, including word-level and comprehension problems. Reading interventions that are implemented should be tested on a continuous basis to determine their effectiveness and whether they need to be modified for individual children.

## 1. INTRODUCTION

Reading is a complex act. It is complex because written language is an abstract representational system of communication. Becoming literate is not a natural process but rather one that needs to be cultivated through various experiences. Children who are reared in families in which many rich verbal interactions occur are more likely to grasp the skills necessary to become literate. A rich language experience may include verbal dialogues between parents and children that consist of extending upon statements made by children. These types of dialogues facilitate vocabulary development as well as verbal reasoning skills. Before real reading takes

place, joint storybook reading with pictures provides children with rich language experiences. Parents can engage children in a discussion about the story by asking them to retell the story, explain why certain events might have occurred, and predict what might happen next as well as other types of comprehension questions. In 1991, Sulzby and Teal found that children progress from a spontaneous storybook telling phase (i.e., looking at pictures and telling a story about the pictures in an unconventional manner) to looking at pictures and telling a story about the pictures in a conventional manner as if they were reading the printed words. Finally, they actually read the printed words on a page without relying on the pictures. Rich language activities are necessary but not sufficient to becoming literate.

There are certainly other critical skills of emergent literacy, such as concepts about print, phonemic awareness, the alphabetic principle, orthographic processing, phonological recoding, morphological understandings, and reading fluency. Acquisition of these skills may serve to prevent future reading problems in children. A description of these critical reading skills is provided.

### **1.1. Concepts about Print**

Before many children enter school, they begin to obtain concepts about print and this continues as children progress in school. Concepts about print involve skills in differentiating print from pictures, knowing that a book has a front and a back, and knowing that reading (in most Western languages) is a left-to-right and top-to-bottom process. At more advanced stages, children grasp that words written on a page can be spoken, and children know when a word and a sentence begin and end (i.e., spatial relationships between words and sentences).

### **1.2. Phonemic Awareness**

Phonemic awareness is the alertness to and operation of discrete sounds in spoken language. It is one of the most important precursor skills to acquiring word recognition. Phonemic awareness activities include rhyme, rhyme production, sound blending, sound deletion and substitution, and sound segmentation. An example of a sound blending exercise is saying the individual sounds in a word such as /c-a-t/ and having children blend the sounds together to form a whole word. Sound segmentation involves saying a word as a whole, such

as /ran/, and having children break the word apart and say each sound sequentially, such as /r-a-n/. These phonemic awareness activities should be explicitly taught to children in kindergarten and first grade according to a synthesis of research presented in the National Reading Panel Report. Among the phonemic awareness exercises, phonemic segmentation and blending were found to be the best predictors of word recognition performance in first grade.

### **1.3. Alphabetic Principle**

In 1990, Adams asserted that children enter into the alphabetic system easily if they have acquired phonemic awareness. The alphabetic principle is knowing that there are 26 letters in the standard English alphabet and that each letter is represented by a sound. The use of alphabet books helps children grasp making letter-sound correspondences, particularly beginning letter-sound association of words.

### **1.4. Orthographic Processing**

Orthographic processing consists of noting letter sequences in words or spelling patterns of words. Therefore, it involves examining the graphic structural properties of words.

### **1.5. Phonological Recoding**

Phonological recoding involves combining phonological with orthographic knowledge about words. It is a process by which children examine graphic representations of a printed word and recode them back into its spoken language form. When children encounter a printed word, they go through a process of sounding out the word and then check to determine if the word matches a word stored in their memory and in their speaking vocabularies. Children initially make sequential one-to-one correspondences with letters and sounds. At more advanced stages in this process, they hierarchically decode words by using some letters in a word to cue the sounds of other letters (e.g., the e at the end of “came” cues the reader to say the /a/ as a long vowel sound).

### **1.6. Morphological Understandings**

Conceptual knowledge about spoken words is critical for beginning readers. Morphemes represent the meaning units of words. Vocabulary development plays a

role in gaining morphological structures (meaning units) of spoken and written language, which is particularly critical when children, for instance, learn to read words that contain prefixes and suffixes as well as compound words.

### 1.7. Reading Fluency

The goal of basic reading instruction is to read words automatically or fluently. This is accomplished by acquiring a repertoire of word identification strategies as well as engaging in repeated readings of words in and out of context under timed conditions. Children are considered to be reading within grade level if they are reading between 75 and 100 words per minute correct on a grade-level passage with three to five errors.

## 2. CHARACTERISTICS OF INDIVIDUALS WITH READING PROBLEMS

Individuals with reading problems are generally characterized as having difficulties in obtaining word recognition skills and gaining meaning from text. These individuals are typically referred to psychologists for evaluation due to these general characteristics. Upon evaluation of several individuals with reading problems, the psychologist may discover that some children display delays across various areas, such as exhibiting low average cognitive ability (i.e., not in mental retardation range) and low average language, reading, math, and written language skills. These individuals are considered to be delayed types of readers, meaning that they tend to be delayed across cognitive and several academic learning domains.

Some individuals with reading problems have a central nervous system condition. These individuals are often classified as dyslexic and have average or higher intellectual abilities but severe reading problems coupled with speech production, rapid naming, phonological processing, and orthographic processing deficits. However, these individuals may not be delayed in math and other academic subject areas. Whether children are classified as dyslexic or the low average type, many children classified as having severe reading delays have deficits in phonological processing. Generally, students identified with learning disabilities in reading due to a central nervous system condition and a discrepancy between intellectual ability and achievement

are not functionally different in their development of reading skills than those who are low average across several cognitive and academic domains. The diagnostic label is less important than the identification of the specific skills for which children experience difficulty. In fact, many researchers have found that low average readers and individuals diagnosed with central nervous system conditions or dyslexia exhibit similar underlying literacy deficits, particularly deficits in phonological awareness.

Other individuals' reading difficulties are due to poor instruction, which often leads to acquisition and fluency problems. Some students recognize and comprehend words accurately but not at a sufficient rate. Other students do not recognize and comprehend words accurately. The most efficient readers are those who recognize and comprehend words accurately and at a sufficient rate (i.e., reading 100 words per minute in the students' grade-level text).

## 3. PHONIC INTERVENTIONS

Phonics is the system by which children learn to make letter–sound correspondences. According to many prominent reading scholars, good phonics instruction does not dominate literacy instruction; it is brisk, lasting 20–30 minutes per day; and it includes phonemic awareness and spelling instruction. There are a variety of approaches to phonics instruction. Synthetic phonics instruction involves breaking words into their constituent sounds and blending the sounds together to form a whole word. An example of synthetic phonics instruction is Englemann and Bruner's Reading Mastery (formerly known as DISTAR). Analytic phonics involves the application of rules to identify words. Analogic phonics approaches involve word study methods of identifying unknown words by making analogies to known words. The research is inconclusive regarding the best approach to teaching phonics; however, some researchers suggest that synthetic approaches are more effective than analytic approaches. A synthesis of the research concluded that systematic direct instruction of phonics is more effective for achieving word recognition and spelling skills than indirect methods.

A number of contemporary phonic techniques involve studying phonological and orthographic features of words. Many of these approaches are not discussed here. However, some are provided as examples of techniques that can be employed to help children examine words. Word sorts are an analogic phonic approach for

helping children categorize words according to shared phonological, spelling, and meaning components. They can be in the form of closed sorts, in which the teacher establishes the categories, or open sorts, in which children induce the categories based on an examination of subsets of given words. Words are printed on index cards, and three or more words are designated as the category words. A stack of words is shuffled and then children are asked to sort the words into one of the categories according to common sound or spelling patterns. Bear *et al.* provide a comprehensive guide to creating various types of word sort lessons and other word study phonic activities. For instance, children can complete phonemic awareness sorts by sorting pictured objects according to beginning and ending sounds. This type of sorting has been found to be especially beneficial and appropriate for preschool and kindergarten children. Children can also spell words in respective categories as the teacher orally presents words. Howard Street and Early Steps, two effective tutoring programs for low achievers, incorporate word sort techniques rather extensively. Word sorts are also being used for children identified with mental retardation, and these exercises have been found to be helpful in improving these children's spelling performance.

Word boxes, a synthetic phonic approach, help children segment sounds of spoken language. This approach has been effective for helping children with learning disabilities identify and spell words and has helped first graders and preschool children achieve phoneme awareness skills. Typically, word boxes have been used in Clay's comprehensive reading recovery program. A word box consists of a drawn rectangle that has been divided into sections (boxes) according to individual phonemes in a word. Initially, children place tokens in respective sections as each sound in a word is articulated slowly. Eventually, children place letters (either magnetic or tile) in respective sections as each sound in a word is articulated. During advanced phases, children are asked to write letters in the respective divided sections of the box.

#### 4. WHOLE WORD INTERVENTIONS

Whole word recognition or "look-say" approaches are effective especially for teaching words with irregular sound and spelling patterns. In look-say approaches, children practice reading words as a whole until they achieve fluency and the words become a part of their sight vocabularies. For instance, words that are printed

on index cards can be read as a whole within a specified time period. Research has shown that whole word methods may produce immediate word recognition achievement effects, whereas phonic approaches may produce more long-term generalized effects, especially when students encounter words unknown to them.

#### 5. REPEATED READINGS

Repeated readings are a very effective whole word recognition strategy for building fluency. Essentially, children read the same passage several times for 1 minute until mastery of the passage has been achieved. Mastery usually means reading the passage fluently—that is, reading all the words accurately and quickly. Reading fluency is highly correlated with reading comprehension.

#### 6. VOCABULARY DEVELOPMENT AND READING COMPREHENSION STRATEGIES

According to McCormick, instruction aimed at developing vocabulary should consist of (i) teaching a small number of words intensively in any given lesson; (ii) exposing the student to the same word in various contexts and in a systematic and continuous fashion; (iii) including gamelike activities to stimulate interest in learning concepts, teaching morphological units (e.g., prefixes and suffixes); and (iv) explicitly, teaching derivations of words through exercises using the dictionary. Categorization activities help children group words that share similar meanings and help students make distinctions about words that are not similar. Cloze procedure exercises involving a passage with every "nth" word missing can be used to help children determine where words are placed contextually within a passage.

Learning strategies instruction approaches were found to be effective for helping children with learning problems attain concepts and improve their reading comprehension performance. Effective learning strategies instruction approaches consist of self-questioning, constructing mental representations to integrate information from text, and identifying text consistencies. Examples of these approaches include semantic mapping, PQ4R (preview, question, read, reflect, recite, and review), and reciprocal teaching.

Semantic mapping typically involves developing word webs that reflect students' understanding of

concepts and constructing a diagram connecting events of a story or connecting facts taken from content area textbooks. Students who struggle to grasp conceptual relationships may find diagrams to be helpful visual aids. Semantic maps can be either process oriented or product oriented. Process-oriented maps are usually completed before students read assigned material to help them establish some background knowledge. This type of mapping requires teacher facilitation of student responses. For example, a teacher may write a concept (e.g., satire) in a drawn box and ask the class members to give examples of when they experience it. As the class members respond, the teacher writes their comments below the concept. The teacher may then ask the class members how they feels when they witness the concept and to write their responses, drawing connecting lines to the other comments, and so on. Product-oriented maps, on the other hand, are produced by students as an outcome activity after they have read material. Students generally work independently or with other peers and construct a map connecting ideas presented in text.

The PQ4R method of comprehending reading material is an extension of SQ3R (survey, question, read, recite, and review). This method involves previewing the reading material, questioning the reading, reading to answer the questions, reflecting on the reading, reciting the reading, and reviewing the material. Previewing the material means surveying the chapter titles, main topics, and subheadings of the text. Students can turn the headings and subheadings into questions. Questions that were developed can be answered by reading the text. Reflections about the content can occur as the material is being read if the students pause to form connections and create images. Reciting is retelling what was read from memory. This form of retelling helps students monitor the information they are obtaining from the reading. If some information is not being retained, students may need to read sections of the text again. Lastly, students review the material by answering questions and referring back to the text for clarification of mistaken responses to questions. This method appears to be most appropriate for older students because it encompasses the application of higher level metacognitive processes. Although PQ4R is not a new method, it continues to be considered an effective technique for helping students organize and retain information from texts.

Reciprocal teaching is a reading comprehension approach that has helped delayed readers catch up to and even exceed the level of typically developing readers.

This approach places heavy emphasis on teacher–student interactions in a rather cognitive apprenticeship fashion. After students and teacher read from common text, they dialogue with each other about the reading material. Initially, the teacher leads the dialogue by modeling strategies of predicting, question generating, summarizing, and clarifying text. The students are then asked to lead the discussions and apply the strategies that were demonstrated by the teacher. Guided practice is provided until students can use the strategies effectively. The goal of these reciprocal teaching interactions is to construct meaning from texts.

## **7. PRINCIPLES OF TEACHING AND LEARNING READING**

The principles of teaching and learning word recognition and comprehension strategies include connecting to and building schema, demonstration and modeling, providing opportunities to learn and practice skills, and shaping reading behaviors through corrective feedback and reinforcement.

### **7.1. Connecting to and Building on Schema**

Word recognition and comprehension skills are acquired more easily if the context from which these skills are presented is personally meaningful. This is especially important for children, who have limited conceptual knowledge. It is important for educators to become aware of children’s prior knowledge and experiences so they can help children associate new information presented in text to their already existing schemas.

### **7.2. Shaping Reading Behaviors**

Shaping behavior refers to modeling/demonstrating, providing corrective feedback, and eliciting reinforcers for successive approximations toward completing a task. Reading behavior is not acquired as naturally as language. In other words, we are not biologically predetermined to acquire reading skills but are biologically wired to acquire language. Therefore, reading behaviors need to be explicitly demonstrated and then modeled. It is important for parents or caregivers to provide opportunities for joint book reading early in children’s development. Joint book reading activities should consist of the parent demonstrating reading

print on a page and looking at the pictures. When parents are not engaged in shared book reading with their children, they should model reading behaviors by reading, for example, a newspaper, magazine, letter, or novel. When children enter school, educators need to explicitly demonstrate prereading and reading skills, such as phonemic awareness, letter naming, and letter-sound correspondences, and whole word reading skills. They also need to model these skills while they are engaged in their own reading activities.

Children need corrective feedback on inaccurate responses made during reading activities. When children mispronounce a word, they should be provided with corrective feedback or prompted to supply the accurate pronunciation of the word. When children identify words incorrectly, they may be corrected through whole word correction procedures or phonic analysis procedures. Whole word correction procedures consist of saying a word correctly when children read a word incorrectly. Phonic analysis correction procedures may involve helping the child break the word apart by sounding out each letter or letter combinations and then blending the sounds to form a whole word.

Delivering reinforcers for efforts made toward achieving a goal can be considered as a way of supporting children through the reading process. Children with reading problems will struggle and become frustrated as they proceed through a trial-and-error process of learning how to read. Reinforcers need to be delivered for efforts made toward reading words accurately and comprehending text and not just for accurate performance. For many older children with reading problems, reinforcers may not have been a systematic part of their instructional histories. In other words, consequences for reading behaviors may have not been emitted or delivered haphazardly rather than in successive approximations to desired reading performance.

### **7.3. Providing Opportunities to Learn**

It is important for children to be provided with sufficient time to engage in reading activities. However, allotted time for reading is not sufficient to ensure active engagement in reading tasks. Instead, children need to be provided with many opportunities to be engaged in tasks during the allotted time for reading exercises. During oral and silent reading activities, children need plenty of opportunities to make responses. This can be achieved if

lessons are structured so as to increase the number of responses children can make. For instance, choral reading of passages rather than round-robin reading increases all children's opportunities for being engaged in reading orally. Children also need many opportunities to practice skills that are learned. This is especially the case for children with learning disabilities and children with mental retardation. Overlearning leads to the ability to generalize skills to other contexts more easily.

### **7.4. Scaffolding**

Scaffolding means providing necessary support and gradually reducing that support as an individual begins to function more independently. Educators need to employ scaffolds or necessary supports for children until they are able to complete tasks independently. Scaffolds include teacher facilitation and feedback as well as appropriate instructional materials that serve to make problems solvable. Examples of scaffolds used during reading activities are verbal prompts, context clues, and manipulative materials. Educators who use scaffolds often view themselves and the materials they design or select as mediators of the learner's development.

### **7.5. Motivating**

Attribution retraining (training students to attribute their performance to the amount of effort rather than other factors) coupled with learning strategy instruction has been found to be effective for helping low achieving students to increase the amount of effort put forth on their reading through use of reading strategies. Children in kindergarten and first grade are more likely to expend effort toward reading even if they do not experience success, whereas older children, especially those with learning disabilities, are less likely to put forth effort if they do not experience immediate success with reading. This is because young children are likely to attribute their failures and successes to the amount of effort they expend, whereas older students are likely to attribute their successes and failures to other factors, such as teacher perceptions or task difficulty level. Likewise, many children lack an awareness of various strategies that would aid their reading skills, which makes it easy for these children to attribute their performance to other factors that may have no or a limited relationship to their performance.

## 8. TARGETING AND CONTINUOUS IMPROVEMENT OF READING SKILLS THROUGH PROCESSES OF LINKING ASSESSMENT TO INTERVENTION

The best approach to targeting specific reading skill problems is to conduct direct systematic assessment of reading performance. This can be accomplished through direct observations of students' oral reading behaviors in the classroom and through the use of curriculum-based reading measures. Curriculum-based reading measures consist of grade-level timed reading passages to assess fluency (accuracy and time) and also may consist of phonemic awareness, letter naming, and reading words out of context. Informal reading inventories and miscue analysis inventories are also helpful assessment measures for targeting specific reading performance. Evaluators should also observe the strategies students do and do not use while attempting to identify words and comprehend reading material. After specific skills and performance levels are assessed and noted, appropriate interventions can be targeted. Interventions should be implemented in a systematic manner and skills should be assessed during intervention periods to monitor student progress on a continuous basis as a function of the intervention. This process helps determine the effectiveness of the intervention. If a student does not demonstrate progress over time, then another intervention needs to be implemented and evaluated. This process of assessing and intervening is cyclical as students progress from performance level to another and from one skill to another.

### See Also the Following Articles

Literacy, Improvement of ■ Reading, Teaching of ■ Teaching Effectiveness ■ Writing, Teaching of

### Further Reading

- Bear, D. R., Invernizzi, M., Templeton, S., & Johnston, F. (1996). *Words their way: Word study for phonics, vocabulary, and spelling instruction*. Englewood Cliffs, NJ: Prentice Hall.
- Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology*, 83, 451–455.
- Clay, M. M. (1993). *Reading recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.
- Connelly, V., Johnston, R., & Thompson, G. B. (2001). The effect of phonics instruction on the reading comprehension of beginning readers. *Reading and Writing: An Interdisciplinary Journal*, 14, 423–451.
- Daly, E. J., III, Witt, J. C., Martens, B. K., & Dool, E. (1997). A model for conducting functional analysis of academic performance problems. *School Psychology Review*, 26, 554–574.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71(3), 393–447.
- Engelmann, S., & Bruner, E. (1998). *Reading mastery I: Distar reading*. Chicago: Science Research Associates.
- Gambrell, L. B., Morrow, L. M., Neuman, S. B., & Pressley, M. (1999). *Best practices in literacy instruction*. New York: Guilford.
- Joseph, L. M. (1998/1999). Word boxes help children with learning disabilities identify and spell words. *The Reading Teacher*, 52(4), 348–356.
- Joseph, L. M. (2000). Developing first grade children's phonemic awareness, word identification, and spelling skills: A comparison of two contemporary phonic approaches. *Reading Research and Instruction*, 39(2), 160–169.
- Joseph, L. M. (2002). Facilitating word recognition and spelling using word boxes and word sort phonic procedures. *School Psychology Review*, 31, 122–129.
- Joseph, L. M., & McCachran, M. (2003). Comparison of a word study phonics technique between students with moderate to mild mental retardation and struggling readers without disabilities. *Education and Training in Mental Retardation and Developmental Disabilities*, 38.
- Maslanka, P., & Joseph, L. M. (2002). A comparison of two phonological awareness techniques between samples of preschool children. *Reading Psychology: An International Quarterly*, 23, 271–288.
- McCormick, S. (1999). *Instructing students who have literacy problems* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Nation, K., & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. *Reading Research Quarterly*, 32, 154–167.
- National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Rockville, MD: NICHD Clearinghouse.
- Nelson, K. (1996). *Language in cognitive development: The emergence of the mediated mind*. New York: Cambridge University Press.
- Stahl, S. A., Duffy-Hester, A., & Stahl, K. A. (1998). Theory and research into practice: Everything you wanted to

- know about phonics (but were afraid to ask). *Reading Research Quarterly*, 33, 338–355.
- Sulzby, E., & Teale, W. (1991). Emergent literacy. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 727–758). New York: Longman.
- Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult to remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities*, 33, 223–238.



# Reading, Teaching of

Virginia W. Berninger

University of Washington, Seattle, Washington, USA

1. Historical Background
2. Relationship between Teaching and Learning
3. Developmental Framework for Creating a Functional Reading System
4. Learning to Read
5. Reading to Learn
6. Coordinating Reading and High-Level Skills
7. Individualizing Instruction for Students with Reading Disabilities
8. Optimizing Reading Instruction for All Students Further Reading

## GLOSSARY

**alphabetic principle** Set of correspondences between a phoneme and the alternative one- or two-letter spelling units that represent it in the writing system.

**automaticity** Nonstrategic, effortless, direct retrieval and production of a word.

**decoding** Converting letters to phonemes and synthesizing the phonemes into a pronounceable word or covert phonological representation of a word.

**fluency** Fast, smooth, coordinated oral reading that reflects the intonation (melody) of the spoken language or rapid, efficient silent reading.

**morphemes** Word parts that, alone or attached to other word parts, convey meaning.

**morphological awareness** Conscious reflection about and manipulation of morphemes, including prefixes, roots or bases (roots plus affixes), and suffixes (inflectional ones that mark number, tense, and comparison as well as

derivational ones that mark grammatical speech part) in short-term and working memory.

**nature–nurture interactions** The notion that reading (and other) behavior is the result of biological and environmental influences.

**orthographic awareness** Consciously reflecting about and manipulating letter units in written words in short-term memory and working memory.

**phoneme** Smallest unit of sound that (a) has no meaning of its own but makes a difference in meaning and (b) corresponds to one or more spelling units.

**phonological awareness** Consciously reflecting about and manipulating sound patterns in spoken words (syllables, phonemes, and remaining parts of syllables after initial phonemes are removed) in short-term and working memory.

**reading comprehension** Text interpretation derived from vocabulary meaning, syntactic awareness, and discourse structures in text and from language and cognitive processes that go beyond what is stated in text.

**rule-governed knowledge versus word-specific knowledge** Abstracted patterns and relationships that can be applied across cases (single exemplars or multi-item constructions) versus learned representations in multiple formats (orthographic, phonological, and morphological codes) used in case (exemplar or instance) learning.

**spelling unit or grapheme** One or two letters that correspond to a phoneme.

**strategies** Nonautomatic reflective approaches applied to reading goals.

**syntactic awareness** Consciously reflecting about and manipulating the structures in sentences or conversational utterances in short-term and working memory.

Teaching reading is a teacher-led process of explicitly guiding students who actively participate in developing and coordinating (a) orthographic, phonological, and morphological awareness of words and their parts that results in accurate and automatic reading of words from the Anglo-Saxon, Latinate, and Greek layers of the language; (b) vocabulary meaning, syntactic awareness, and discourse structures that contribute to accurate and fluent reading and reading comprehension; and (c) interrelationships of reading with listening, speaking, and writing. Effective reading instruction also shows students how to apply the emergent knowledge and procedural skills to a wide variety of goals, including (a) learning from and thinking about written texts, (b) enjoying reading, and (c) using reading for pragmatic functions at school, home, and work as well as in everyday living.

## **1. HISTORICAL BACKGROUND**

### **1.1. Writing Systems and Teaching Methods**

Reading became possible when humans discovered three ways in which to represent the meaning or sounds of words in writing: ideograms/pictographs portraying concepts with line drawings, visual syllabaries standing for spoken syllables, and alphabets representing phonemes in spoken words with one- or two-letter spelling units. Three instructional methods corresponding to each of these writing systems were developed: whole word method (also referred to as look-say or sight words), syllable method and related word family method (syllable without its initial sound), and phonics (correspondences between phonemes and alphabet letters), respectively.

### **1.2. Debates**

During the 20th century, scientific research resolved the debate about phonics versus the whole word. The conclusion was that phonics is the most effective, at least for the primary grades when basic word decoding skills are typically mastered. Research also showed that decodable text is effective in beginning reading instruction but that other methods, such as repeatedly reading the same whole words and teaching word families, are effective supplements to phonics. During the 20th century, scientific research also resolved another debate about whole language versus phonics that was about the optimal unit of instruction and instructional goal.

Advocates of whole language argued for the whole text (based on good children's literature) as the unit and meaning as the goal, whereas proponents of the phonological method argued for the phoneme in spoken words as the unit and phonological decoding as the goal so that students can read independently. Scientific research supported the conclusions that (a) the phonological method is more effective for beginning reading instruction, (b) meaning is motivating to beginning readers who benefit from listening to and discussing good children's literature rather than reading it independently, (c) children exhibit individual differences in the amount and intensity of phonological instruction they require, (d) phonological instruction may be necessary but it is not sufficient, and (e) a balanced approach focused on both decoding and meaning is the most effective. Because language has multiple dimensions, effective instruction directs children's attention not only to phonology and orthography but also to the morphology, syntax, and discourse structure of spoken and written language.

The debate about direct instruction (which is teacher controlled with rote student responding) versus constructivism (which has minimal direct teacher input and relies on student construction of personal written language knowledge) is also being resolved with scientific research. Both teacher-led explicit instruction and strategy instruction for self-regulated learning benefit struggling readers, who learn from (a) direct explanations about the structure and rules underlying language and (b) teacher-led questions, prompts, or activities that encourage students to engage in active hypothesis testing in applying those explanations to their own reading. At the beginning of the 21st century, there is an increasing recognition that instructional methods for teaching reading should be grounded in scientific research rather than in philosophical belief.

## **2. RELATIONSHIP BETWEEN TEACHING AND LEARNING**

Teaching has three instructional components: pedagogy (instructional behaviors of the teacher), curriculum (instructional goals and plans to reach those goals), and instructional materials (text and other instructional resources for achieving those goals). Teaching is related to learning but not in a simple one-to-one fashion. Teachers cannot program students' minds/brains through instruction the same way in which a

programmer can program a computer through input. Ultimately, students create their own learning (mental representations and organization) from the instructional cues they receive from pedagogy, curriculum, and instructional materials. Teacher knowledge of the developing brain, cognitive psychology, linguistics, and child development can influence the quality and appropriateness of the instructional cues provided. These instructional cues, in turn, are influenced by multiple biological (e.g., genes, neural architecture, maturation) and environmental (e.g., home literacy, school instruction, cultural, linguistic, socioeconomic) factors. Hence, learning is the by-product of nature–nurture interactions, and internal cognitive, affective, and motivational processes in the learner’s mind/brain mediate the instructional cues and constrain but do not determine learning independent of environmental input.

### 3. DEVELOPMENTAL FRAMEWORK FOR CREATING A FUNCTIONAL READING SYSTEM

The functional reading system, which supports independent reading for a variety of purposes, is constructed over a long course of development from multiple language and nonlanguage systems in the brain and their interactions with the instructional environment. Although research has established correlations between beginning reading and subsequent reading skill, this relationship is not causal; developmentally and instructionally appropriate intervention at any time in reading development may improve the functioning of the reading system. In designing and implementing developmentally appropriate intervention, it is important not to confuse the novice and expert stages. Effective reading instruction aims just above the current instructional level to promote further growth along the developmental path to literacy. Children may follow different developmental paths to literacy. Some begin the journey at home during the preschool years, with parents using the lap method (reading storybooks to their children on a regular basis). Others do not begin the journey until school entry, but instructional methods have been developed to simulate the lap method in school settings (e.g., big books and little books with teacher-led finger point reading). However, federal programs are providing home literacy programs and incentives for explicit literacy instruction in preschools for those children who, heretofore, came to

**TABLE I**  
**Developmental Framework with Partially Overlapping, Cascading Stages**

---

<b>First stage: Learning to read words and understand text</b>
1. Subword level: alphabetic principle, syllables, and morphemes
2. Word level: rule-governed decoding, word-specific representations, automatic word recognition
3. Text level: fluent oral reading, fluent silent reading, reading comprehension
4. Integration of reading with listening, speaking, and writing
<b>Second stage: Reading to learn, enjoy, and work in school and nonschool settings</b>
1. Domain-specific content areas in school curriculum
2. Leisure time activities in and out of school
3. Everyday reading for functioning in society
4. Vocation-specific reading for functioning at work
<b>Third stage: Executive coordination of reading and high-level functions</b>
1. Integration with technology (e.g., word processing systems, Web-based information systems)
2. Integration with thinking (e.g., critical thinking), mental world building (e.g., creative thinking), and problem solving

---

school unprepared for reading. *Table I* provides an overview of partially overlapping, cascading developmental stages and their instructional components.

### 4. LEARNING TO READ

Basic decoding skills are typically taught during the first three grades. Fourth grade is an important transition time. Written text changes from mostly one- and two-syllable words of Anglo-Saxon origin that characterize primary grade reading material (and conversation) to a high density of words of more than two syllables and mostly of Latin (and Greek) origin that characterize reading material in the upper grades—when students continue to benefit from explicit instruction in decoding longer, more complex words. Also, the nature of the reading vocabulary, syntax in written text, and discourse structures changes. High-frequency vocabulary gives way to low-frequency complex vocabulary. Syntax in written material becomes more complex. Discourse structure changes from primarily narrative to primarily expository. Oral reading gives way to silent reading as the primary mode of instruction.

### 4.1. Teaching Word Reading Skills

Table II lists the instructional goals and associated instructional activities organized in the developmental progression from initial to mastery phases in learning word reading skills. In general, English word reading is predictable but not completely regular on the basis of alphabetic principle alone. Alphabetic principle can be applied to decoding both structure words (high-frequency words such as prepositions, conjunctions, pronouns, and articles that contribute to sentence meaning but have no meaning by themselves and are not always perfectly decodable) and content words (nouns, verbs, adjectives, and adverbs that tend to be more decodable). In English, a word might not be governed by regular rules of word identification at one level of analysis (e.g., single letter and associated phoneme) but might be when larger units (multiletter word families) or a different level of analysis (morphological) is taken into account. Although alphabetic principle benefits beginning word decoding, subsequent reading of polysyllabic

words requires attention to additional linguistic cues (e.g., morphological) in word structure as well.

### 4.2. Teaching Reading Comprehension

In 2002, Carlisle and Rice provided an overview of teaching reading comprehension (see also Table III). Because lack of word automaticity (Table II) or text fluency (Table III) can interfere with normal reading comprehension, instructional activities aimed at developing oral and silent reading fluency are important components of reading comprehension instruction. From fourth grade onward, reading vocabulary instruction is necessary because many of the words are not part of students' speaking vocabulary as was the case with beginning reading texts. Effective vocabulary instruction requires far more than teaching the meaning of each word in the language. Both morphological awareness and syntactic awareness contribute to oral

TABLE II  
Teaching Word Reading

<i>Reading instructional goal</i>	<i>Instructional activities</i>
Phonological awareness	Rhyming, syllable and phoneme analysis with spoken words
Orthographic awareness	Letter naming and letter writing; looking games requiring careful attention to (and memory for) all letters in a written word
Alphabetic principle	Producing the phoneme corresponding to a spelling unit containing one or two letters
Decoding monosyllabic Anglo-Saxon words (structure and content words)	Finding letters or spelling units that correspond to phonemes (applying alphabetic principle) and synthesizing the phonemes to pronounce the word
Decoding two-syllable Anglo-Saxon words (structure and content words)	Identifying type (e.g., closed, open, silent e, vowel team, r-controlled, -le) of each syllable in word and then decoding each syllable using alphabetic principle
Morphological awareness	Differentiating initial common spelling patterns that do or do not correspond to prefixes and final common spelling patterns that do or do not correspond to suffixes marking tense, number, or comparison or derivational suffixes marking grammatical part of speech; analyzing words into roots, affixes, or bases (affixed roots)
Decoding polysyllabic words from the Latinate and Greek layers of the language of Latinate and Greek origin	Introducing correspondences between spelling units and phonemes, syllable patterns (e.g., schwa), and morpheme patterns specific to the Latinate or Greek layers; practice applying multiple strategies based on alphabetic principle, syllable types, and morphemes to words
Automaticity	Practice reading monosyllabic and polysyllabic words from Anglo-Saxon, Latinate, and Greek word origin out of sentence context (on lists or flash cards) with and without rate criteria

TABLE III  
Teaching Reading Comprehension

<i>Reading instructional goal</i>	<i>Instructional activities</i>
Oral reading fluency	Encourage children to attend to units larger than single words and to reflect the musical melody of the spoken language when reading aloud. Compute and record rate of accurate reading of repeated readings of same passages on which children can identify at least 90% of running words.
Silent reading fluency	Transition to silent prereading before oral reading, and follow same procedures as for oral reading fluency. Eventually transition to silent reading and wide reading of many passages. Record time for silent wide reading as a function of comprehension criteria (answering factual and inferential questions with 90% accuracy and story recall with main idea and supporting details).
Vocabulary	Guide students in inferring word meaning from sentence context. Define new vocabulary and use each word in a sentence to illustrate its meaning in context. Provide morphological patterns for generating new vocabulary items.
Sentence interpretation	Draw students' attention to the importance of word order by rearranging scrambled words into grammatically acceptable sentences. Discuss how meaning changes by reordering words slightly in a sentence. Help students paraphrase sentences to communicate the same idea.
Discourse understanding	Relate students' background (world) knowledge to content of the text to be read. Ask questions that prompt students to think about both information stated in text and information inferred from text. Teach strategies for self-regulating and monitoring the comprehension process, including prediction, summarization, clarification, and elaboration. Show how discourse structure can be used to build a mental representation of text. Describe the discourse structures of different genres: narrative (setting, characters, plot, and outcome) and expository (informational: lists, descriptions, comparisons, cause and effect, problem-solution; persuasive: one argument with evidence and conclusion with reasons, counterarguments with evidence, point-counterpoint, conclusions with reasons). Call attention to the different kinds of paragraph organization within the different genres (e.g., events for narrative, main idea and supporting detail for expository, mix of more than one genre).

vocabulary development as well as reading comprehension. Prior focus on question asking to assess comprehension has given way to pedagogy for explicit instruction in cognitive strategies for self-regulating and monitoring the reading comprehension process and for increasing active engagement of the reader in processing written text.

### 4.3. Integration of Reading with Other Language Skills

#### 4.3.1. Listening

Students learn about reading by listening to the teacher's instructional language. Compared with conversational speech that is highly contextualized (embedded in social routines that support turn taking and communication repair), instructional language is decontextualized, with the focus redirected from language as a means of achieving an end (social communication) to language as an end in itself (language as the object of attention and

reflection). Listening to the teacher read good children's literature stimulates linguistic development (vocabulary meaning, syntactic awareness, and discourse understanding) and cognitive development, and this in turn benefits reading development.

#### 4.3.2. Speaking

Oral discussions about what has been read promote elaboration and deeper understanding of a text. Pedagogical methods for encouraging talk about reading material range from teacher-guided, collaborative, whole group discussions; to reciprocal teaching in which students assume teaching roles with some teacher guidance; to book clubs in which discussion is totally under student control.

#### 4.3.3. Writing

Writing alphabet letters increases orthographic awareness, which in turn may facilitate learning to decode

words. Word spelling may exert reciprocal, mutually enhancing influences on the development of word reading. Written composition develops as children write about the material they read for various assignments.

## 5. READING TO LEARN

### 5.1. Teaching Domain-Specific Reading Skills in Content Areas

Although many middle school and high school teachers view their role as merely transmitting content matter, students learn more in content subjects when teachers also teach the vocabulary terms, textbook genre, and study strategies unique to the discipline.

### 5.2. Teaching Positive Affect Toward Reading

Unless children are successful in reading and enjoy reading, they are unlikely to engage in sufficient sustained practice of reading to become skilled readers. Therefore, reading instruction must create opportunities for children to learn to derive pleasure from reading. Some effective techniques for developing a reading habit and positive affect toward reading are including reading choice time at school and including leisure reading time at home. Children enjoy keeping reading logs of what they have read (e.g., books, magazines, newspapers) and competitions to see who can read the most books.

### 5.3. Teaching Everyday Reading Skills

Reading skills might not transfer automatically from the classroom to the real world without explicit instruction in specific reading tasks such as reading labels on food and household products, medicine labels and pharmacy instructions, menus, newspapers, magazines, advertisements, and the like.

### 5.4. Teaching Vocation-Specific Reading Skills

Vocation-specific and job setting-specific reading requirements vary considerably. High schools and colleges cannot prepare students fully for every kind of job-related reading goal (e.g., reading and applying

information in highly technical manuals for operating heavy equipment or high-tech instruments). Job-related reading goals require not only decoding skills but also considerable domain-specific content knowledge. Some on-the-job reading comprehension strategies might be necessary.

## 6. COORDINATING READING AND HIGH-LEVEL SKILLS

### 6.1. Teaching Literacy for Technology Users

Although most elementary schools have computer laboratories, few have fully integrated technology into their literacy programs. By middle school, many students use word processing programs to complete written assignments based on reading source material but without explicit instruction in how to do so. As word processing and voice activation programs become more widely available and integrated with the instructional program, the need for explicit instruction in keyboarding, dictation, and integrated reading-writing systems for word processing will increase. During the information age, students also need explicit instruction in using Web-based information systems.

### 6.2. Teaching Critical and Creative Thinking and Problem Solving

Reading instruction should prompt readers to go beyond understanding a text at an initial level of analysis. For example, instruction may challenge readers to take a critical stance toward the text (e.g., its arguments, the author's intent), create new ideas that extend those in the text, and/or solve problems posed in writing assignments using reading source material. Many believe that teaching reading as a thinking skill is critical to the survival of democracy as a form of government.

## 7. INDIVIDUALIZING INSTRUCTION FOR STUDENTS WITH READING DISABILITIES

Effective strategies for teaching students with reading disabilities have been developed. However, many challenges remain in implementing individually tailored, appropriate instruction for all students with learning disabilities.

## 8. OPTIMIZING READING INSTRUCTION FOR ALL STUDENTS

At the beginning of the 21st century, expectations are rising for the educational achievement of all students. Considering the enormous normal variation among students at the same grade level, the expectation that all students will be readers and meet high-stakes standards will require reading instruction to become increasingly sophisticated.

### See Also the Following Articles

Learning ■ Literacy, Improvement of ■ Reading Interventions  
■ Writing, Teaching of

### Further Reading

- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Balmuth, M. (1982). *The roots of phonics: A historical introduction*. Baltimore, MD: York Press.
- Berninger, V., & Richards, T. (2002). *Brain literacy for educators and psychologists*. San Diego: Academic Press.
- Birsh, J. (Ed.). (1999). *Multisensory teaching of basic language skills*. Baltimore, MD: Paul H. Brookes.
- Calfee, R., & Patrick, C. (1995). *Teach our children well: Bringing K–12 education into the 21st century*. Stanford, CA: Stanford Alumni Association.
- Carlisle, J., & Rice, M. (2002). *Improving reading comprehension: Research-based principles and practices*. Baltimore, MD: York Press.
- Chall, J. (1996). *Learning to read: The great debate* (3rd ed.). Fort Worth, TX: Harcourt Brace. (Originally published in 1967).
- Clark, D., & Uhry, J. (1995). *Dyslexia: Theory and practice of remedial instruction* (2nd ed.). Baltimore, MD: York Press.
- Ehri, L., Nunes, S., Stahl, S., & Willows, D. (2002). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71, 393–447.
- Nagy, W. (1988). *Teaching vocabulary to improve reading comprehension*. Newark, DE: International Reading Association.
- Nagy, W., Osborn, J., Winsor, P., & O'Flahavan, J. (1994). Structural analysis: Some guidelines for instruction. In F. Lehr, & J. Osborn (Eds.), *Reading, language, and literacy* (pp. 45–58). Hillsdale, NJ: Lawrence Erlbaum.
- Palincsar, A., & Brown, A. (1986). Interactive teaching to promote independent learning from text. *The Reading Teacher*, 39, 771–777.
- Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, 2, 31–74.
- Swanson, H. L. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York: Guilford.
- Thompson, G. B., & Nicholson, T. (Eds.). (1999). *Learning to read: Beyond phonics and whole language*. Newark, DE: International Reading Association.
- Venezky, R. (1999). *The American way of spelling*. New York: Guilford.
- Wong, B. (Ed.). (1998). *Learning about learning disabilities* (2nd ed.). San Diego: Academic Press.



# Reciprocity Norm

Edwin P. Hollander

Baruch College and University Graduate Center, City University of New York, New York, New York, USA

1. Background
  2. Mutual Interest
  3. Interdependence
  4. Social Exchange
  5. Equity and Inequity
  6. Bargaining and Negotiation
  7. Implications
- Further Reading

## GLOSSARY

**adaptation level** The standard of adjustment to a pattern of experience that a person has become used to.

**comparison level** A standard against which a person judges the attractiveness of a relationship with regard to whether he or she receives what is deserved, relative to his or her costs.

**comparison level for alternatives** The reference point employed in deciding whether to remain in an interaction. It can be considered as the lowest level of reward that an individual will accept in order to continue in the relationship. It also relates to the concept of adaptation level, with respect to what one has come to expect.

**distributive justice** A concern for whether an individual's investments in an ongoing relationship are or are not balanced by rewards.

**egoism** A self-centered basis for judging experience as good or bad in terms of one's benefits.

**injustice** A perceived negative outcome between actual rewards and expected rewards.

**reference group** A group that an individual takes as a standard for attitudes and actions, whether or not he or she belongs to it.

Social interaction between individuals involves an exchange of benefits, the most distinctive feature of which is the "give and take" of reciprocal behavior. This norm of reciprocity has been summarized by Gouldner as a universal norm stating that a benefit should be returned and, above all, that one who gives a benefit should not be harmed. These are often captured in everyday speech by such statements as "do unto others as you would have them do unto you" and "don't bite the hand that feeds you." Gouldner states that reciprocity fulfills two basic functions: to stabilize relationships and to serve as a check on egoism. These show concern for interdependence in relationships and the need to consider more than one's self-interest.

## 1. BACKGROUND

The rewards exchanged in relationships need not be of the same kind or of equal amount. Also, the anticipation of receiving a reward can account for such things as sacrifices, where goods are given to "appease the gods," which occur in many societies.

Another example is the distinction between a leader's election or appointment. An election has been found to give followers a great sense of responsibility and higher expectations for the leader's performance. In effect, the group gives the leader a benefit in advance by electing him or her to a position of higher status, and then the group feels a claim on the leader to return the benefit by producing favorable outcomes to the group.

## 2. MUTUAL INTEREST

When there is mutual interest, in most social settings one can assume that the other person understands the relevance of the reciprocity norm. This makes for a smoothing of relationships, if commonly understood. However, this need not always be the case, even with the need for interdependence. Most often, however, the concept of community of interest requires that reciprocity prevail, although it is conditional. There are variations among individuals and groups in the way in which it is applied. Although the broad culture may support the norm, its application still depends on the particular people in a given interaction.

## 3. INTERDEPENDENCE

Group support is closely tied to the motivation to take part in the task of a group. Whenever individuals come together to do things, they rely on one another for suitable performance. To the extent that members are rewarded by their attempts to achieve group goals, they will help work toward the goal and will be supported by the others. From a psychological standpoint, it is crucial that the individual care about what the group seeks to attain. Therefore, in joining groups, individuals are often attracted by the group's activity. Then they play a role that is rewarded by others' approval, which brings task and affiliation motives together.

Not just any source of support will influence a person's actions, however. The favorable effects of group support require two conditions: The individual must be positively motivated toward the activity, and he or she must be favorable about identification with the group members as relevant others. They must be people whose standards and support are desired—that is, a “reference group.”

It is possible for an individual to do something he or she finds personally distasteful because others whose approval is sought are doing it. However, where a choice exists, the usual inclination is to be associated with those who support activities that the individual finds satisfying.

## 4. SOCIAL EXCHANGE

Although stated mainly in economic terms, social exchange goes beyond the simple notion of straight profit.

Two early contributions to an understanding of social exchange derive from the 1961 work of Homans and the 1959 work of Thibaut and Kelley. Both involve certain concepts associated with a rewards–costs relationship.

Homans views social exchange as the underlying factor accounting for certain features of face-to-face interaction. He states that the rules for social interaction apply in various settings and can be understood primarily in terms of reinforcement, where the quantity and the value of an activity or a sentiment can affect the behavior of another individual. His propositions for what he calls elementary social behavior can be simply stated as follows: The more a person's activity is rewarded, the more likely the activity is performed; the more similar a present activity is to one rewarded in the past, the more it is performed; the greater the value of an activity, the more it is performed; the more rewarded an activity, the less any further unit of reward is valued; and the failure to receive an expected reward for an activity results in anger, which makes the results of aggressive behavior rewarding. In addition, he explains that a lack of reward leads to apathy, whereas too much of that reward decreases its value. Profit is defined as total rewards minus costs, with costs including values given up or forgone.

Thibaut and Kelley proposed that rewards for the parties involved are needed to maintain the interaction. In ongoing relationships, there is a trading off of rewards and costs for both parties. Thibaut and Kelley define rewards as

The pleasures, satisfactions, and gratifications the person enjoys. The provision of a means whereby a drive is reduced or a need fulfilled constitutes a reward. We assume that the amount of reward provided by any such experience can be measured and that the reward values of different modalities of gratification are reducible to a single psychological scale. (p. 12)

They define costs as

Any factors that operate to inhibit or deter the performance of a sequence of behavior. The greater the deterrence to performing a given act—the greater the inhibition the individual has to overcome—the greater the cost of the act. Thus cost is high when great physical or mental effort is required, when embarrassment or anxiety accompany the action, or when there are conflicting forces or competing response tendencies of any sort. (pp. 12–13)

In line with Homans' idea of profit as rewards minus costs, Thibaut and Kelley use the concept of comparison level. This is a standard against which the person judges

the attractiveness of the relationship with regard to whether he or she receives what is deserved, relative to his or her costs. Related to this concept is the comparison level for alternatives. It represents the reference point employed in deciding whether to remain in the interaction. It can be considered as the lowest level of reward that the individual will accept in order to continue in the relationship. It also relates to the concept of adaptation level, with respect to what one has come to expect.

The relationship of comparison level and the comparison level for alternatives can be stated simply: In general, an individual will be highly satisfied with outcomes if his or her comparison level and the comparison level for alternatives are both exceeded by actual outcomes. The opposite holds true with regard to dissatisfaction.

An injustice can also be thought of as an imbalance between an individual's rewards and costs in an activity. This is the fundamental point in Homans' concept of distributive justice. It deals with the condition in which an individual's investments are either balanced or not by his or her rewards. As previously noted, he considers profit to be rewards minus costs. Among these costs may be the investments of the individual—that is, what the individual brings to the situation from his or her background. This includes status, education, and experience, in addition to capability.

## 5. EQUITY AND INEQUITY

Equity is another term for fairness. An inequity is more likely to be perceived as an unfavorable comparison between one's own and others' rewards relative to costs. In the latter case, individuals will compare their rewards with others and take account of their own and the others' investments. In the workplace, these include such factors as seniority, responsibility, and productivity. Compensation is not the only element in assessing equity in the workplace. In 1983, Greenberg and Ornstein found that other variables, such as promotions, interesting work, pleasant co-workers, and increased status, contribute to the sense of equity. In 1987, Bandura more broadly asserted that goal accomplishment is a major feature of equity as well.

An earlier conception presented by Adams in 1965 defines equity as a function of the balance effected between a person's inputs and outcomes on the job. Inputs include qualifications and how hard one works

$$\frac{\text{Outcomes}_P \text{ (gains)}}{\text{Inputs}_P \text{ (costs)}} \qquad \frac{\text{Outcomes}_O \text{ (gains)}}{\text{Inputs}_O \text{ (costs)}}$$

**FIGURE 1** Comparison of a person's (P) own ratio of outcomes to inputs with an other's (O) ratio. Adapted from Adams (1965).

at the job, whereas outcomes include pay, fringe benefits, status, and the intrinsic interest of the job.

In his equity theory, Adams contends that the major expectancy is that an equitable input–outcome ratio will prevail. Inequity exists for a person (P) whenever he or she perceives that the ratio of his or her own outcomes to inputs and the ratio of other's (O) outcomes to inputs are unequal. This comparison of ratios can occur when P and O are in a direct exchange relationship or when both are in an exchange relationship with a third party, such as working in the same group or organization (Fig. 1).

In any consideration of justice or injustice, and equity or inequity, there is likely to be a difference of view. It is to be expected that P's perception of his or her own investments, costs, and rewards may not be the same as O's perceptions of them. However, it is not the actual fact nor the other person's perception that matter most in determining action. It is the individual's own definition of the situation. Therefore, inequities may be perceived by individuals despite the fact that others see no inequity in their version of social reality.

## 6. BARGAINING AND NEGOTIATION

Fundamentally, the process of interaction has a bargaining component. Even as they compete to gain desired outcomes, there are advantages to be achieved as P and O cooperate to trade rewards. Cooperation within competition is facilitated by negotiation. It remains the one method most widely used, and it is seen routinely in labor management relations, in which there may be awkwardness associated with power differences. To be effective, rules are necessary for an agreed procedural structure that depersonalizes the conflict. These help to avoid pitfalls such as perceiving the conflict as a zero-sum game, in which there must be a winner and a loser. Instead of accommodation, there may be an adherence to fixed positions, which may make both parties losers.

There is also the prospect of misunderstandings that derail progress. A basic feature of successful negotiation is to understand the opponent's position, which is helped by having open communication with the mediation of an outsider. A mediator can be useful for encouraging such things as limits on hostile language. This risk occurs when leaders desire to solidify their support base, with costs to the resolution of the conflict. Osgood's 1962 GRIT (Graduated Reciprocation in Tension-Reduction) grew out of his concern with the threat posed by the arms buildup in the Cold War between the United States and the Soviet Union. The strategy involves making concessions, usually small at first, to set in motion a process of reciprocation by the other side, ultimately leading to tension reduction of wider scope. At the negative extreme, in 1962 Tuchman described the misunderstandings and downward spiral that led to the arms buildup, always said to be "defensive," prior to the onset of World War I.

## 7. IMPLICATIONS

Running through these points is the norm of reciprocating, whether positively or negatively. The strategy utilized is dependent on the circumstances and the goals sought, including power relationships. If there are not mutual benefits in a solution, then the likelihood of a successful outcome is remote. Therefore, the norm is important, but it is sometimes not as important as the attitudes and motives of the parties involved.

## See Also the Following Articles

Intergroup Relations and Culture

## Further Reading

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press.
- Bandura, A. (1987). *Social foundation of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Deutsch, M., & Coleman, P. (Eds.). (2000). *The handbook of conflict resolution: Theory and practice*. New York: Wiley.
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25, 161-179.
- Greenberg, J., & Ornstein, S. (1983). High status job title as compensation for underpayment: A test of equity theory. *Journal of Applied Psychology*, 68(2), 285-297.
- Hollander, E. P., & Julian, J. W. (1969). Contemporary trends in the analysis of leadership processes. *Psychological Bulletin*, 71, 387-397.
- Hollander, E. P., & Julian, J. W. (1970). Studies in leader legitimacy, influence, and innovation. In L. L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 5, pp. 33-69). New York: Academic Press.
- Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt, Brace & World.
- Osgood, C. E. (1962). *An alternative to war or surrender*. Urbana, IL: University of Illinois Press.
- Pruitt, D. (1981). *Negotiation behavior*. New York: Academic Press.
- Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.
- Tuchman, B. (1962). *The guns of August*. New York: Macmillan.



# Recruitment

James A. Breugh

University of Missouri, St. Louis, Missouri, USA

1. Problems Resulting from Ineffective Employee Recruitment
  2. Deciding on Recruitment Objectives
  3. Recruitment Strategy Development
  4. Considering the Job Applicant's Perspective
  5. Assessing Recruitment Results
  6. Conclusion
- Further Reading

*recruiting methods* Approaches such as job fairs and mailing lists used by employers to reach potential applicants for the purpose of informing them of job openings.

Research has shown that various recruitment decisions can have a significant impact on such factors as the diversity of job applicants and whether they accept job offers and remain on the job for a reasonable period of time. This article addresses such important topics as deciding on recruitment objectives, choosing a recruitment method, selecting recruiters, deciding on a recruitment message, and evaluating a recruitment campaign.

## GLOSSARY

*applicant population* The type of job candidates (e.g., retirees and former employees) an organization hopes to reach by its recruiting efforts.

*applicant self-insight* Refers to the extent to which a job candidate has an accurate understanding of his or her abilities, interests, etc.

*employee referrals* A recruitment method that involves current employees of an organization referring people they know as candidates for open positions. Frequently, an employer will offer incentives to motivate such referrals.

*realistic job preview (RJP)* An attempt by an employer to convey accurate information concerning a position (e.g., job duties and work hours) to job candidates. RJPs can involve a variety of different approaches (e.g., a video, a brochure, and a visit to the work site) and are designed to improve the accuracy of recruits' job expectations.

## 1. PROBLEMS RESULTING FROM INEFFECTIVE EMPLOYEE RECRUITMENT

The importance of effective employee recruitment has become increasingly recognized by organizations. This recognition at least partly stems from the fact that ineffective recruitment efforts (e.g., using unprofessional recruiters and presenting an overly positive view of an open position) have been linked to a number of serious problems that employers commonly face. For example, ineffective recruitment activities have been linked to (i) desirable job applicants turning down job offers; (ii) firms' Web sites being inundated

with resumes from unqualified job seekers; (iii) job advertisements resulting in applicant pools that do not reflect the demographic diversity desired; and (iv) newly hired workers, who had unrealistic job expectations, resigning after spending only a few weeks on the job. In order to understand how organizational recruitment activities are linked to these and other outcomes, several issues need to be considered, including what applicant population to target, how to reach these individuals, what recruitment message to convey, and whom to use as recruiters.

## 2. DECIDING ON RECRUITMENT OBJECTIVES

Prior to undertaking recruitment activities such as drafting a job advertisement or selecting individuals to serve as recruiters, it is important that an organization carefully considers its recruitment objectives. Although Breaugh reviewed numerous objectives that could be important to a given organization, this article focuses on four key questions that should be addressed in establishing recruitment objectives, and it shows how decisions concerning these objectives should guide subsequent recruitment actions. The first question an organization needs to answer is whether its focus is on filling a given position or on recruiting individuals who will progress through a series of positions over the course of a career with the organization. If the focus is on successfully filling a given position, then the employer does not need to be concerned with whether job applicants possess the talent and the motivation that would enable them to successfully perform higher level jobs.

A second important question is whether an employer prefers to fill jobs internally or externally. If the employer's focus is on recruiting from within its current workforce, it will not be difficult for the organization to reach the targeted population of potential recruits, nor is it likely that the firm will need to communicate information concerning organizationwide characteristics such as its values. In contrast, if an organization seeks to target applicants who do not currently work for it, more consideration will need to be given to how to reach desirable job candidates and what information to convey about the organization.

A third question meriting consideration in establishing recruitment objectives is whether an employer wants to attract a large number of applications or

fewer applications from a select group of individuals who possess certain job-related attributes. If the former is the case, then posting a job advertisement that lists few applicant requirements for hiring on a well-known job board may be an appropriate strategy. In contrast, if an employer only wants to attract applications from a limited number of individuals who meet restrictive requirements such as expertise in SAS programming, then a different recruitment approach (e.g., a direct mail campaign) may be appropriate.

A fourth question that merits consideration is whether demographic diversity is important. To the extent that it is, an employer would be wise to incorporate certain steps into its recruitment plan. Such steps might include recruiting at colleges with diverse student bodies, having a diverse group of recruiters, and including pictures of minorities and women in recruitment brochures. Having given serious consideration to these four questions should enable an employer to more precisely define the applicant population it is interested in reaching.

## 3. RECRUITMENT STRATEGY DEVELOPMENT

Having established the types of individuals it seeks to attract (e.g., persons with experience working in a team or retirees who may be interested in working part-time), an organization needs to consider what method(s) to use to reach them. If an employer decides to target individuals who currently work for it, the organization may decide to post job announcements on the company's Web site. Alternatively, in order to limit the number of applications received, some employers rely on supervisors to nominate subordinates who are seen as both qualified and interested in a job opening.

To reach individuals who do not work for an organization, organizations have utilized several different recruitment methods, including newspaper advertisements, employment agencies, school placement offices, job boards, employee referrals, radio ads, direct mail campaigns, and job fairs. Although research has not been entirely consistent, research reviewed by Barber showed that employee referrals are frequently among the best recruitment methods. The benefits associated with employee referrals may be due to current employees (i) providing potential referrals with a realistic preview of what a job involves, (ii) prescreening potential referrals so that only individuals who are seen as

qualified and motivated are referred, and (iii) helping the referred person succeed once on the job. A potential drawback of using an employee referral program is that it may not result in a diverse pool of applicants if the current workforce lacks diversity.

Some employers are able to generate job candidates without any formal recruitment activities. For example, some individuals apply for jobs (i.e., mail a resume) without knowing whether a job opening exists. As noted by Breaugh and Starke, such direct applicants are more likely if an organization is highly visible and has a positive reputation. A number of highly visible organizations, such as Southwest Airlines, have relied heavily on their Web sites to generate job applications. Little research exists on whether Web sites generate more desirable applicants than more traditional recruitment methods.

A number of recruitment methods, such as job fairs and the use of college placement offices, involve the use of corporate recruiters. As discussed by Rynes and Cable, recruiters play an important role because they convey information concerning a job opening, they prescreen job candidates, and they answer questions from prospective employees. The selection and training of recruiters are also important because job candidates may interpret how recruiters treat them as a signal of how the organization treats its employees. Given the significance of their role, it is not surprising that considerable research on recruiters has been conducted. Although results are not consistent, in general, job candidates rate recruiters as being more effective to the extent that they are interpersonally skilled and informative. Research has shown that although training can have a modest effect on the performance of recruiters, the selection of recruiters is likely to have a greater impact. In particular, an employer should give consideration to whether a staff recruiter is used as opposed to using someone from the department with the job opening. The advantage of using a staff recruiter is that the person may be better able to answer questions about such issues as employee benefits and may be less likely to ask a candidate inappropriate questions. However, utilizing someone from the department in which the job opening exists (e.g., the supervisor to whom the person would report or a prospective coworker) offers important advantages. For example, in general, someone from the department will be seen as more informative concerning job duties, work group climate, etc. In addition, someone from the department, especially a potential coworker, is likely to be seen as a credible source of information. Given that many job seekers are

cautious about accepting at face value all the information presented during the recruitment process, such credibility is important. Given their relative strengths and weaknesses, during the course of the recruitment process many employers use a team of recruiters composed of both line and staff recruiters. Regardless of the type of recruiter chosen, this decision should not be made lightly. Research by Rynes *et al.* showed that unprofessional treatment by a recruiter, such as failing to return phone calls or condescending behavior, can cause job candidates to withdraw as candidates.

After deciding on one or more recruitment methods to use to reach prospective employees, an organization needs to consider the information it wants to communicate to these individuals. In considering its message, one question an employer must address is whether it is interested in presenting a realistic view of a position or an overly positive one. A quick scan of the help wanted advertisements in a newspaper shows that many organizations convey an exaggerated view (e.g., can the pay for every job be competitive?). An employer should provide a realistic view of what a job involves. Presenting an exaggerated view of a position is unethical. Furthermore, it is likely to result in individuals accepting jobs that they quickly come to dislike. Wanous showed that such dissatisfaction is associated with a high rate of turnover for new employees. Such turnover can be quite expensive, especially if an employer has invested considerable training in the new employees or has reimbursed them for relocation expenses.

In phrasing a recruitment message, an employer should consider other factors in addition to realism. A key concern should be whether recruitment communications convey information that is considered important by recruits for making job choice decisions. Research has shown that most job candidates perceive information concerning the job (e.g., job duties and remuneration), the work group (e.g., coworker competence and whether interactions are cooperative or competitive), the supervisor (e.g., managerial style and technical competence), and the organization (e.g., corporate culture and economic conditions) as being particularly important. In addressing such topics, research has shown that recruits desire information that is sufficiently specific to allow for an informed job choice to be made. Thus, it is not sufficient for an organization to transmit information concerning how supervisors in general act or how coworkers in general relate. Instead, such information needs to be tied to a given job opening. Unfortunately, the recruitment messages sent by many organizations

fail to include the level of information specificity sought by many job candidates.

#### **4. CONSIDERING THE JOB APPLICANT'S PERSPECTIVE**

Filling an open position requires the agreement of both the hiring organization and the job candidate. Therefore, it is important that an organization consider the job applicant's perspective in planning the recruitment process. Because the types of information that many job applicants desire were discussed previously, it suffices to state here that an organization should do whatever it can to present such information in a credible manner. Off-the-record conversations with co-workers can be a particularly effective vehicle for conveying information that may be sensitive in nature (Does this organization truly have a merit-based reward system? Does the supervisor really involve employees in decision-making?). Such conversations are also effective for addressing the unique information needs of some job applicants, such as how easily a position will accommodate the observance of religious holidays.

If an organization seeks to hire new employees for whom there is good person–job fit, it is not sufficient for the organization to provide realistic and detailed information about important aspects of the position being filled. It also needs to be concerned with whether the job candidate can make sense of the information it conveys. For example, new employees may have been told that a job requires 60-hour workweeks or dealing with angry customers. However, if they lack relevant experience, some new hires may not be capable of fully grasping the meaning of the information conveyed by the organization. If an organization believes that inexperienced individuals will not be able to fully comprehend what a job involves, it may want to target individuals who can better appreciate what the job being filled demands. For example, individuals with prior experience working long hours should be better able to comprehend how a long workweek can interfere with family responsibilities. In order to maximize person–job fit, organizations should consider recruiting former employees whose prior performance with them was viewed favorably. If successfully recruited, former employees should have an excellent understanding of job and organizational attributes and thus have realistic job expectations.

#### **5. ASSESSING RECRUITMENT RESULTS**

After an organization has completed its recruitment activities for filling one or more job openings, it is important for it to evaluate the effectiveness of these activities. For example, an employer could compare recruiters, colleges visited, or advertising outlets (e.g., newspaper versus radio) in terms of the number of job applicants generated, the number of individuals hired, the cost per hire, the diversity of those hired, and/or the first-year performance of those hired. Unless an assessment of recruitment activities is conducted, an organization may not discover that certain activities have more value than others. For example, although employers frequently rely on job advertisements as a method for generating job candidates, research has shown that the use of employee referrals is a better method for generating employees who perform well and who remain on the job for a longer period of time. A particularly important source of information for assessing the effectiveness of recruitment actions can be newly hired employees and job applicants who were not hired. These individuals can provide insight into how they were treated during various stages of the recruitment process (e.g., whether phone calls were returned in a timely manner and whether a site visit was professionally conducted). Such information should provide those responsible for recruitment decision making with valuable ideas for improving the effectiveness of future recruitment efforts.

#### **6. CONCLUSION**

In staffing an organization, attention is frequently given to the soundness of a firm's selection system. Although using valid selection devices is clearly important, it is increasingly being recognized that poor recruitment activities can result in a pool of job applicants who lack the ability and/or motivation to successfully fill job openings. Such an applicant pool can restrict the value of even the best selection system. This article addressed a number of issues that can have a major impact on the overall success of a recruitment campaign. Specifically, it discussed key issues that an organization should consider in deciding whom to target for recruitment, how to reach them, what recruitment message to convey, and whom to use as recruiters. Thoughtful consideration

of these issues is likely to result in a more desirable applicant pool (e.g., more diverse, more qualified, and more interested) and in a more desirable group of new employees (e.g., higher job satisfaction and longer tenure).

### See Also the Following Articles

Organizational Socialization ■ Person–Environment Fit  
 ■ Personnel Psychology ■ Personnel Selection

### Further Reading

- Barber, A. E. (1998). *Recruiting employees*. Thousand Oaks, CA: Sage.
- Breaugh, J. A. (1992). *Employee recruitment: Science and practice*. Boston: PWS-Kent.
- Breaugh, J. A., & Starke, M. (2000). Research on employee recruitment: So many studies, so many remaining questions. *Journal of Management*, 26, 405–434.
- Greenhaus, J. H., Callanan, G. A., & Godshalk, V. M. (2000). *Career management*. Fort Worth, TX: Dryden Press.
- Rynes, S. L., Bretz, R. D., & Gerhart, B. (1991). The importance of recruitment in job choice: A different way of looking. *Personnel Psychology*, 44, 487–521.
- Rynes, S. L., & Cable, D. M. (2003). Recruitment research in the twenty-first century. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology* (Vol. 12, pp. 55–76). Hoboken, NJ: Wiley.
- Taylor, M. S., & Collins, C. J. (2000). Organizational recruitment: Enhancing the intersection of research and practice. In C. L. Cooper, & E. A. Locke (Eds.), *Industrial and organizational psychology: Linking theory with practice* (pp. 304–334). Oxford, UK: Blackwell.
- Wanous, J. P. (1992). *Organizational entry*. Reading, MA: Addison-Wesley.



# Rehabilitation Counseling

Timothy R. Elliott

University of Alabama, Birmingham, Alabama, USA

Doreen Miller

Southern University, Baton Rouge, Louisiana, USA

1. Introduction
  2. Defining Roles and Functions
  3. History and Evolution
  4. Critical Legislation
  5. Settings
  6. Research Base
  7. Summary
- Further Reading

Rehabilitation counseling is a counseling profession noted for its emphasis on serving persons with various physical, intellectual, behavioral, sensory, and developmental disabilities. The profession has been characterized by its counseling model and its responsiveness to state and federal mandates to coordinate the delivery of services to eligible individuals with disabling conditions to promote their employability and independence.

## GLOSSARY

**vocational evaluation** A systematic and comprehensive assessment process to identify an individual's vocational potential; the evaluation may include, but is not limited to, self-report and observational assessment of a person's vocational and career interests, values, temperaments, work-related behaviors, aptitudes, skills, physical capacities, learning style, and training needs.

**vocational rehabilitation** The provision of services to enhance the employability and functional independence of a person who has been adversely impaired in these roles by a physical, intellectual, sensory, behavioral, intellectual, or developmental disability.

**work adjustment** The degree of congruence and correspondence between an individual's values and behavior and the behavior that is valued, accepted, and expected in a work environment.

## 1. INTRODUCTION

Rehabilitation counseling is a relatively young profession. In contrast to most human service professions, the birth and development of rehabilitation counseling are rooted in federal legislation. The genesis of the profession can be traced to the early 20th century in the United States. The establishment of the profession was precipitated by a number of social factors, the most fundamental of which was the need to provide rehabilitation services for an increasing number of people who were unable to work due to physical disabilities that were often incurred in the rapidly changing workplace. Although there was a need to provide services for these persons, there was no existing professional field available. Thus, the federal government mandated that the profession provide expert coordination and

administration of necessary services to citizens who were eligible for assistance.

## 2. DEFINING ROLES AND FUNCTIONS

Rehabilitation counseling initially focused on helping clients to become employable so that they could return to the workforce. This goal has expanded over the years to subsume a systematic process that assists persons with various disabilities to achieve their personal, career, and independent living goals in the most integrated settings possible. The rehabilitation counseling process involves communication, goal setting, and beneficial growth or change through self-advocacy, psychological, vocational, social, and behavioral interventions. The specific techniques and competencies used in rehabilitation counseling may include, but are not limited to, the following:

- Assessment, vocational evaluation, and appraisal
- Diagnosis and treatment planning
- Career (vocational) counseling
- Individual and group counseling treatment interventions focused on facilitating adjustment to the medical and psychosocial impacts of disability
- Case management, referral, and service coordination
- Program evaluation and research
- Interventions to remove environmental, employment, and attitudinal barriers as well as to enhance work adjustment
- Knowledge of disability-related legislation and regulations and of community resources
- Consultation services among multiple parties and regulatory systems
- Expert witness testimony
- Job analysis and modification, job development, and placement services, including assistance with employment and job accommodations
- The provision of consultation about access to relevant rehabilitation technologies

Currently, most rehabilitation counselors possess a master's degree in rehabilitation counseling (or in a related field) and are credentialed. The types of professional credentials available to a rehabilitation counselor fall into three general categories:

- Registration
- Licensure
- Certification

## 3. HISTORY AND EVOLUTION

Vocational rehabilitation (VR) emerged during the early years of the 20th century as federal policy-makers wrestled with the consequences of an increasingly industrialized and urban economy as well as the ravages of international conflict. The Smith–Hughes Act of 1917 provided money to states on a matching basis to establish vocational education programs and also created the Federal Board of Vocational Education to administer VR programs. Relevant training and rehabilitation programs were provided under this legislation to meet current workforce demands. Parenthetically, this act established the first rehabilitation education programs within colleges of education in universities; rehabilitation counseling traditionally has been housed within this academic environment. The Soldier's Act of 1918 created VR programs for returning soldiers with disabilities, and the Federal Board of Vocational Education administered these programs.

Training programs were established in academic institutions, and considerable debate concerning the appropriate “intellectual home” for rehabilitation counseling ensued. One contingent viewed rehabilitation counseling within the larger domain of counseling psychology because rehabilitation counseling was perceived as a logical extension of the field's emphasis on vocational counseling, career development, and recognition of everyday people experiencing transitional stress in personal and social roles. Over time, however, critics expressed disapproval of “psychologized” coursework in graduate programs and argued that esoteric theories were of little use in the work setting. In addition, the multidisciplinary nature of VR, generally had little tolerance for theories that could not be easily conveyed or understood across disciplines. A greater emphasis on labor market issues, job assessment, work adjustment, and job placement was encouraged. The premium on practitioner competencies and issues permeates graduate training in rehabilitation counseling. Although coursework in areas considered “core” to psychology training may be offered, these are usually confined to areas of applied interest (e.g., assessment) and others may be permitted as electives (e.g., personality, abnormal behavior). In some institutions, rehabilitation counseling students may require the permission of the psychology department to enroll in any graduate psychology course.

#### 4. CRITICAL LEGISLATION

Subsequent legislation increased the role of the federal government in preparing persons with disabilities for employment. For example, the Smith–Fess Act of 1920 increased the amount of money provided to state programs, and the 1935 Federal Social Security Act made the federal–state VR programs permanent. Other legislation extended VR services to blind individuals (the Randolph–Shepard Act of 1936) and to persons with mental retardation or psychiatric disabilities (the Barden–Lafollette Act of 1943).

The Vocational Rehabilitation Act of 1954 ushered in the “golden age” of rehabilitation and essentially created the rehabilitation counseling profession. It also dictated the model for systematically providing VR services and provided money to expand facilities. Services were also expanded to persons with substance abuse problems, public offenders, and the socially disadvantaged. The Rehabilitation Act of 1973 made services to persons with the most severe disabilities a top priority for state programs. It instituted the use of individualized written plans for clients and welcomed consumer input into the VR process. It also advanced a civil rights agenda for persons with disabilities by promoting access, prohibiting architectural barriers, and prohibiting discrimination in educational institutions and in the workplace.

Contemporary legislation has had a dramatic impact on rehabilitation counseling in terms of training and opportunities. The passage of the Americans with Disabilities Act in 1990 expanded the civil rights of persons with disabilities and reinforced the consultative role of rehabilitation counselors. However, a series of legislative acts appear to blur the support and role of rehabilitation counseling and its traditional role in the VR process. Specifically, the Workforce Investment Act in 1998 and the Ticket-to-Work/Work Incentive Improvement Act in 1999 were passed in part to circumvent perceived “red tape” that was associated with the traditional VR process as well as to encourage greater client input and options for rehabilitative services. These policies provide incentives for other entities in the private sector to offer and provide competitive rehabilitation services to consumers.

#### 5. SETTINGS

From its inception, rehabilitation counseling has been virtually synonymous with state VR agencies, as

originally mandated by federal legislation. This association continues to the current time. Traditionally, rehabilitation counselors have also found employment with Veterans Administration agencies, state employment agencies, state health departments, hospitals, mental health centers, prisons and correctional facilities, independent living centers, educational institutions (and training programs for students with disabilities), job training centers, social security offices, and workers’ compensation programs. For decades, rehabilitation counselors were considered among the chief advocates for persons with disabilities and were seen as experts concerning these persons’ rehabilitation and integration into society. Many rehabilitation counselors actively participated in the independent living movement for persons with disabilities.

However, changes in legislation and increases in the incidence of chronic disease and disability over the past three decades opened new opportunities in private business, substance abuse treatment programs, and employee assistance programs. Many rehabilitation counselors obtained positions in health care management programs, with insurance companies, and with third-party payers as case management specialists. Those in private practice often provide expert testimony, consultation, life planning services (for workers who incur lifelong severe disabilities), and vocational evaluation for legal firms, health care systems, businesses, and other agencies.

Individuals who possess a master’s degree in rehabilitation counseling may pursue licensure as licensed professional counselors. With this licensure, they may be able to provide a wide range of mental health services for reimbursement. Most doctoral-level rehabilitation counselors retain their status as certified rehabilitation counselors. Others may qualify for psychological licensure in some states (depending on the graduate program content and the accreditation status of the department conferring the doctorate). These individuals may identify with rehabilitation psychology and pursue the diplomate in rehabilitation psychology offered under the auspices of the American Board of Professional Psychology. Doctoral-level rehabilitation counselors also assume academic positions where they are distinguished by their research, service, and training programs. Most rehabilitation counseling training programs are housed in colleges of education, a few are found in colleges associated with allied health professions, and some are found in psychology departments.

## 6. RESEARCH BASE

As a multidisciplinary enterprise that is constantly responsive and sensitive to the vagaries of federal legislation, fluctuating state budgets, and labor market trends, the VR literature has been criticized in academic circles for lacking theory-driven research (undermining its scholarly relevance to other fields) and an archival base (in which current knowledge builds on, advances, and refines prior knowledge). In a seminal text, Wright argued in 1980 that rehabilitation counseling has an established literature base that has steadily advanced the knowledge and sophistication of the profession. The peer-reviewed outlets associated with the field showcase a great diversity of topics, methodological rigor, and practitioner concerns. The *Journal of Rehabilitation* and the *Journal of Applied Rehabilitation Counseling* (both affiliated with the National Rehabilitation Association) have a clear emphasis on practitioner interests. The *Rehabilitation Counseling Bulletin* (affiliated with the American Rehabilitation Counseling Association) publishes studies of theoretical and practical interest and has a reputation for high standards in research design and methodology. The *Vocational Evaluation and Work Adjustment Bulletin* historically has focused on psychometric and practice issues germane to vocational assessment. *Rehabilitation Psychology* (affiliated with the American Psychological Association) publishes some work relevant to rehabilitation counseling. *Disability and Society* features work germane to rehabilitation counselors from a perspective associated with academic programs in disability studies and proponents of independent living programs for persons with disabilities. Other relevant outlets include the *Journal of Vocational Rehabilitation*, *Journal of Rehabilitation Administration*, and *Rehabilitation Education*.

The rehabilitation counseling literature has attended to the competencies and issues that face the field, including training, credentialing, and roles and functions. In addition, this literature has valued studies that address practice concerns associated with measurement, program evaluation, specific techniques and their application, and studies of psychosocial adjustment and vocational outcomes of clients with specific disability conditions. Rehabilitation counseling research often scrutinizes vocational evaluation practices and effectiveness as well as program and outcome evaluations of state VR programs. Recent research has

attended to multicultural issues and practice implications of recent legislation. Many rehabilitation counselor researchers have considerable expertise in psychometric theory because the field has systematically evaluated the psychometric properties and appropriateness of psychological measures for use with persons with various physical, sensory, and intellectual disabilities.

A perusal of citation behaviors in available search engines suggests that research published in the outlets traditionally associated with rehabilitation counseling is somewhat insular because relatively few outlets outside of the field cite work published in these journals. A notable exception can be found in the contributions that rehabilitation counseling has made in the area of job placement. The practice of supported employment, which developed as a placement technique for persons with severe disabilities, has emerged as an effective placement strategy in several published clinical trials. The approach requires job specialists to observe and correct problematic client behaviors in the workplace over staggered and intermittent periods of time to promote work adjustment and model appropriate work behaviors. The practice has demonstrated efficacy in studies of persons with traumatic brain injuries, mental retardation, severe mental disorders, and sensory disabilities. Clinical trials supporting this technique have been published in mainstream (and influential) psychology and psychiatric journals.

## 7. SUMMARY

Rehabilitation counseling has evolved in response to federal legislation and labor market trends. Rehabilitation counselors in state agencies have been mandated to prioritize services for eligible clients with the most severe disabilities. During the time frame in which this mandate has been active, available data indicate that competitive employment placements increased substantially despite declining state budgets and increased expense for these clients. The profession will likely face competition from other professions that respond to recent legislation and provide VR services and placement to persons with disabilities, but it is probable that the field will continue its association with state agencies and thrive in case management roles in the private sector. Rehabilitation counseling has imbued its practitioners with unique skills in case management, job placement, and vocational evaluation,

and these skills serve the profession well in the public and private sector. Those with advanced training will also provide invaluable service in the program evaluation, training, and consultation that ensure the future of the profession.

### See Also the Following Articles

Assessment and Evaluation, Overview ■ Job Analysis, Design, and Evaluation ■ Psychological Rehabilitation Therapies ■ Work Adjustment

### Further Reading

- Bond, G. R., Becker, D., Drake, R., Rapp, C., Meisler, N., Lehman, A., Bell, M., & Blyler, C. (2001). Implementing supported employment as an evidence-based practice. *Psychiatric Services*, *52*, 313–322.
- Chan, F., & Leahy, M. (Eds.). (1999). *Health care and disability case management*. Lake Zurich, IL: Vocational Consultants Press.
- Elliott, T., & Leung, P. (in press). Vocational rehabilitation: History and practice. In W. B. Walsh, & M. Savickas (Eds.), *Handbook of vocational psychology* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Growick, B. (2000). The political implications of the TTW–WHA. *Rehabilitation Education*, *15*, 89–93.
- McGowan, J., & Porter, T. (1967). *An introduction to the vocational rehabilitation process: A training manual*. Washington, DC: U.S. Department of Health, Education, and Welfare.
- Parker, R. M., & Szymanski, E. (Eds.). (1998). *Rehabilitation counseling: Basics and beyond*. Austin, TX: Pro–Ed.
- Rubin, S. E., & Roessler, R. T. (2001). *Foundations of the vocational rehabilitation process* (5th ed.). Austin, TX: Pro–Ed.
- Shrey, D. E., & Lacerte, M. (1995). *Principles and practices of disability management in industry*. Winter Park, FL: GR Press.
- Szymanski, E. M., & Parker, R. M. (Eds.). (1996). *Work and disability*. Austin, TX: Pro–Ed.
- Wright, G. N. (1980). *Total rehabilitation*. Boston: Little, Brown.



# Religion and Mental Health

Adam B. Cohen and Harold G. Koenig

Duke University Medical Center, Durham, North Carolina, USA

1. Introduction
  2. Benefits to Mental Health
  3. Applications of Religion
  4. Caveats to Researchers
  5. Testing Prayer?
  6. Negative Implications of Religion
  7. Training Counselors and Consultants
  8. Conclusion
- Further Reading

## GLOSSARY

**caregiver** A person who takes care of patients with end-stage cancer or Alzheimer's disease; often a relative.

**positive psychology** Psychological research on positive emotions, such as joy, wonder, happiness, satisfaction, meaning, purpose, and hope.

In the past, the predominant view was that religion is inherently unhealthy. Recent research has begun to convincingly challenge this assumption. In hundreds of studies, religion has been associated with positive mental and physical health outcomes and with positive emotions. Although most studies have been cross-sectional, some have been longitudinal and several randomized clinical trials have also been completed. Despite these provocative associations, it is important to keep in mind several caveats. There may be significant scientific and theological concerns surrounding the

prescription of religion and the empirical investigation of religion's effects on health. Moreover, not all religions may relate to health in the same ways, and in some cases religion may have decidedly negative consequences for health. Psychiatric and psychological training is beginning to emphasize a need to address religious and spiritual issues as part of clinical practice.

## I. INTRODUCTION

For quite some time, many influential mental health professionals conceptualized religion and mental health as being diametrically opposed, and only within the past 10 years have such notions been challenged. Religion and spirituality have many components that are of potential relevance to mental health, including religious attendance, private religious activities (e.g., prayer and reading of holy scriptures), a feeling of connection or relationship with God or a higher power, religious beliefs, and religious coping. In addition, religion and spirituality may cause mental health via healthy lifestyles and behaviors and the promotion of social support. Some writers distinguish religion from spirituality by claiming religion is more organized, whereas spirituality has more to do with one's personal relationship with God or a higher power.

In 2001, Koenig *et al.* performed a comprehensive, systematic review of research on religion and mental health and identified 724 quantitative studies that

examined this association. Many different populations—from young to old, from sick to healthy—have been examined. Studies have been conducted in almost every area of the world, including the United States, Canada, England, continental Europe, Australia, China, Malaysia, Egypt, India, and Israel. The majority of studies (478; 66%) reported statistically significant relationships between religious involvement and better mental health, greater social support, or less substance abuse. Most of the studies were cross-sectional, but a number have been longitudinal and there are several clinical trials that have examined the effects of religious interventions on mental health. Longitudinal studies have also shown faster recovery from depression in community samples and more rapid adaptation to stressful life circumstances in religious caregivers. Several clinical trials examining the effects of a religious-based psychotherapy on treatment outcomes in depression, bereavement, and anxiety have likewise documented benefits for both Christian and Muslim patients.

## 2. BENEFITS TO MENTAL HEALTH

Mental health involves not only the freedom from mental disorders, such as depression, anxiety, psychotic conditions, or personality problems, but also the regular experience of positive emotions, such as joy, wonder, happiness, satisfaction, meaning, purpose, and hope. Relatively little research has been done on positive emotions and mental states. This is changing with the growth in “positive psychology,” which has been led by Martin Seligman, former president of the American Psychological Association. Religiousness appears to be related to positive emotions, where the relationships may be even stronger and may have implications for the prevention of mental disorder. Of the approximately 100 studies conducted during the past century on religiousness and well-being, hope, and optimism, nearly 80% have found that religious persons experience significantly more positive emotions compared to those who are less religious. Similarly, in a recent study, measures of religious belief, coping, and spirituality were associated even more strongly than was social support with measures of life satisfaction among Protestants and Catholics, which is consistent with a role of religion-influenced positive emotion in well-being.

## 3. APPLICATIONS OF RELIGION

It is important to emphasize that clinical applications of religion must be theologically and ethically informed. For example, it is not only impractical but also may be theologically and ethically inadvisable to prescribe religious behaviors on the part of patients, for a clinician to impose his or her religious beliefs onto patients, or to create an atmosphere of intolerance for patients of diverse faith communities.

## 4. CAVEATS TO RESEARCHERS

From a research perspective, it is crucial to be theologically savvy. One aspect of this is to recognize that not all religions are the same, and that people from different faith communities may differ in psychologically relevant ways. One example is the salience of certain mental states (e.g., faith) to Jews and Protestants. For example, in the religion/well-being studies mentioned previously, social support from religious sources correlated modestly with well-being for Jews, Protestants, and Catholics. However, religious belief, spirituality, and religious coping were much more related to well-being among Catholics and Protestants than among Jews.

## 5. TESTING PRAYER?

Another aspect of being theologically informed in examining religion and mental health connections has been raised in the context of research on the effects on health of distant intercessory prayer. Chibnall *et al.* argued that such research may border on heresy because it tests or challenges God, that operational definitions of prayer may fly in the face of theological definitions, and that results are incredibly difficult to interpret. For example, do positive effects of prayer on health prove the existence of God? What do negative effects indicate? What if prayers by members of one faith community turn out to be more efficacious than prayers by those of another faith community?

## 6. NEGATIVE IMPLICATIONS OF RELIGION

Religious beliefs or activities may sometimes be associated with worse mental health or neurotic behavior.

Most clinicians today know from experience that religion may be utilized in an unhealthy manner or manipulated to serve defensive functions that ultimately impair mental health and prevent healthy growth. For example, religious concerns are common among people with obsessive-compulsive disorders and even people in nonclinical samples, including fears of sin, fears of God, and viewing the moral status of thoughts to be equal to that of actions. Vivid examples can be seen when individual cases are examined: the rigid fundamentalist; the obsessive-compulsive who attends church daily and prays 10 or 20 times per day; the psychotic patient who believes he is God, Jesus, or the devil; and the severely depressed person who is overwhelmed with guilt, believing that she has committed the unpardonable sin. This does not necessarily mean that religion promotes mental ill health, since ill people often turn to religion for comfort.

## 7. TRAINING COUNSELORS AND CONSULTANTS

Training programs in both psychiatry and psychology are beginning to emphasize the need for mental health professionals to address religious or spiritual issues as part of routine mental health care. This is now a requirement for accreditation of psychiatric residency programs by the American College for Graduate Medical Education (1994), which has mandated that training programs include didactic training in the “presentation of the biological, psychological, sociocultural, economic, ethnic, gender, religious/spiritual, sexual orientation and family factors that significantly influence physical and psychological development in infancy, childhood, adolescence, and adulthood” (pp. 11–12). Training programs in psychology are also required to ensure that the curriculum “includes exposure to theoretical and empirical knowledge bases relevant to the role of cultural and individual diversity,” and religion is included in the definition of cultural and individual diversity.

As a result of increased training and exposure to research on the relationship between religion and health, there appears to be more widespread integration of religion into clinical practice, with therapists now utilizing patients’ religious beliefs and social connections within the faith community to facilitate healing and recovery. Not only is the area of pastoral counseling growing rapidly but also there are indications that

secular therapists are beginning to address spiritual issues when treating patients. Unfortunately, systematic research documenting when, where, and how frequent such integration is taking place has not been done.

## 8. CONCLUSION

Religion and mental health appear to be related—either for better or for worse. Clinicians need to be aware of the research that has been done in this area and understand how to both sensibly and sensitively apply this knowledge to clinical practice.

## Acknowledgments

This work was partially supported by National Institute on Aging Grant AG00029 and a grant from the Positive Psychology Network to Adam Cohen. The views expressed do not necessarily represent those of these organizations.

## See Also the Following Articles

Coping ■ Well-Being

## Further Reading

- Abramowitz, J., Huppert, J., Cohen, A. B., Cahill, S. P., & Tolin, D. F. (2002). Religious obsessions and compulsions in a non-clinical sample: The Penn Inventory of Scrupulosity. *Behaviour Research and Therapy*, 40, 825–838.
- Accreditation Council on Graduate Medical Education. (1994, March). *Special requirements for residency training in psychiatry*. Chicago: Accreditation Council on Graduate Medical Education.
- American Psychological Association. (1995). Guidelines and principles for accreditation of programs in professional psychology. In *American Psychological Association Committee on Accreditation site visitor handbook* (pp. 5, 1-27). Washington, DC: American Psychological Association.
- Azhar, M. Z., & Varma, S. L. (1995a). Religious psychotherapy in depressive patients. *Psychotherapy & Psychosomatics*, 63, 165–173.
- Azhar, M. Z., & Varma, S. L. (1995b). Religious psychotherapy as management of bereavement. *Acta Psychiatrica Scandinavica*, 91, 233–235.
- Azhar, M. Z., Varma, S. L., & Dharap, A. S. (1994). Religious psychotherapy in anxiety disorder patients. *Acta Psychiatrica Scandinavica*, 90, 1–3.
- Braam, A. W., Beekman, A. T. F., Deeg, D. J. H., Smith, J. H., & van Tilburg, W. (1997). Religiosity as a protective or

- prognostic factor of depression in later life; Results from the community survey in The Netherlands. *Acta Psychiatrica Scandinavica*, 96, 199–205.
- Chibnall, J. T., Jeral, J. M., & Cerullo, M. A. (2001). Experiments on distant intercessory prayer: God, science, and the lesson of Massah. *Archives of Internal Medicine*, 161, 2529–2536.
- Cohen, A. B. (2002). The importance of spirituality in well-being for Jews and Christians. *Journal of Happiness Studies*, 3, 287–310.
- Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology*, 81, 697–710.
- Hill, P. C., Pargament, K. I., Hood, R. W., McCullough, M. E., Swyers, J. P., & Larson, D. (2000). Conceptualizing religion and spirituality: Points of commonality, points of departure. *Journal for the Theory of Social Behaviour*, 30, 51–77.
- Koenig, H. G. (2002). *Spirituality in patient care: Why, how, when, and what*. Radnor, PA: Templeton Foundation Press.
- Koenig, H. G., McCullough, M., & Larson, D. B. (2001). *Handbook of religion and health: A century of research reviewed*. New York: Oxford University Press.
- Levin, J. S., & Chatters, L. M. (1998). Research on religion and mental health: An overview of empirical findings and theoretical issues. In H. G. Koenig (Ed.), *Handbook of religion and mental health* (pp. 33–50). San Diego: Academic Press.
- Pargament, K. I. (1997). *The psychology of religion and coping: Theory, research, practice*. New York: Guilford.
- Propst, L. R., Ostrom, R., Watkins, P., Dean, T., & Mashburn, D. (1992). Comparative efficacy of religious and nonreligious cognitive-behavior therapy for the treatment of clinical depression in religious individuals. *Journal of Consulting and Clinical Psychology*, 60, 94–103.
- Seligman, M. E. P. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. New York: Free Press.
- Shafran, R., Thordarson, D. S., & Rachman, S. (1996). Thought-action fusion in obsessive compulsive disorder. *Journal of Anxiety Disorders*, 10, 379–391.
- van Oyen Witvliet, C., Ludwig, T. E., & Vander Laan, K. L. (2001). Granting forgiveness or harboring grudges: Implications for emotion, physiology, and health. *Psychological Science*, 12, 117–123.
- Zinnbauer, B. J., Pargament, K. I., & Scott, A. B. (1999). The emerging meanings of religiousness and spirituality: Problems and prospects. *Journal of Personality*, 67(6), 889–919.



# Residential Preferences and Attachment across the Lifespan

M. Vittoria Giuliani

Consiglio Nazionale delle Ricerche, Rome, Italy

1. Theories and Concepts
2. Housing Attributes and Preferences
3. Personal and Social Influences in Preference Ratings
4. Moving Home and Attachment  
Further Reading

## GLOSSARY

**actual–aspirational gap theory** A model that postulates that people perceive salient attributes of their physical environment and evaluate them based on certain standards of comparison, especially standards defined by what people believe they may reasonably aspire to.

**appropriation of place** The process by which people become more closely associated with or involved in places through taking possession, sharing and communicating social meanings, and manipulating and personalizing the environment.

**cognitive dissonance** A state of tension or discomfort resulting from inconsistency among beliefs or between one's beliefs and one's actions.

**housing bundle** The set of physical (e.g., dwelling size, building type, and neighborhood facilities) and nonphysical attributes (e.g., tenure) defining a housing alternative.

**life values** Desirable end states that the individual strives to attain in his or her life.

**person–environment fit theory** A model that assumes that congruence between personal preferences or needs and environmental press fosters environmental satisfaction and psychological well-being.

**place attachment** An affective bond between an individual and a particular place not interchangeable with another with the same functional quality.

**place identity** All aspects of self-identity having place-related implications.

**proactive behavior** Acting in advance to prevent potential stressors and to promote personal growth.

**settlement identity** Pattern of cognitions, feelings, behavioral tendencies, and skills that relate the identity of a person to a type of settlement and provide dispositions for future engagement with that type of settlement.

To what extent do places where people live meet their needs and desires? Family life and social life are changing dramatically in many areas of the world. How are these changes reflected in the meaning people attribute to their residence and the way they assess it? During the past few decades, researchers from different disciplines have addressed these issues from different perspectives, sharing the assumption that the quality of residential environment plays a central role in the global quality of life.

## 1. THEORIES AND CONCEPTS

What is the residential environment? As an environmental domain, the residential environment is usually

defined in contrast to other environmental domains, such as the working environment and the public environment. In both cases, the reference is to a concept of residence that is private, based on the model that has evolved in Western societies since the industrial revolution.

However, the home/work contrast has been questioned in the Western world both from a feminist perspective and in relation to the advent of telework, which has focused attention on a phenomenon—home-based work—that has in fact always existed but is currently growing.

Regarding the public/private contrast, cross-cultural research has revealed that the boundaries between the two spheres can be very fluid and do not necessarily coincide with the domestic confines. Institutional residences, such as student dormitories or nursing homes, can hardly be considered private environments.

Finally, it should be kept in mind that the residential environment has no definable spatial boundaries. One can approach residential preference at the level of room and dwelling unit or include the immediate surroundings, the neighborhood, or even the larger community. The term housing bundle, used to indicate the pattern of attributes defining a housing alternative, includes items pertaining to different scale units (i.e., from both the micro- and the macroenvironment). Research has shown that for different groups of people, salient environmental units may be different; furthermore, when assessing housing bundles, including both the dwelling unit and the neighborhood dimension, people modify the importance assigned to the attributes of housing or neighborhood when evaluated in isolation.

This article focuses on the domestic environment. Aspects related to neighborhood satisfaction and preferences are reviewed elsewhere in this encyclopedia and will only be mentioned here in reference to the location of the dwelling unit.

### **1.1. The Meaning of Residential Environment and Home Attachment**

In the study of residential environment, a distinction is regularly made between house and home. A house (or apartment) is defined by its physical structure, and a home is defined by a set of meanings. Research on housing is focused on the spatial and behavioral aspects of living, whereas that on the home deals with the symbolic and emotional meaning of the living

experience. The term dwelling has been put forward as a link between house and home: Dwelling means the process by which a physical entity takes on psychological, symbolic, and social meaning, transforming it into a home.

Many psychological and sociocultural dimensions have been identified to distinguish between home and nonhome and to grasp personal differences in the home meaning: security and control, reflection of one's ideas and values, acting upon and modifying, permanence and continuity, place for relationships with family and friends, center of activity, refuge from the outside world, indicator of personal status, material or physical structure, and a place to own.

A comparison of home and nonhome reveals that the meanings attributed to the former generally have a positive connotation, whereas those associated with the latter are generally negative in physical as well as social or personal terms. Recently, there has been an increasing focus on the negative aspects of home experience, viewed, for example, as a constraint (for the old or the unemployed) or as a place of possible violence or oppression, particularly for women and children. In addition, extending analysis of the meaning of home to different cultures, other meanings emerge, such as religious meanings as evidenced, for example, by the beliefs of Hindu and Muslim women.

To characterize the variety of representations and use of space in homes in different cultures, and also the changes that occur in the same culture over time, Altman and colleagues adopted a dialectic approach in which the sociopsychological meaning of the home is given by a limited number of bipolar dimensions that characterize the individual's relationship with society, such as identity/communality and accessibility/inaccessibility (or openness/closedness). In relation to preferences, it can therefore be said that cultural and individual differences are represented by the changing balance in the relative strength of the oppositional poles that define each dimension.

Many researchers have argued that what defines the very nature and essence of home, as distinguished from house, is the emotional relationship with the place. Based on earlier theoretical work on place and home by phenomenologists and humanistic geographers, a growing body of psychological literature has been devoted to the study of affective bonds that people develop with places. The concept used in this context is place attachment. Although it is argued that people may develop feelings of attachment to places of different spatial scale—residence, neighborhood, city,

region, country, etc.—and to a variety of types of setting, from the work environment to natural settings, empirical research in this area has focused on residential settings. Although from a spatial standpoint environmental units are “nested” within each other, to be attached to one’s city does not necessarily imply being attached to one’s neighborhood, house, or apartment: The degree of attachment to different environmental units varies in relation to sociocultural and demographic variables, such as social position, age, and gender.

Attachment has been considered in relation to residential satisfaction—even though the direction of the relationship is not clear (e.g., is one more satisfied because one is more attached, or is one more attached because one is more satisfied?)—to residential mobility, and to residential choices. With respect to satisfaction, attachment represents a comprehensive measure, including even behavioral and emotional aspects that go beyond a mere affective response. In fact, what qualifies attachment is not the positive valence of affects but the fact that it is perceived as a bond, with an enduring quality, directed toward a specific target, not interchangeable with any other with the same functional quality. Consequences of attachment have been postulated at the level of social involvement and mobility and at the level of well-being and mental health.

Moreover, attachment is not necessarily positive; it can represent a source of individual hardship or be the result of conformist attitudes or social rejection. It has also been suggested that place attachment develops and supports place identity; that even in mobile countries such as the United States, the majority of people form affective bonds with places; and that changes of place consistent with changes occurring in their needs and identity during their lifetime may actually favor rather than hinder the formation of affective bonds.

### **1.2. Preferences, Satisfaction, and Housing Choices**

What kind of residential environment is good for what kind of person? Environmental psychologists have tried to answer this fundamental question by studying residential satisfaction and preferences. Although the terms satisfaction and preferences are often used interchangeably to indicate what people like, they are two distinct concepts. Both have to do with an assessment, but in the former case the fulfillment of needs is evaluated, whereas in the latter case the fulfillment of

desires is evaluated. Furthermore, although satisfaction is generally measured in relation to the actual environment, preferences can also refer to an ideal environment (i.e., not bound by the limitations of the individual or the constraints of the market). Finally, strictly speaking, preferences imply a choice among alternatives.

Preferences and satisfaction have been put in direct relation in different models of residential satisfaction, such as the actual–aspirational gap theory or the person–environment fit theory. Both models postulate that satisfaction derives from the perceived congruence between personal preferences or aspirations (which are in turn influenced by the personal characteristics of residents) and the objective and subjective characteristics of a place. Satisfaction has therefore often been used as an indirect indicator of preference. Moreover, both preferences and satisfaction have been considered to be predictive of behavioral intentions and actual behavior. However, various studies have shown a low correlation both between choices and preferences and between satisfaction and preferences. Although the explanation in the first case refers especially to external factors (e.g., market restrictions) and only to a lesser degree to other factors, such as the negotiating process with other members of the household, the discrepancy between satisfaction and preference is generally explained by psychological factors, such as processes of adaptation and assimilation or mechanisms of cognitive dissonance. Highly positive responses often reflect or even conceal an adjustment to inevitable inadequacies.

How is the meaning attributed to the environment related to the preferences toward the objective attributes of the housing? The subject has not been investigated systematically. Most research on housing preferences has simply been aimed at identifying correlations between objective environmental attributes and subjective evaluation (satisfaction and preferences) and/or personal and social influences on subjective evaluation. The two lines of research on subjective evaluation and the psychological meaning of the residential environment have largely remained separate.

A more general question is whether residential preferences and choices can be predicted from subjective beliefs about the effect of different housing alternatives on the attainment of life values (e.g., comfort, family, freedom, security, togetherness, and money). Unfortunately, limited attention has been devoted to this issue. Results provide evidence that beliefs determine the evaluations. In addition, they suggest

important differences between residential preferences and simulated choices, particularly that preferences do not completely explain choices and that choices, unlike preferences, are better explained by the overall evaluation of all attributes than the evaluation of single attributes.

### **1.3. Evaluative Perspectives and Aesthetic Preferences**

Asking someone to express his or her preference for a residential environment is like asking whether that person would like to live in a certain place. However, preferences can be different depending on whether the place would be a permanent residence or a vacation home, a place to live alone or with the family, etc. Furthermore, as already mentioned, a residence has to satisfy a multiplicity of needs, not the least of which is aesthetic. Satisfaction of aesthetic needs is a salient component of housing satisfaction with regard to the inside, the outside, and the neighborhood, and the aesthetic qualities of the living environment are given high importance across all age and social groups. However, empirical research on preferences for interior decoration and furnishing has shown that aesthetic preferences can enter into conflict with other needs, such as comfort, practicality, or upkeep, and that some persons will change their preferences depending on whether they are asked to evaluate a house for living purposes or merely from an aesthetic standpoint.

The explanatory models adopted in environmental psychology are classified into two different traditions: (i) the experimental and motivational aesthetics of Berlyne, who traces aesthetic preferences to the physiological arousal response elicited by the "collative properties" of the objects, which include structural properties (e.g., complexity and coherence) and aspects of experience (e.g., novelty and the ability to surprise), and (ii) the sociopsychological tradition that derives preferences from the social meaning of the objects (i.e., their ability to express social identity in terms of belonging and distinction). Other theories have emphasized the cognitive components of aesthetic preference: For example, according to the schema theory, preferences relate to the discrepancy between the perceived example and the abstract schema, or prototype, representing the appropriate category. Increases in discrepancy evoke arousal and interest, but preference is highest for moderate to low discrepancies. Considerable empirical work has been carried out to study the physical and symbolic

properties of a building and what individual variables affect preferences.

### **1.4. Methods for Assessing Preferences**

In studying preferences, a distinction widely made is that between revealed and expressed preferences: Revealed preferences are inferred from actual behavior, whereas expressed preferences are directly derived from the answers people give to questions about what they like or dislike. Although revealed preferences are used considerably in the sociological and economic domains, in psychology the second technique is preferred. Persons may be asked to provide preferential judgments that record their personal likes or dislikes concerning the places to be assessed (e.g., through adjective checklists and rating scales), or they may be asked to record comparative appraisals (i.e., evaluations made with regard to some implicit or explicit standard of comparison). Preferences can also be surveyed indirectly by measuring residential satisfaction, although attention must be paid to possible discrepancies between the two measures, as previously mentioned.

One useful approach is based on various forms of trade-off games that allow for priorities for competing interests to be identified. For example, answers to the question of how persons would change their home if they had one room more or less provide a good ranking of priorities.

Variety in methods and techniques is also determined by the presentation of the stimulus to be evaluated: verbal descriptions, drawings, photographs, scale or life-size simulations, or direct observation. There is a lack of data concerning the relative validity of different methods and measurement procedures, even though it has been shown that observers' perceptions of room size and crowding vary significantly in three different modes of environmental display (actual room, video tape, and scale model).

Finally, there is a major methodological difference between asking for the evaluation of separate housing attributes and asking for a global evaluation of different housing alternatives. Adoption of one or the other method is of considerable theoretical importance. In the former case, it is assumed that the preference for a particular housing alternative is determined by the sum of the evaluations of its attributes, whereas in the latter case emphasis is placed on the holistic

dimension of the evaluation, hence the preference for a single attribute depends on the other attributes with which it is associated.

## 2. HOUSING ATTRIBUTES AND PREFERENCES

What are the objective qualities that make one residence preferable over another? The question has been dealt with in a variety of ways, by a number of disciplines, and in the theoretical and applied spheres, ranging from market research to public agencies responsible for housing policies. Considered here are selected aspects that best exemplify the psychological approach at three environmental levels: domestic interior, architectural style, and housing location.

### 2.1. Interior Arrangement and Decoration

A house becomes a home through a process of “appropriation,” which largely consists of choosing and changing the functional organization of space, furniture, and decoration—in other words, by personalizing it. A place that can be personalized becomes a symbolic mirror of certain aspects of the personality of the person residing in that place. It has been shown that people can derive hypotheses regarding the social and personal aspects of the inhabitants from photographs of their living rooms, and that observers’ inferences on the personality and lifestyle of the homeowners correspond significantly to their self-concept.

In a survey of all members of 80 families in Chicago, valued and cherished objects in the home were found to have different meanings based on age and gender. Women and older adults tended to value them as symbols of personal relationships and family ties. Younger people and men, on the other hand, were more utilitarian.

Studies carried out in Italy and France to identify the descriptive dimensions of the physical environment that best differentiate interior decorating choices found that in addition to the kind and style of furniture (modern vs old-fashioned, rich vs plain, etc.), the arrangement of the furniture plays an important role. The items characterized by particular complexity and richness (e.g., antique and large or highly decorated furniture) are generally associated with a symmetrical or central position, whereas more sober items are

associated with a more elaborate and unconventional composition. The close correspondence between preferences for interior design and furnishings and cultural background and social status suggest that individual taste is largely shaped by cultural ideologies.

If interior decorating choices largely reflect the sociocultural context, this holds true even more for the internal organization of space, over which individuals can exercise less control. Anthropological studies have shown that division of the domestic space and the functional restrictions by gender, age, or activity are congruent with the social and political organization of a culture. In addition, architectural and organizational characteristics have a precise psychological meaning. For example, by asking participants to construct models of houses that would promote in turn privacy, security, and social interaction for the occupants, researchers found that privacy houses had significantly more rooms and corridors, security houses were smaller, and social interaction houses had greater visibility between rooms.

Environmental psychology has emphasized societal trends related to socialization patterns and household roles that are reflected in the actual configuration of the house (room size and functional organization) and in changes in needs and desires. For example, a study of prototypical house plans in the United States in the second half of the 20th century found that increasing the size of the kitchen and incorporating it more openly into the rest of the house seems to respond to women’s desire for greater sharing of domestic tasks and greater interaction among family members. More informality in social interaction is reflected in the widespread preference for multifunctional spaces, though with considerable variations related to household composition (e.g., the presence and age of children) and life stage.

### 2.2. Architectural Style

Research on aesthetic evaluations of buildings is both of theoretical interest, because it allows for the verification of theories on aesthetic perception in a real context, and of applied interest in relation to urban design policies. Various studies in North America, Australia, the United Kingdom, and Turkey have investigated the preferences for different house styles. As with interiors, at least with regard to single houses, preferences for the exterior were largely influenced by the symbolic meaning transmitted by the architectural form. Affective and social meanings are attributed to the different styles: A particular style may be considered more pleasing or interesting than

another, and inferences can be drawn about the characteristics of the residents.

The architectural traits associated with social status are resistant to change and learned at an early age. By looking at photographs of homes built at the end of the 19th century in Boston, observers were able to identify the occupational status and social class of the original owners; schoolchildren (6–12 years) showed high internal agreement in the cost ranking of several house styles, with preferences being correlated with cost.

In addition to confirming the U-shaped relationship between complexity and liking theorized by Berlyne, of particular interest is the lack of correlation found in many studies, if not all, between local experience and preferences. In evaluating representative examples of Australian and British houses, students in Glasgow and Sydney, for example, liked the ones to which they were less accustomed.

Using a set of stimuli including drawings of six different house styles (contemporary, colonial, farm, Tudor, saltbox, and Mediterranean), ample convergence of preferences of laypersons for the Tudor and farm styles was found in different geographic areas of the United States. Differences were found, using the same and/or other stimuli, between adults and children, men and women, architects or students of architecture, and people without a background in design.

### **2.3. Housing Location**

An important characteristic of the place of residence is the location. Residential location preferences have been approached from two perspectives: (i) the importance attributed to a variety of services and functions (e.g., the proximity of parks, schools, friends, and relatives) that characterize a specific location and (ii) the typology of the setting, particularly in relation to the distance from city center. The two aspects are not independent, but since the evaluation of neighborhood attributes has been studied mainly in relation to neighborhood satisfaction, attention is concentrated here on the second aspect.

A wide literature, especially in North America, has repeatedly noted the strong preferences for small towns or suburban living that accompany the desire for a single-family, self-owned home. If the individuals' age and role in the family are taken into consideration, the picture becomes much more complex. Preferences for one location or another are determined by familiarity and expectations, largely guided by cultural norms,

about the proper type of place for a person to live at a particular stage in life. The psychological ties established with certain kinds of settings have been conceptualized in terms of settlement identity. Settlement identity is formed through personal experience of specific types of residential environments and the generalization of the characteristics to prototypes of places. Thus, city center is associated with greater opportunities for leisure and cultural activities, a wider range of contacts, and greater accessibility to certain services. The suburbs are associated with more quiet, better air quality, greater homogeneity among residents, etc. Small towns seem to be associated with the best characteristics of both the city center and the suburbs. Identification with one kind of setting or another and the consonance or dissonance between settlement identity and actual place of residence influence both preferences and assessment of different settings.

## **3. PERSONAL AND SOCIAL INFLUENCES IN PREFERENCE RATINGS**

As mentioned previously, in addition to environmental characteristics, a wide variety of individual characteristics, ranging from personality and personal history to demographic and sociocultural factors, influence the preferences and meanings attributed to a house. Here, attention is focused on only a few characteristics, keeping in mind that the separation between them is rather artificial and that differences arise mainly through the interaction between different dimensions.

### **3.1. Designers vs Users**

Numerous studies have documented that architects or other design experts assess built environments differently from laypeople and that architects are unable to predict lay evaluations. Regarding housing style, there is a higher consensus in preferences among trained people than among untrained laypersons, with architects preferring more "high style," "atypical" houses. It has been suggested that this is the result of a greater predisposition among architects for higher levels of sensory input and their familiarity with atypical examples acquired during training, which leads to the formation of prototypical schemes that are different from those of untrained persons.

### **3.2. Gender Role**

In all societies, men and women have different roles in relation to the environment and the opportunities offered by it. This is reflected in both the significance attributed to the home and the needs and preferences related to the residential environment at various levels. Although there is a high degree of consensus between men and women regarding the essential qualities of a home, there are differences in the salience of different attributes: For many women, the home symbolizes self, family, and social relationships; for many men, it symbolizes ownership and childhood memories. Women attribute greater value to objects of “contemplation” (photographs, plants, etc.); men attribute greater value to objects symbolizing action, such as sports equipment and stereos. Regarding housing design, modern style is preferred by males, whereas women tend to favor more traditional design.

Women are generally less satisfied with housing than men and desire more privacy. However, this difference in satisfaction with a living situation often does not depend on a difference in assessment criteria but, rather, on the subjective fact that many women, particularly single mothers or older women living on their own, are more poorly housed than men in analogous situations.

Women, particularly single mothers, are more open to alternative housing arrangements, such as cohousing, that allow them to share some domestic tasks. Other differences are found at the neighborhood level: Proximity to friends and accessibility to services are more important for women, who are also more in favor than men of denser communities and multifamily housing.

### **3.3. Age and Stage of Life**

Some of the most important transitions in people’s lives are either defined by a change in where they live (e.g., leaving the parental home) or constitute a reason for moving (e.g., the birth of a child or an older child leaving the home). Furthermore, changes in the significance attributed to the residential environment and the needs it has to satisfy are associated with different stages in life.

Use of the spaces making up the residential environment and their salience at the psychological level vary during the course of a lifetime. For children, interest is centered on play: The home has to be safe and provide space inside and outside for play. As adolescence approaches, privacy and self-expression take on a

preponderant role. Older teenagers and young adults living in their parents’ home show little interest for the characteristics of the house, as long as privacy and control over their private space are ensured. Loss of control seems to characterize places that young adults experience as the most depressing. Families with children are often dissatisfied with the space in the home and want to move to residential neighborhoods far from the city center, whereas many older couples and single residents prefer to live closer to the city center.

Housing for the elderly has increasingly become a key issue in environmental psychology. However, placing an undue emphasis on the concept of the elderly as a homogeneous group is a mistake. First, old age encompasses an arc of life that, with the increase in life expectancy and advances in health care, is becoming elongated and ever more differentiated—to the point that distinctions are now made between young old, old, and old old. Second, interindividual differences tend to accentuate with age in relation to the variety and quality of life experience. Finally, cultural differences are heightened: For example, whereas independent living is a desire shared by the vast majority of the elderly in the United States, in Japan the solution preferred by an elderly widow is to live with her son.

Some features nevertheless seem to characterize the relationship between the house and advancing age: most notably, a strong attachment to the house, seen as a symbol of independence and a material trace of the pleasant and unpleasant events of one’s own life. Numerous studies have shown the negative consequences on health from moving from one house to another and, even more, from a house to an institution; however, studies have also demonstrated the moderating effects of the circumstances in which that change takes place, particularly whether the choice is voluntary, its preparation, the characteristics of the new environment, and the social context before and after. Increasing emphasis is being placed on the proactive aspects of housing choices in old age in an attempt to identify the strategies for optimization activated by a change in environment.

## **4. MOVING HOME AND ATTACHMENT**

Prolonged association between an individual and a place is widely recognized as one of the distinctive features of

attachment to place. As happens with attachment to people, individuals may not be conscious of their attachment to a place and may only become aware of it in particular circumstances, such as when the bond is threatened. Moving means breaking this bond; therefore, relocation has been considered as providing unique insight into the attachment processes.

The early theories on attachment take two opposing approaches. The first approach stresses the emotional experience of places: Affective ties are established automatically and unconsciously with one's place of residence and are destroyed by mobility. The second approach focuses on cognitive rather than emotional aspects of people–environment relationships and stresses the positive function of change. Affective bonds with place are the result of a positive assessment of the congruence between the attributes of the environment and the inhabitant's needs. Because needs are liable to change during one's lifetime, if one's needs are fully satisfied by a new place, attachment can be expected to develop even after a move.

Recent research, while acknowledging the importance of emotional aspects, emphasizes the dynamic relationship between person and place and the modification of affective bonds throughout a person's lifespan. However, still lacking are longitudinal studies on how new bonds are formed, to what extent one place of attachment can be replaced by another, or whether multiple attachments can coexist.

Although leaving a place one loves is almost always stressful and painful, not all people appear to develop feelings of attachment with their place of residence and the reaction to a change is closely linked to a person's time and place orientations (i.e., their perception of themselves as anchored more to a past or to a future timeframe).

Finally, two factors of fundamental importance are the possibility of remaining in contact with the places of the past and the perception of control with respect to the change. Enforced relocations, caused by natural and technological disasters, urban renewal, or transfer into institutions, constitute a break in the individual's and community's sense of continuity that is often underestimated and cannot be compensated by the physical quality of the new environment.

## See Also the Following Articles

- Environmental Design and Planning, Public Participation in
- Environmental Stress ■ Person–Environment Fit
- Residential Satisfaction and Perceived Urban Quality
- Urban Environments and Human Behavior

## Further Reading

- Altman, I., & Low, S. M. (Eds.). (1992). *Place attachment. Human behavior and environment. Advances in theory and research* (Vol. 12). New York: Plenum.
- Altman, I., & Werner, C. M. (Eds.). (1985). *Home environment. Human behavior and environment. Advances in theory and research* (Vol. 8). New York: Plenum.
- Aragonés, J. I., Francescato, G., & Garling, T. (Eds.). (2001). *Residential environments: Choice, satisfaction and behavior*. Westport, CT: Bergin & Garvey.
- Arias, E. G. (Ed.). (1993). *The meaning and use of housing: International perspectives, approaches and their applications*. Aldershot, UK: Avebury.
- Bonnes, M., Giuliani, M. V., & Werner, C. M. (Eds.). (1987). *Home interior: A European perspective* [Special issue]. *Environment and Behavior*, 12(2).
- Cooper Marcus, C. (1995). *House as a mirror of self*. Berkeley, CA: Conary Press.
- Craik, K. H., & Zube, E. H. (Eds.). (1976). *Perceiving environmental quality*. New York: Plenum.
- Csikszentmihalyi, M., & Rochberg-Halton, E. (1981). *The meaning of things. Domestic symbols and the self*. Cambridge, UK: Cambridge University Press.
- Giuliani, M. V. (2003). Theory of attachment and place attachment. In M. Bonnes, T. Lee, & M. Bonaiuto (Eds.), *Psychological theories for environmental issues* (pp. 137–170). Aldershot, UK: Ashgate.
- Lawrence, R. J. (1987). *Housing, dwellings, and homes: Design theory, research and practice*. Chichester, UK: Wiley.
- Lawrence, R. J. (Ed.). (1991). The meaning and use of home [Special issue]. *Journal of Architectural and Planning Research*, 8(2).
- Nasar, J. L. (Ed.). (1992). *Environmental aesthetics: Theory, research, and applications*. Cambridge, UK: Cambridge University Press.
- Rapoport, A. (1990). *The meaning of the built environment: A nonverbal communication approach* (R. ed.). Tucson, AZ: University of Arizona Press.
- Tognoli, J. (1987). Residential environments. In D. Stokols & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 655–690). New York: Wiley.



# Residential Satisfaction and Perceived Urban Quality

Marino Bonaiuto

University of Rome La Sapienza, Rome, Italy

1. Introduction
2. Urban Environment Quality in Experts' and Laypersons' Evaluations
3. Describing Residential Satisfaction and Perceived Residential Environment Quality
4. Explaining Residential Satisfaction
5. Applications  
Further Reading

## GLOSSARY

*(perceived) environmental quality indexes [(P)EQIs]* Standard set of (perceived) indicators for evaluating a specific place (e.g., residential PEQIs) that can also be used for policy and monitoring functions.

*place attachment* Forms and degrees of affect (e.g., feelings, moods, emotions) that people develop over time and come to experience with reference to the places in which they are born, live, and act.

*(residential) postoccupancy evaluation (POE)* An examination of the effectiveness for human users of occupied designed (residential) environments.

*residential satisfaction* The experience of pleasure or gratification deriving from living in a specific place, that is, the global evaluations that inhabitants give of their housing that can be considered at various levels of scale (e.g., house, building, neighborhood).

*residential (urban) quality* Evaluation of the various specific attributes of the residential (urban) environment; this

evaluation can be technical and expert based (i.e., "objective") or observer based (i.e., perceived or "subjective").

The quality of the urban environment, and of the residential environment in particular, is one of the main components of people's overall quality of life. Global satisfaction with people's own residential environment has both affective and cognitive components. On the one hand, people experience feelings with respect to their own houses/neighborhoods (i.e., place attachment); on the other hand, they assess the degree of quality for each different feature of their own houses/neighborhoods (i.e., residential quality). Knowing residential satisfaction, residential attachment, and residential quality with reference to the inhabitants of a specific place can help its design, management, and policy.

## 1. INTRODUCTION

The quality of the urban environment, and of the residential environment in particular, is one of the main components of people's overall quality of life, although it tends to rank lower (in terms of importance in the overall quality of life) than do items such as leisure activities, economic conditions, work, friendships, and marriage/family life.

Residential satisfaction can be defined as the experience of pleasure or gratification deriving from living in a specific place, that is, the global evaluation that inhabitants give of their housing that can be considered at various levels of scale (e.g., house, building, neighborhood). Conceptually, it includes the three main components of the psychological construct of attitude: cognition, affect, and behavior. Operationally, residential satisfaction is often measured through inhabitants' answers (ranging from "not at all" to "completely") to items such as "How satisfied are you with the house/building/neighborhood?" Satisfaction scores lack absolute validity; a high score cannot be automatically considered as an indicator of optimal performance of a housing program. However, satisfaction scores do have relative validity; other things being equal, a setting with a high satisfaction score performs better than one with a low score.

Residential satisfaction can be considered an overall global evaluation of the residential environment from inhabitants' perspectives. Such a construct includes different aspects belonging to three main domains: the realm of people's affective response toward the residential environment, the realm of their cognitive evaluations of the same environment, and the realm of the action carried out in that environment. Very few studies have dealt with the behavioral component, whereas the first two components have been studied in greater depth. The first two aspects are grasped by two important constructs that are strictly related to, and partly overlapping with, residential satisfaction: place attachment and perceived residential environment quality (PREQ), respectively.

Place attachment, or more specifically residential attachment, refers to the feelings that people develop over time and come to experience with respect to their homes or their residential neighborhoods. It encompasses cognitive, motivational, and behavioral aspects, reflected in psychological correlates such as the tendency of people to give favorable evaluations of their dwelling places, reasons for improving them, and reluctance to leave them. Place attachment can be operatively defined in terms of inhabitants' answers to items such as (in the case of the residential neighborhood scale) "This neighborhood is part of me" and "It would be very hard for me to leave this neighborhood." The aim is to grasp people's overall affective responses to their own residential environments.

PREQ refers to people's evaluations of the degree of quality of a range of single features of their homes or neighborhoods. PREQ can be operatively defined in terms of inhabitants' answers to items such as (again

referring to the residential neighborhood scale) "Buildings have unpleasant colors" and "The air is clean." The aim is to grasp specific cognitive evaluations with respect to a range of certain residential features from people's points of view.

Moreover, environmental quality is typically considered to be a multidimensional construct. Residential quality can include an array of features (e.g., from architectural and town planning features to human and social features, from services to contextual elements). Each of these categories can include a range of subcategories, with each consisting of many different items.

Finally, environmental quality can be viewed in terms of both technical and observational assessments. In terms of the former, environmental quality is evaluated on the basis of technical systems of measurement (so-called objective or expert assessment). In terms of the latter, perceived environmental quality is evaluated on the basis of consensual impressions expressed in an ordinary language framework by the inhabitants (so-called subjective or layperson assessment).

## **2. URBAN ENVIRONMENT QUALITY IN EXPERTS' AND LAYERPERSONS' EVALUATIONS**

Thus, the quality of the very same urban environment can be evaluated from at least two different perspectives: the technical expert's assessment and the observer-based layperson's assessment. The former is often referred to as objective because it involves tools and measures (e.g., metrics, weights) of scientific and technical disciplines, that is, mechanical monitoring equipment or other physical means to produce a reading of environmental quality (it does not require individual exposition). The latter perspective may be referred to as subjective because it relies on self-report tools through which people express their own perceptions, observations, and impressions; that is, it employs the perceptual abilities of humans to judge environmental quality. This means that observational assessment offers a measure of the quality of the environment as it is experienced. However, both kinds of environmental assessments aspire to objectivity inasmuch as they strive to have reproducible measures that are valid, reliable, sensible, and useful. Similarly, they both share subjectivity because technical assessment relies on human decisions regarding which dimensions to examine and sampling times and places as well as on

interpreting results, whereas observational assessment by definition is derived from individual experience-based responses (although these responses sometimes can show strikingly high intersubjective agreement).

Both kinds of evaluation aim to evaluate the environment in which people live, but it is likely that different values, ideals, and goals concerning the environment underlie the expert's technical assessment and the layperson's observational assessment. Theoretically, such underlying discrepancies reflect the processes of social and cultural construction of the environment and of environmental issues, where social conflicts and controversies affect environmental descriptions and evaluations. For example, ethical, aesthetic, and ideological forces play a role in defining what should be considered as "polluted" or "dangerous" for the environment and for people.

Therefore, the degree of correspondence between the different types of evaluations (technical and observational) can be problematic. For example, when assessing the quality of urban green areas, experts and laypersons are often in strong disagreement because laypersons place strong importance on green area accessibility and usability, whereas experts' assessments tend to overlook these aspects and emphasize the quality of the biological element per se. Thus, different systems of values can generate different, and sometimes opposite, evaluations of the quality of the same urban residential environment. For this reason, it is important to compare and integrate technical assessments with observational ones in designing and managing the environment.

The following sections discuss the general argument of observational assessment by referring specifically to the residential neighborhood scale.

### **3. DESCRIBING RESIDENTIAL SATISFACTION AND PERCEIVED RESIDENTIAL ENVIRONMENT QUALITY**

A basic issue in observational assessment is to establish criteria and devise valid and reliable tools to operatively describe and measure constructs such as residential satisfaction and perceived quality of the residential environment. The general aim here is to know how laypersons (i.e., residents) perceive and evaluate these aspects. Such a problem can generally be framed within a broader "place theory" that considers any place

experience, at any given scale, as psychologically composed of three main elements: the physical attributes of the setting, the activities carried out in the setting, and the descriptions or conceptions that people have of the setting. Therefore, describing inhabitants' residential satisfaction and perceived residential quality is a prerequisite to obtaining a complete environmental-psychological picture of their residential place (which should, however, be complemented by a technical assessment of the place's physical features as well as by an assessment of inhabitants' uses of the place).

In describing how inhabitants perceive, conceive, and evaluate a residential place, two main methodological approaches can be adopted: a more "inductive" approach, which moves from how inhabitants spontaneously think and/or talk about their own residential environment, and a more "deductive" approach, which moves from the features of the residential environment that the researcher considers to be theoretically relevant and then elicits inhabitants' evaluations of these features. Often, the two approaches represent two successive phases of a broader research project.

On the basis of the first approach, inhabitants can take part in unstructured or semistructured interviews, or in small group discussions, and the researcher can then try to describe the way in which they talk about their residential environment. This approach usually privileges qualitative evidence to depict the common-sense ways in which to conceive and evaluate the residential environment.

On the basis of the second approach, inhabitants are typically administered questionnaires or checklists that present them with several items, referring to their residential environment, selected as relevant within the researcher's theoretical framework. Then, they are requested to respond to the items by filling in agreement-disagreement or satisfaction-dissatisfaction graduated scales. This approach typically privileges quantitative methods, gathering large samples of inhabitants and using statistics to infer the population's main trends.

The first approach can offer an in-depth and fresh description of the "residential ethnography," that is, inhabitants' ways of giving meaning to, and organizing, their social-physical residential reality. The second approach can offer quick, systematic, and broad information on how inhabitants position themselves on the whole array of features that the researcher considers to be crucial to evaluate the PREQ.

Each method can emphasize different aspects. For instance, consider perceived residential neighborhood

quality. If one concentrates on inhabitants' free expressions, one will probably notice that they often bring together the evaluation of aspects that are technically or theoretically considered separate (e.g., architectonic or urban planning features with social features, services with social features). On the other hand, if one asks inhabitants directly, through ad hoc prepared scales, to evaluate each feature of the residential environment, one will end up with separate pure indicators for each feature. In the first case, one can appreciate the fact that the method allows one to "reconstruct" the web of inhabitants' residential perceptions and evaluations, discovering how they psychologically bring together different features on the basis that, for example, they all pertain to urban safety. In the second case, one can appreciate the possibility of knowing inhabitants' positions on exactly the same matters in which one is interested, and these can then be used to make systematic comparisons between different groups or places or between these perceptions and other expert-based assessments of the very same features.

Having said this, it must be recognized that the results of different studies carried out in several countries over the past 30 years or so converge to show a stable structure underlying people's evaluations of their residential environment. Turning again to the residential neighborhood scale, so far as neighborhood satisfaction is concerned, its description and measurement has traditionally been pursued in a very simple way: through a single item. On the other hand, so far as neighborhood environment quality is concerned, consistent evidence in the literature confirms that a theoretical model that includes three or four main areas can account for inhabitants' perceived neighborhood quality. Specifically, the first three areas are traditionally established in the literature, whereas the fourth area has been proposed in more recent contributions:

1. Spatial features (architectonic and town planning, including green areas)
2. Human and social features (population and type of social relations)
3. Functional features (available services)
4. Contextual features (lifestyle, pollution, maintenance, and care)

Each of these macro-areas consists of several aspects, and measurement scales are available to describe inhabitants' perceived quality of each specific aspect. The following is an example of a list of PREQ indicators for which specific valid and reliable

scales exist (all referring to the neighborhood or the neighbors):

1. Building aesthetics
2. Building density
3. Building volume
4. Internal practicability
5. External connections (with the city)
6. Green areas (presence and care)
7. Discretion and civility
8. Security and tolerance
9. Sociability and cordiality
10. School services
11. Social care services
12. Sport services
13. Social-cultural activities
14. Commercial services
15. Transport services
16. Relaxing versus distressing pace of life
17. Stimulating versus boring pace of life
18. Environmental health (unpolluted environment)
19. Upkeep and care

#### **4. EXPLAINING RESIDENTIAL SATISFACTION**

In general, two main models have been proposed in the literature to explain residential satisfaction. A first model considers residential satisfaction as a criterion variable and seeks to identify its most relevant predictors so as to provide useful information for town planning and management. A second model uses residential satisfaction as a variable predicting other relevant urban behaviors such as residential mobility and residential choice.

Residential satisfaction can depend on different kinds of factors: personal, social, and physical. Among personal factors, residential satisfaction tends to increase with age, socioeconomic level, and improvement expectations. It tends to decrease when people are characterized by an aggressive personality trait or when people compare their residences with high-standard solutions.

Among social factors, residential satisfaction tends to increase with good neighbor relations, with similarity to neighbors, with a balance of separation and togetherness with neighbors that suits them, and with adherence to and conformity with social norms regarding "usual" housing arrangements for people like themselves.

Among physical factors, single-family dwellings are more satisfying than apartments (at least in North America), and green areas determine higher residential satisfaction. The preferred architectural style, as well as the interior, tends to vary with the cultural background of inhabitants. Color has complex relations with residential satisfaction as well, depending not only on hue but also on saturation and brightness. In general, residential satisfaction is greater when residences match those of inhabitants' own culture, probably because residences that reflect the culture are designed to complement behavior patterns typical of that culture.

Moreover, all of these factors can affect not only the level of residential satisfaction but also very specific preferences for residential solutions and choices at different scale levels.

In sum, within a general organizing framework or model, all of the previously mentioned factors can affect inhabitants' impressions of their residential places, and these in turn affect their cognitive, affective, and behavioral responses to their residences. These responses can also be affected (directly or indirectly) through the inhabitants' impressions by the objectively measurable features of their residences.

Again considering the example of inhabitants' cognition and affect at the residential neighborhood scale, it has been shown that neighborhood satisfaction and neighborhood attachment are related to a variety of neighborhood features, specifically to personal, social, and residential factors as well as to certain aspects of PREQ. So far as personal or residential features are concerned, age, home ownership and size, and length of residence in the neighborhood, for example, are usually positively related to neighborhood satisfaction and/or attachment, whereas home crowding is usually inversely related to it.

Among the features of PREQ, some are generally positively associated with residential satisfaction and attachment: general social and psychological climate (opportunities and quiet), building aesthetics and presence and maintenance of green areas, social relations (sociability, discretion, safety, and tolerance), and presence of, and access to, services and facilities. In the case of attachment, compared with residential satisfaction, the quality of social relations bears a stronger importance. However, both residential satisfaction and attachment are associated to a "cocktail" of quality indicators belonging to different areas: architecture and town planning features, social relations, services, and general context. That is, they both share a multi-component or multidetermined nature.

## 5. APPLICATIONS

Residential satisfaction, residential attachment, and PREQ indicators can be used with applied aims in specific settings for residential design or evaluation. Typically, important applied research issues pertain to (a) the relationship between experts' and laypersons' environmental assessments, (b) the relationship between different evaluations (e.g., cognitive and affective, molecular and molar), and (c) the use of these constructs and measures in postoccupancy evaluation (POE) procedures.

The first two issues were mentioned in the opening paragraphs. So far as (residential) POE is concerned, it can be defined as an examination of the effectiveness for human users of occupied designed (residential) environments. It should be a systematic process, including all user groups, all important activities occurring in a place, and rigorous methods of collecting and analyzing data. It can refer to any relevant constructs and instruments, but when residential environments are being considered, it would certainly assess both residential satisfaction and some PREQ indicators and probably residential attachment as well. It would try to understand, for example, what is the degree of residential satisfaction and what are the environmental quality indicators—or other personal, social, residential, and physical factors—most associated with it for the inhabitants of the considered place.

In sum, knowing how residential environment quality (both perceived and technical) relates to residential satisfaction and residential attachment, with reference to the inhabitants of a specific place, can help to inform its design, management, and policies.

### See Also the Following Articles

Environmental Design and Planning, Public Participation in ■ Environmental Stress ■ Extreme Environments and Mental Function ■ Person–Environment Fit ■ Restorative Environments ■ School Environments ■ Urban Environments and Human Behavior ■ Work Environments

### Further Reading

- Aragónes, J. I., Francescato, G., & Garling, T. (Eds.). (2001). *Residential Environments: Choice, Satisfaction, and Behavior*. Westport, CT: Bergin & Garvey.
- Bonaiuto, M., Aiello, A., Perugini, M., Bonnes, M., & Ercolani, A. P. (1999). Multidimensional perception of residential environment quality and neighborhood

- attachment in the urban environment. *Journal of Environmental Psychology*, 19, 331–352.
- Bonaiuto, M., Bonnes, M., & Continisio, M. (2004). Neighborhood evaluation within a multi-place perspective on urban activities. *Environment & Behavior*, 36, 41–69.
- Bonaiuto, M., Fornara, F., & Bonnes, M. (2003). Indexes of perceived residential environmental quality and neighborhood attachment in urban environments: A confirmation study on the city of Rome. *Landscape and Urban Planning*, 65, 41–52.
- Bonnes, M., & Bonaiuto, M. (1995). Expert and layperson evaluation of urban environmental quality: The “natural” versus the “built” environment. In Y. Guerrier, N. Alexander, J. Chase, & M. O'Brien (Eds.), *Values and the environment: A social science perspective* (pp. 151–163). New York: John Wiley.
- Craik, K. H., & Feimer, N. R. (1987). Environmental assessment. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (Vol. 2, pp. 891–918). New York: John Wiley.
- Gifford, R. (2002). *Environmental psychology: Principles and practice* (3rd ed.). Colville, WA: Optimal Books.
- Lawrence, R. J. (2002). Healthy residential environments. In R. B. Bechtel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 394–412). New York: John Wiley.
- Tognoli, J. (1987). Residential environments. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (Vol.1, pp. 655–690). New York: John Wiley.



# Restorative Environments

Terry Hartig

Uppsala University, Gävle, Sweden

1. Introduction
2. Essential Components of Theories about Restorative Environments
3. Extant Theories about Restorative Environments
4. Conclusions and Prospects  
Further Reading

## GLOSSARY

**restoration** The process of renewing, recovering, or reestablishing physical, psychological, and social resources or capabilities diminished in ongoing efforts to meet adaptive demands.

**restorative environment** An environment that promotes (and not merely permits) restoration.

**stress** A process of responding to an excess of demands relative to the resources needed to cope with those demands.

People inevitably deplete adaptive resources in everyday life. Persistent failure to restore needed resources will ultimately harm psychological and physical health. Restoration proceeds more effectively in some environments than in others due to a relative absence of demands as well as to qualities that promote restoration. Knowledge of the restoration-promoting qualities of environments can serve applied health fields.

## 1. INTRODUCTION

People inevitably deplete physical and psychological resources as they meet self- and externally imposed demands within changing environments. Failure to reestablish vital capabilities for effective action can harm health through multiple pathways. Some of those pathways have received close attention in research on the effects of chronic stress. As commonly defined, stress arises from an excess of demands relative to the resources needed to cope with those demands. This formulation implies that stress becomes chronic when excessive demands persist and the person can neither acquire the new resources needed to obviate those demands nor apply available resources more effectively. Less obviously, stress may also persist when the person cannot access an environment that supports sufficiently rapid or complete restoration of necessary resources diminished in the effort to cope.

In searching for the sources of chronic stress, researchers have identified a variety of social and physical environmental stressors such as crowding and noise. A relative absence of such demands may permit restoration. However, restorative environments warrant definition in positive terms rather than in negative terms. Environmental psychologists, in extending research on topics such as landscape aesthetics and psychological benefits of wilderness experience, have proposed qualities of environments that promote restoration. This theorizing has value for fields concerned with psychosocially

mediated relations between environment and health, such as health psychology, public health, social epidemiology, and psychosomatic and preventive medicine.

## 2. ESSENTIAL COMPONENTS OF THEORIES ABOUT RESTORATIVE ENVIRONMENTS

To better understand the current status and avenues for further development of theories about restorative environments, consider what such theories must include. First, a theory must specify some condition from which a person becomes restored. Second, it must describe the process of restoring some set of resources. Finally, it must characterize the environments that promote a process of restoration.

### 2.1. Antecedent Condition

Restoration cannot occur unless the possibility for restoration exists. Necessarily, before a person can become restored, he or she must have depleted some of those resources that are useful for maintaining and improving adaptation to the environment. Whether those resources are biological, psychological, or social, their availability is essential to continued adaptation. Over the long run, an inability to renew depleted resources may have grave consequences for effective action, subjective well-being, and physical health.

That the resources of interest here regularly and predictably become diminished helps one to distinguish the driving concerns of restorative environments theory from those of therapies and rehabilitative strategies. True, restorative environments applications can serve goals of therapy and rehabilitation defined by one health professional or another. Yet therapy often has to do with capacities that a person never had and, thus, that could become “restored” only with reference to normative criteria based on some population. Alternatively, therapy and rehabilitation may focus on capacities that a person has lost to an accident or some pathological process rather than in the normal course of adaptation to a continuously changing environment. This means that restorative environments theory has a broader reach than do therapy and rehabilitation in general. It does not necessarily exclude relatively unusual events that may reduce some capacity for action, but its area of concern extends to states that reflect normal “wear and tear.”

### 2.2. Restorative Process

The term “restoration” denotes some set of processes through which one or more individuals renew or reestablish adaptive resources or capacities that have become diminished. Whether the resources of interest have a biological, psychological, or social character, the processes included under the restoration rubric here have in common a psychosocial character. Having specified an antecedent condition, a theory ought to define the given restorative process in terms of the resources that become restored and the psychological, physiological, and/or social mechanism(s) for their renewal. Furthermore, because all processes extend through time, a complete description of a restorative process will refer to temporal features and parameters such as stages in restoration and the time required for restoration. Measurements of appropriate variables at various time points provide evidence as to whether or not restoration is actually under way.

### 2.3. Environmental Context

All restoration occurs in the course of some activity, and all activity occurs in some environment. Restorative processes are not necessarily specific to a particular environment, but they may proceed more readily or smoothly in some activities and environments than in others. Those environments that promote (rather than merely permit) restoration can be referred to as “restorative.” Whether with respect to their physical, social, activity, temporal, or spatial features, some form of contrast with another relatively demanding environment is implicit in the description of an environment as restorative.

#### 2.3.1. Significance of the Natural–Urban Distinction

One set of contrasts aligns with a coarse distinction between natural and urban environments. In studying restorative environments, some researchers have focused on people living in cities who occasionally go to relatively natural environments for restoration. Convergent practical concerns motivate this focus. First, most people today reside in urban areas. Conditions common in urban areas can impose heavy demands on people. Second, it might not be easy to ensure that people who live in cities have easy access to natural environments. Other people may prefer to close off access to those environments and disturb their natural character in the pursuit of economic gain.

Although these problems are related, work on them within environmental psychology has, for the most part, proceeded independently. A large body of research attests to harmful effects of noise, crowding, air pollution, and other demands in urban areas. Another large body of research describes psychological and social amenity values of natural areas placed at risk by urbanization and the extraction of timber, minerals, and so forth. In restorative environments research, the practical and theoretical connections between these bodies of work become distinct.

### 3. EXTANT THEORIES ABOUT RESTORATIVE ENVIRONMENTS

Two theories have guided most of the research on restorative environments to date. Both are rooted mainly in research on the psychological values of natural environments, but in specifying an antecedent condition from which people might need restoration, both have referred to work performed under the general “environmental stress” rubric.

#### 3.1. Directed Attention Restoration

Building on some 30 years of research on environmental cognition, environmental preferences, and the psychological benefits of nature experiences, Stephen and Rachel Kaplan developed a theory concerned with the capacity for directing attention.

##### 3.1.1. Antecedent Condition

The Kaplans assumed that a person’s ability to direct attention depends on a central inhibitory capacity. To focus on something that is not of itself interesting, the person must inhibit competing stimuli that are more interesting. Doing so requires effort, and with prolonged or intensive use, the person’s ability to inhibit competing stimuli will diminish or become fatigued. The Kaplans described a variety of negative consequences that the person may suffer when this inhibitory capacity diminishes, including irritability, failure to recognize interpersonal cues, reduced self-control, and increased error in performance of tasks requiring directed attention. Their account of directed attention fatigue has much in common with accounts of cognitive effects and aftereffects of stress or informational overload.

##### 3.1.2. Restorative Process

According to attention restoration theory, a person can restore a diminished capacity for voluntarily directing attention when he or she experiences fascination, a mode of attention that the Kaplans assumed to have an involuntary quality, to be effortless, and to not have capacity limitations. When a person can rely on fascination in ongoing activity, demands on the central inhibitory capacity are relaxed and a capacity for directing attention can be renewed.

Regarding the time course of attentional restoration, the Kaplans proposed that a restorative process may continue through several stages, ranging from clearing one’s head of random thoughts and cognitive “clutter,” to renewing directed attention capacity, and ultimately to reflecting on matters of personal importance.

##### 3.1.3. Environmental Context

As described by the Kaplans, fascination is engaged by objects or events or by processes of exploring and making sense of an environment. Yet fascination is not sufficient for restoration. Attention restoration theory also specifies being away, or getting psychological distance, from the work one usually does and from the pursuit of given goals and purposes—hence, from further demands on directed attention. Another factor is extent, that is, the sense that a physical or conceptual environment available for restoration is sufficiently large in scope to entertain continued exploration and is sufficiently coherent so that one can make sense of what is seen going on around himself or herself and relate it to some larger frame of reference. A fourth factor, compatibility, encompasses a person’s inclinations, environmental supports for his or her activities, and environmental demands. It rests on the match among what a person wants to do, what the person can do, and what the person must do in the given environment. The Kaplans proposed that high compatibility allows for deeper levels of restoration.

Although many environments might afford the experience of being away, fascination, extent, and compatibility, the Kaplans argued that natural environments should do so more readily than do other environments. For example, natural environments may afford being away more readily due to a scarcity of reminders about work demands and a relative absence of people (with whom interactions may require directed attention). The Kaplans also asserted that natural environments are rich in aesthetically pleasing features, such as

scenery and sunsets, which evoke moderate or “soft” fascination that permits a more reflective mode.

### **3.1.4. Empirical Tests**

Several quasi- and true experiments have tested the proposition that experiences of natural environments should promote better restoration of a capacity to direct attention than should experiences of other environments. In these studies, the researchers have operationalized directed attention capacity in terms of performance on tasks that require a participant to focus attention. For example, Hartig and colleagues reported a field experiment in which differential proofreading performance was seen after 40 minutes spent in either a nature reserve, a city center, or a passive relaxation condition. On average, posttest proofreading performance of participants randomly assigned to the natural environment condition was better than that of participants assigned to the other two groups.

## **3.2. Psychophysiological Stress Recovery**

A second theory, developed by Roger Ulrich, focuses on patterns of affective and aesthetic response to visual stimulus characteristics of an environment. Of the range of possible reactions to what one sees in an environment at a given moment, Ulrich had a particular interest in those that promote psychophysiological stress reduction.

### **3.2.1. Antecedent Condition**

Ulrich defined stress as a process of responding to a situation perceived as demanding or threatening to well-being. He further assumed the operation of an evolved system for directing behavior in situations relevant to continued survival. That system depends on “hard-wired” affective responding in the selection of a behavioral strategy (i.e., approach or avoidance) and the simultaneous mobilization of physiological resources needed to execute that strategy. Stress becomes manifest in self-reports of negative emotions and short-term changes in physiological systems that indicate negative affect and heightened autonomic arousal.

### **3.2.2. Restorative Process**

This theory focuses primarily on restoration from stress as a potential mode of affective responding to a visual

stimulus array. It proposes that restoration can occur when a scene elicits feelings of mild to moderate interest, pleasantness, and calm. For a person who enters a situation experiencing psychophysiological stress and needing to renew resources for further activity, it could be adaptive to continue viewing the scene in a nonvigilant manner. While the scene is being viewed, negative affects are replaced by positive ones, the person’s interest is held, and physiological arousal declines.

### **3.2.3. Environmental Context**

It is believed that such restorative responses are initiated very rapidly by the perception of certain visual patterns in the environment. Ulrich described these as aspects of the visual stimulus array that might not provide enough information for conscious cognitive judgments but that, nonetheless, can effectively elicit a generalized affective response. He proposed that qualities of the visual stimulus array, such as moderate depth, moderate complexity, and the presence of a focal point, will promote restoration. Nature enters into this theory as well in that particular environmental contents—vegetation and water—may rapidly evoke positive affective responses. People may be biologically prepared to quickly acquire and retain a liking for environmental features that would have been significant for the survival of our early ancestors. These include water, uniform grassy ground cover, and the forms and distribution of trees characteristic of savannah landscapes—a setting of human evolution.

### **3.2.4. Empirical Tests**

Experiments guided by this theory have involved 10- to 18-minute photographic simulations of natural and urban environments and have documented differential change in emotional and physiological outcomes measured during or immediately after the period of the simulation. For example, Ulrich and colleagues reported a study in which participants viewed a stressful industrial accident film followed by a 10-minute video of either a natural setting, urban traffic, or an outdoor pedestrian mall. Poststressor recovery trajectories for frontalis muscle tension, skin conductance, heart period, and pulse transit time differed as a function of the environment viewed, with nature simulations promoting the fastest returns toward baseline and the lowest overall levels. Changes in self-reported affect converged with the physiological results in showing greater restorativeness with the nature videos.

## 4. CONCLUSIONS AND PROSPECTS

The theories just reviewed propose a variety of restorative qualities of environments, including complexity, depth, and focality in the visual stimulus array as well as being away, fascination, extent, and compatibility in the experience of the purposive actor. Both theories emphasize the relative restorative power of some natural environments, assuming that people remain adapted, to some degree, to the environments of our prehistoric ancestors and observing that those evolutionary settings would contrast sharply with today's demanding urban environments. Research guided by these theories has found, with considerable consistency, relatively greater restorativeness of commonplace, nonthreatening natural environments as compared with urban outdoor environments.

### 4.1. Prospects for Theory and Empirical Research

One perspective on the future of restorative environments research encompasses the possibilities for working with the two theories just overviewed. First, their claims about the mechanisms at work in restoration may well need refinement or revision in the light of independent developments in other areas of psychological inquiry. Second, tests of each theory can more closely consider attributions about the restorative qualities they propose. For example, to what extent does fascination versus being away account for a possible difference in attentional effects of sampled urban and natural environments? Third, studies can test predictions about how a given restorative process leads to some novel outcome, as was done in Kuo and Sullivan's study of directed attention fatigue and aggression by residents of urban low-income housing with barren versus green surroundings. Fourth, the moderating role of individual differences in attention restoration and stress recovery remains little studied by environmental psychologists. Just as individuals consistently distinguish themselves in the face of demands, as with the heightened physiological reactivity of a repressive coping style, so too may individuals show distinctive patterns of restoration. To take a final example, future work can further explore the complementarity of the two theories. The theories appear to deal with processes that can run simultaneously, albeit not for the same duration. They specify antecedent conditions that might not necessarily be related to one another; that is,

a reduced capacity for directing attention may occur independently of the elevated physiological arousal characteristic of stress and vice versa. And although both of the theories note how particular environmental configurations hold attention, they then diverge in following the implications of this for different resources, namely, directed attention capacity versus physiological response capabilities.

The prospects for restorative environments research involve more than further work with the extant theories. The general theoretical framework provided at the outset opens a broader perspective. It implies possibilities for theoretical extension through reference to other antecedent conditions, restorative processes, and environmental variables. For example, the theories reviewed previously have rather little to say about how social factors might work in promoting (vs permitting) restoration. Noting this gap, Staats and Hartig discussed how the presence of another person can enable and enhance restoration in a given environment. Another person can enable restoration by making it possible for an individual to enter an environment of high restorative quality, for example, by providing protection. Issues of access aside, once in a given environment, another person might enhance the restorative quality of the environment by providing the individual with welcome observations and interpretations that increase interest in the environment.

To take another example, consider the cumulative effect of repeated restorative experiences on a specific person's relationship to a given place. Recognizing that people do commonly establish such relationships, Korpela and colleagues proposed that a person becomes attached to places relied on for restorative experiences and that such attachments, in turn, contribute to that person's identity. Furthermore, a place may become more restorative for that person due to the personal bonds. Thus, restorative experience, place attachment, and place identity become reciprocally influential.

### 4.2. Practical Prospects

Beliefs about restorative qualities of environments have appeared over millennia and across cultures. Some such beliefs, expressed in paradisiacal religious imagery, arguments for wilderness preservation, claims about the values of national and urban parks, and philosophies of institutional health care, have attributed restorative quality to natural scenery and features such as vegetation and water.

Recent practical efforts have sought to capitalize on the growing body of empirical research on the restorative values of nonthreatening natural settings and scenery in combination with improved understanding of the biological and behavioral pathways through which chronic stress harms health. At the scale of buildings and their immediate surroundings, architects, interior designers, and landscape architects have used “healing gardens,” landscape paintings and photographs, and window views of natural elements to make highly technological, (often) anxiety-arousing health care settings more humane. Design measures sensitive to user needs for restoration may translate into improved patient outcomes such as less pain and shorter stays. They may also yield benefits to worried family members and hard-working staff.

At the scale of neighborhoods and cities, local decision makers weighing the value of competing land uses can consider increasingly well-grounded claims that green areas serve health. A well-articulated urban green structure can provide urban residents with easily accessible opportunities for exercising and for interrupting the stress process during the course of everyday life in the city. That said, visual access to natural scenery through windows in residences and workplaces may provide restorative benefits independent of physical activity. The increment paid for a pleasant view from a dwelling or hotel room speaks to the value that people attach to such benefits.

People value views with other than natural content, to be sure, and not all restorative environments are natural environments. By considering the potential restorative value of human environments in general, and translating the knowledge secured into policy, planning, and design measures, one can better support adaptation and, thus, promote health and well-being.

### Acknowledgments

The author thanks Florian Kaiser, Kalevi Korpela, and Henk Staats for comments on an earlier version of this article.

### See Also the Following Articles

Conservation Behavior ■ Environmental Assessment ■ Environmental Design and Planning, Public Participation in ■ Environmental Psychology, Overview ■ Pro-environmental Attitudes and Behavioral Change ■ Residential Satisfaction and Perceived Urban Quality ■ Urban Environments and Human Behavior

### Further Reading

- Cimprich, B., & Ronis, D. L. (2003). An environmental intervention to restore attention in women with newly diagnosed breast cancer. *Cancer Nursing*, 26, 284–292.
- Hartig, T. (2001). Guest editor's introduction [to special issue on restorative environments]. *Environment and Behavior*, 33, 475–479.
- Hartig, T., Evans, G. W., Jamner, L. D., Davis, D. S., & Gärling, T. (2003). Tracking restoration in natural and urban field settings. *Journal of Environmental Psychology*, 23, 109–123.
- Hartig, T., Mang, M., & Evans, G. W. (1991). Restorative effects of natural environment experiences. *Environment and Behavior*, 23, 3–26.
- Hartig, T., & Staats, H. (2003). Guest editors' introduction: Restorative environments. *Journal of Environmental Psychology*, 23, 103–107.
- Hartig, T., Johansson, G., & Kylin, C. (2003). Residence in the social ecology of stress and restoration. *Journal of Social Issues*, 59, 611–636.
- Herzog, T. R., Black, A. M., Fountaine, K. A., & Knotts, D. J. (1997). Reflection and attentional recovery as two distinctive benefits of restorative environments. *Journal of Environmental Psychology*, 17, 165–170.
- Kaplan, R. (2001). The nature of the view from home: Psychological benefits. *Environment and Behavior*, 33, 507–542.
- Kaplan, R., & Kaplan, S. (1989). *The experience of nature: A psychological perspective*. New York: Cambridge University Press.
- Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, 15, 169–182.
- Kaplan, S., & Talbot, J. F. (1983). Psychological benefits of a wilderness experience. In I. Altman, & J. F. Wohlwill (Eds.), *Human behavior and environment: Advances in theory and research*, Vol. 6: *Behavior and the natural environment* (pp. 163–203). New York: Plenum.
- Knopf, R. C. (1987). Human behavior, cognition, and affect in the natural environment. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (Vol. 1. pp. 783–825). New York: John Wiley.
- Korpela, K. M., Hartig, T., Kaiser, F. G., & Fuhrer, U. (2001). Restorative experience and self-regulation in favorite places. *Environment and Behavior*, 33, 572–589.
- Kuo, F. E., & Sullivan, W. C. (2001). Aggression and violence in the inner city: Impacts of environment via mental fatigue. *Environment and Behavior*, 33, 543–571.
- Staats, H., & Hartig, T. (2004). Alone or with a friend: A social context for psychological restoration and environmental preferences. *Journal of Environmental Psychology*, 24, 199–211.
- Ulrich, R. S. (1983). Aesthetic and affective response to natural environment. In I. Altman, & J. F. Wohlwill (Eds.), *Human behavior and environment: Advances in theory and*

- research, Vol. 6: *Behavior and the natural environment* (pp. 85–125). New York: Plenum.
- Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science*, *224*, 420–421.
- Ulrich, R. S. (1993). Biophilia, biophobia, and natural landscapes. In S. Kellert, & E. O. Wilson (Eds.), *The biophilia hypothesis* (pp. 73–137). Washington, DC: Island Press.
- Ulrich, R. S., Simons, R., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, *11*, 201–230.



# Risk-Taking and Sexual Behavior

Vita C. Rabinowitz

Hunter College, New York, New York, USA

1. Defining Sexual Risk-Taking
2. What Are Risky Sexual Behaviors?
3. What Groups Are at High Risk for Teen Pregnancy and STIs?
4. What Social, Interpersonal, and Personality Factors Are Associated with Risky Sexual Behaviors?
5. Why Do People Engage in Risky Sex?
6. How Can We Reduce the Incidence of Risky Sex?  
Further Reading

## GLOSSARY

*acquired immunodeficiency virus (AIDS)* A condition caused by the human immunodeficiency virus and characterized by destruction of the immune system, which leaves the infected person vulnerable to a host of serious opportunistic diseases.

*human immunodeficiency virus (HIV)* The virus that causes AIDS.

*human papilloma virus* The virus that causes warts on the anus or genitals.

*risky or unsafe sexual behaviors* Behaviors that make it more likely that one will contract a sexually transmitted infection or that will contribute to an unwanted pregnancy, especially for adolescents.

*sexually transmitted infection (STI)* Any of a number of parasitic, bacterial, or viral infections that can spread through sexual contact.

How do you choose your sexual partner(s)? When do you initiate, accept, and reject sexual advances? When

should you and your partner become sexually intimate, and what kinds of sexual practices will you engage in? What factors into your decisions about condom use? When should you get tested for HIV infection, and when should your partner be tested? What sexual risks do you take, and when and why do you take them?

Like eating and sleeping, sex is a natural function, but alone among the natural functions, sex is widely regarded as risky to people's moral, physical, and psychological well-being. Of the natural functions, it is the one most often feared, proscribed, regulated, and punished. Indeed, many of us learn that sex is dangerous before we learn that sex is natural, pleasurable, or beautiful. This article discusses what psychologists have learned about sexual risk-taking. First, sexual risk-taking and its possible negative consequences are defined. Then, the factors that make sex more or less risky are identified. This is followed by a discussion of some groups who are at special risk for negative sexual outcomes and the contexts or conditions that increase sexual risk-taking. Finally, a discussion of what psychological research can teach us about how to make more responsible sexual decisions is presented.

## 1. DEFINING SEXUAL RISK-TAKING

Although sex is a natural function, motives for sexuality, expressions of sexuality, and the consequences of sexual behavior vary widely within and among individuals, societies, and cultures, and change over time. All societies have an incest taboo, but masturbation,

prostitution, premarital sex, and polygamy, among other sexual activities, are condemned in some societies, tolerated in others, and championed in still others. In 1989, evolutionary psychologist Buss conducted a major survey of more than 10,000 participants from 37 cultures on six continents on the characteristics that people sought in mates. He found that the value that people place on virginity in prospective mates ranged from "indispensable" in China to "irrelevant" in Sweden and the Netherlands. Within societies, the meaning of a sexual act often depends on the context in which it takes place. For example, when access to women is severely restricted, as in male prisons, men having sex with men is not considered deviant and certainly does not cause these men to identify themselves as gay. Just as the meaning of a sexual act depends on many factors, what constitutes a sexual risk also depends on many factors and varies with societal and interpersonal contexts. After the sexual revolution of the 1960s and 1970s, the threat of AIDS in the 1980s has brought with it a return to more conservative, less risky forms of sexual behavior among many groups.

Sex can be risky because it threatens people's social status, emotional or psychological well-being, or physical health. In some societies, sex outside of marriage can threaten a woman's social standing as a wife and mother (if not also her life). Sexually intimate relationships are often consuming and complex, and they offer extraordinarily rich opportunities for feelings of disappointment, frustration, heartbreak, and loss. However, the main focus of psychological research on sexual risk-taking has been on unintended pregnancies and sexually transmitted infections, or STIs. For most psychologists who study human sexuality, risky sex refers to those sexual behaviors that render people vulnerable to threats to their physical health or well-being or that render women and girls, especially teenage girls, vulnerable to unintended pregnancies.

Psychologists' current emphasis on sexually transmitted infections and teen pregnancies in part reflects societies' heightened concerns about these conditions, especially since the latter part of the 20th century. Nothing has provided more impetus for the study of risky sex than the epidemic of AIDS, which is associated with sexual activity and threatens the lives of significant portions of humanity. Because it is sexually transmitted, incurable, lethal, and rapidly spreading, AIDS has garnered more attention worldwide than any new disease in recent history.

Teen pregnancy has long been regarded as a problem for society because teens, especially unmarried teens, are seen as incapable of providing children with the care and resources that they need in today's world. Although the rates of pregnancy, births, and abortions among girls aged 15–19 began to decline in the industrialized world, including the United States, in the early to mid-1990s, teen pregnancy continues to be a very real public health problem. Its rate in the United States remains much higher than that in most other developed countries. The consequences of unplanned teen pregnancies, and 9 in 10 pregnancies among unmarried teens are unplanned, can devastate young mothers, fathers, their children, and society at large. Unmarried teenage mothers are more likely to quit school, live in poverty, and go on public assistance than their peers. As Fraser *et al.*'s 1995 research suggests, they are more likely to have medical complications during pregnancy and childbearing than any group of fertile women except those in their late forties. They suffer high rates of miscarriage, stillbirths, birth complications, and infant mortality. The children of teenage mothers are at greater risk of physical, emotional, behavioral, and intellectual problems. They are more likely than other children to experience poor nutrition and health care; to suffer from family instability, abuse, and neglect; and to perform poorly in school.

Aside from the problems of teen pregnancy, childbearing, and parenting per se, adolescent sexuality raises other concerns. Studies show that approximately 30% of adolescent women do not use any form of contraception during their first sexual intercourse, and fewer than half of sexually active girls use condoms every time they have vaginal intercourse. Because sexually active teenagers use condoms inconsistently, if at all, teens are exposed to a wide range of STIs, including AIDS. According to the Centers for Disease Control and Prevention (CDC), in 2000, young people were among the fastest growing groups of those infected with HIV. Twenty percent of people in the United States who are diagnosed with AIDS are in their twenties, and most of them were infected in their teens. Among teens and young adults, African American youth account for two-thirds of all cases of HIV/AIDS in people younger than age 25. Research reviewed by Rathus *et al.* suggests that adolescent girls are particularly vulnerable to STIs such as HIV/AIDS because the cervix at their age is not fully mature and is thus more susceptible to infection.

Sexual behavior among adolescents has increased more or less steadily throughout the past century. Today, approximately 60% of high school seniors, both girls and boys, have had sexual intercourse at least once, and girls and boys are having intercourse at increasingly younger ages. Indeed, many scientists and laypeople regard all sex among young teens as inherently risky, no matter what forms of contraception and STI prevention are used. Many young people are unprepared for the complex consequences and responsibilities that come with sexual relationships and suffer negative consequences for sexual expression. As Rathus *et al.* note, there is evidence that girls who have sex early, particularly those who become pregnant and have children, exhibit lower self-esteem and are more likely to be depressed than their peers. Girls are often pressured into sexual activity by boys or their female peers, yet, at the same time, condemned for being sexual, especially outside of a committed relationship. This makes it difficult to understand how sexual expression for many girls is a freely chosen, empowered act. In fact, girls who become sexually active early often do not do so voluntarily. Younger teenage girls, especially those younger than age 14, are highly likely to be coerced into sexual intercourse by older males. Tragically, the sexual use of young girls by older men is common in some developing countries with high rates of HIV/AIDS, where sex with very young girls, particularly virgins, is seen as a way to avoid or even “cure” AIDS.

Despite the understandable concern about sexual decision making and sexual coercion among young people, some experts on adolescent sexuality oppose the presumption that adolescent sexuality is invariably problematic. Both Fine and Tolman, for example, view sex among teens as developmentally normal and appropriate rather than inherently pathological. They decry the extent to which schools, parents, and religious institutions unwittingly conspire to scare young people, especially young girls, away from sex by equating sexuality entirely with negative outcomes—violence, victimization, pregnancy, disease, immortality, and low-class behavior. In their view, sexual desire and exploration are vital parts of the developmental processes of adolescence for girls and boys. They cite the high rates of adolescent sexual behavior throughout the world, and the increasingly responsible sexual behavior among key groups of teens in the United States, as evidence for this. For most people, a sexual life begins in adolescence, and there is increasing evidence in the United States and the rest of the industrialized

world that teens are not getting into as much trouble as they did in the recent past because of their sexual activity.

## 2. WHAT ARE RISKY SEXUAL BEHAVIORS?

As noted earlier, sex with another person carries with it a variety of potential risks. However, for the purpose of this analysis, risky or unsafe sexual behaviors are those that make it more likely that one will contract a STI or will contribute to an unwanted pregnancy, especially for adolescents. These include, but are not limited to, (a) having vaginal and especially anal sexual intercourse without a latex condom; (b) having multiple sex partners; (c) having sex outside of a committed relationship, especially with strangers, casual acquaintances, or people who have used intravenous drugs; (d) having sex under the influence of alcohol or other drugs, when sexual decision making is impaired; and (e) having sex under coercive conditions, when the possibility of meaningful choice, including the option of practicing safe sex, is eliminated.

## 3. WHAT GROUPS ARE AT HIGH RISK FOR TEEN PREGNANCY AND STIs?

It needs to be stated at the outset that it is sexual behavior that is risky, not being a member of a group. Being young or gay per se does not put a person at risk for STIs, for example, but engaging in anal sex without a latex condom does. That said, researchers have uncovered several correlates of risky sexual behavior, some of them involving subgroups of people. The fact that some geographic, social, ethnic, or other characteristics of people are associated with higher rates of teen pregnancies or STIs means only that some variables are statistically related and does not mean that there is anything inherent in group membership that causes these behaviors or negative outcomes.

Young people are vulnerable. Sexually active teens, especially young, unmarried teens, are generally at heightened risk for both unplanned pregnancies and STIs. This is in part because teens have less information about consequences of sexual behavior, less access to contraceptives, less skill in sexual communication and negotiation, and less ability to control a sexual

situation. As Rathus *et al.* suggest, another factor—the sense that young people have that they are invulnerable and can live forever—may also be at work. Many adolescents have the attitude that bad things, including pregnancies, STIs, and even death, simply cannot happen to them. However, as the Kaiser Foundation reported in 2002, more than 60% of new HIV infections in the United States occur in people aged 13–24.

People who have sexual practices involving sharing blood, semen, or feces are at high risk for HIV and other health problems. In the United States and much of the industrialized world, the AIDS epidemic first manifested in the 1980s among gay men and intravenous drug users. According to the CDC, in the United States AIDS predominantly affects men who have sex with men (50%) and people who share needles when they inject drugs (25–30%). These groups are at risk because some of their members engage in practices that involve sharing blood, semen, or feces, all of which can contain HIV. Researchers have found that gay men are more likely than lesbians and straight men and women to practice sexual behaviors that efficiently transmit the virus, for example, those in which semen containing the virus is directly transferred into the bloodstream through tears and open sores in the anus or rectum. These include practices that cause fissures, such as anal–genital sex and fist–anal activity (fisting). HIV is also contained in feces, so anilingus or oral–anal contact (rimming) is a high-risk activity. Beyond specific sexual practices, gay men are also more likely than others to have multiple sex partners.

Many changes in the gay community have occurred since the onset of AIDS, and many gay men have changed their behavior to prevent contracting or spreading AIDS. However, although male–male anal intercourse has been on the decline since the epidemic took hold, there has recently been an increase in one particularly dangerous form of anal intercourse, “barebacking,” which is anal intercourse without a condom. Unprotected receptive anal intercourse is especially risky because if the insertive partner is infected, the virus contained in the semen easily enters the bloodstream through the many small tears in the colon caused by anal penetration and thrusting. Anal sex is also associated with hepatitis B infection, anorectal human papilloma virus, and cancer. Again, it is behavior and not group membership that puts people at risk: Many heterosexual women are the receptive partners in anal sex, and, if their partners do not use latex condoms, they are at high risk of HIV infection.

Constellations of geographic, cultural, religious, and economic conditions have formed to make sex much more risky in certain places and times than others. HIV/AIDS is a global epidemic, but sub-Saharan Africa, which has approximately 10% of the world’s population, bears the burden of approximately 66% of the world’s AIDS cases. As Kelly notes, in some African countries one in four adults is infected, and it is estimated that 35% of boys now aged 15 will ultimately die of AIDS. However, in Africa, AIDS is mainly spread through unprotected vaginal intercourse and perinatally (i.e., around the time of birth) from mother to infant. Heterosexual transmission in Africa is facilitated by high rates of female genital mutilation among women, cultural practices and beliefs that discourage condom use, dry sex—sex without lubrication, migratory work patterns in men, and possibly low rates of circumcision among men. In Africa and, increasingly, in southern and Southeast Asia, men, women, and children are affected by AIDS in staggering numbers. The rates of both infection and death from infection are much higher in the poorer nations of the world. In the developed world, AIDS is still largely a disease of gay men and intravenous drug users, although heterosexual women infected by their male partners are the fastest growing group.

Within wealthy nations such as the United States, urban centers are the epicenter of the epidemic, with poverty, unemployment, alcohol use, and intravenous drug use all contributing to the problem. In the United States, race is confounded or correlated with many other social and economic factors. The Kaiser Foundation found that although African Americans make up only 13% of the U.S. population, they account for approximately 60% of new HIV infections and 54% of deaths from AIDS.

Women are at higher risk for some STIs, including HIV/AIDS, than men. There are many ways, from the biological to the social, in which sexuality is made more problematic for females than males. Of course, among the biological factors, one stands out above all others: Sexual intercourse is far more consequential for females than males because females alone bear children. However, being female also renders one more vulnerable to a variety of STIs. According to recent estimates, males transmit the HIV virus much more efficiently—approximately 10–18 times more efficiently—to females than females transmit the virus to males. Thus, acts of unprotected vaginal intercourse are in fact much more hazardous to women’s health than they are to men’s health.

#### **4. WHAT SOCIAL, INTERPERSONAL, AND PERSONALITY FACTORS ARE ASSOCIATED WITH RISKY SEXUAL BEHAVIORS?**

There are a number of factors that can predispose some people toward high-risk behaviors. One of the most reliable of all predictors of unsafe sex is the use of alcohol and other noninjectable drugs before engaging in sexual relations. Many studies show that substance use disinhibits people sexually and facilitates having multiple sex partners, sex with strangers, and sex without condoms. Teens who binge drink—who have five or more drinks on one occasion—are three times less likely to use a condom during sex than those who do not.

Relationship context matters. Women who believe that they are in monogamous relationships tend to view those relationships as safe, and they do not insist on condoms during sexual intercourse, regardless of what or how much they know about their partners' sexual histories or, for that matter, their partners' current sex lives outside of the primary relationship. Getting women in such relationships to see their behavior as risky means that they need to consider the possibility that their partner may not be "safe."

There is evidence that perceptions of the physical attractiveness of one's potential sex partner affect willingness to have unsafe sex. The more attractive a potential partner is perceived to be, the more likely people are to state an intention to have unprotected sex with that partner. This finding is robust and has been found among gay and straight respondents. It suggests that other desirable features of a partner besides physical attractiveness may lead people to abandon safe sex practices, for example, wealth or high social status.

Some personality factors appear to be related to sexual risk-taking. Self-regulation, defined as the ability to regulate one's emotions, attention, and behavior, is positively correlated with safe sex practices. On the other hand, people who are high in thrill or sensation-seeking are less likely to practice safe sex. For example, within the gay community, a minority of men who have sex with men deliberately seek out known high-risk behaviors, such as barebacking.

Despite all of the information available about safe sex, many people whose self-esteem is low do not accurately perceive risky behavior as such and consistently underestimate their vulnerability to diseases and conditions. Acknowledging self-destructive behavior is

anxiety provoking, is damaging to one's view of one-self, threatens self-esteem, and possibly is threatening to a valued sexually intimate relationship as well. To avoid negative feelings about the self while continuing to enjoy the rewards associated with sexual behavior, some people deny that they are engaging in high-risk behavior; indeed, some people deny that they are sexually active at all. These people use an avoidant coping style that enables them to continue engaging in the risky behavior. Some studies show that women who have been victims of emotional, physical, and sexual abuse and have low self-esteem are particularly vulnerable to having risky sex.

Finally, some people have personal beliefs and attitudes about sexual behavior that make it difficult to practice safe sex. Some of them are deeply rooted in religious teachings or cultural practices. For example, the Catholic Church does not condone the use of any kind of birth control, including condoms, and many observant Catholics follow this teaching. In some cultures, particularly ones in which gender roles are rigidly enforced, women have little choice in their sexual expression and must follow the lead of their men, who may regard the use of a condom as an affront to their masculinity, a sign that their partners have been unfaithful, are sick, or do not love or trust them, and so on. Some men regard sex with condoms as less spontaneous or sexually arousing and, regardless of their or their partners' HIV status, insist on "skin-to-skin contact."

#### **5. WHY DO PEOPLE ENGAGE IN RISKY SEX?**

As already noted, sex is a natural function, and one that perpetuates the species. Sex is also one of the most rewarding experiences available to people. Given these fundamental truths, we would not appear to need further explanation for why people have sex, even risky sex. However, recent psychological research suggests that sexual behavior is not simply motivated, and that sexual behaviors, whether risky or safe, serve a wide range of psychological functions that have little to do with perpetuating the species, health protection, or disease avoidance. As Cooper *et al.* argue, people have many complex motives for having sex, and these motives need to be understood before effective interventions to decrease risky behaviors can be implemented. Specifically, Cooper *et al.* argue that having sex to meet different needs may be associated with distinctive patterns of

risky behavior. These investigators have developed a motivational model of sexual behavior that considers two dimensions: whether the motivation is the pursuit of positive experiences versus the avoidance of negative experiences, and whether the focus is the self or the social environment. The crossing of these two dimensions yields four constellations: using sex to escape or avoid threats to self-esteem (coping motives), using sex to avoid or escape negative social experiences (approval motives), using sex to enhance one's own positive emotions or experiences (enhancement motives), and using sex to enhance social connections with others (intimacy motives). Research suggests that these motives do indeed account for sexual behavior and predict sexual risk-taking, particularly when relationship variables and sexual self-esteem are taken into account.

## **6. HOW CAN WE REDUCE THE INCIDENCE OF RISKY SEX?**

Sexual behavior is extremely complex, satisfies many motives, and is affected by many forces. Effective strategies for encouraging safe sex among young people not yet sexually active, and inducing people who practice risky sex to change their behavior, will need to include both individual and societal efforts. As McAnulty and Burnette argue, ultimate success will require a range of efforts involving the health care system, the government, the schools, religious organizations, peers, parents, lovers, and friends. Everything we know about changing personal behavior from smoking to exercising and eating suggests that changing sexual behavior will not be easy. Like these other personal behaviors, many risky sexual behaviors have the seductive feature of being rewarding in the short term but damaging in the long term. If anything, sexual behaviors are possibly more resistant to change than these other behaviors precisely because they are more complex, inherently interpersonal, highly rewarding, and private. McAnulty and Burnette suggested a number of behavioral patterns that can reduce or eliminate risky sexual practices.

### **6.1. Practicing Abstinence and Monogamy**

The only prevention method that is 100% effective in eliminating the threats of unwanted pregnancy or STIs is abstinence. Abstinence can mean many things, from total avoidance of sexual activity to avoidance of

penetrative sex. Many sexual activities that involve others, including kissing, petting, mutual masturbation, and the use of sex toys, are considered completely safe.

The next most effective STI prevention method is monogamy—having a sexually exclusive relationship with one partner. Obviously, for monogamy to be effective in preventing STIs, both partners must be free of disease. To prevent pregnancy, some form of birth control must be used.

### **6.2. Making a Commitment to Practice Safe Sex**

Many STI prevention programs in schools and communities encourage people to make an explicit commitment to practice safe sex. This means encouraging people to identify their personal sexual goals and identify and implement the strategies that they will need to follow to achieve those goals in their own lives. The more public the commitment, and the more preparation that goes into the strategies, the more likely people are to follow through on their commitments.

### **6.3. Sexual Communication Skills**

People, especially women, can reduce their vulnerability in sexual situations by communicating more effectively and assertively with their partners. First, this involves educating oneself about STIs, their risks, and what one needs to do to protect oneself. Next, this means initiating desired activities and refusing unwanted activities, and negotiation about condom use. Aside from condom use, a couple may negotiate such sexual matters as sexual exclusivity and getting HIV antibody testing together. Being direct and honest with potential partners about one's own health and risk status is key, as is getting relevant, honest information from one's potential partner about whether the partner has, for example, had a recent HIV test or has ever had unprotected anal sex, sex with commercial sex workers, or engaged in intravenous drug use. Unfortunately, even honest answers to direct questions may not be enough because many people are unaware of the risks associated with some behaviors and of their own disease status. Moreover, many surveys show that a significant minority of people—approximately 20–35%—lie about their sexual history or disease status to avoid rejection, further complicating negotiated safety.

The most important communication about sex takes place between people who are contemplating sexual relations, but open communication about sex should

occur in other forums and among other groups of people as well. Research suggests that parents can play a constructive role in affecting their children's sexual risk-taking. Mothers appear to be particularly critical in the development of responsible adolescent sexual behavior. Open communication in a nonjudgmental way between mothers and teens about sexual matters and parental monitoring of teen behavior predicts lower sexual risk-taking in adolescence.

#### 6.4. Avoiding Known High-Risk Behaviors

Alcohol and other drugs do not cause STIs, but they are consistently associated with high-risk sexual behavior because they can be disinhibiting and interfere with a person's judgment and resolve. If one remains sober, one is more likely to use condoms and less likely to engage in high-risk sexual behaviors.

#### 6.5. Getting Regular Medical Checkups

Early detection of STIs is critical to treating them effectively. Moreover, the presence of untreated STIs makes new infections, including HIV, more likely. Many STIs, such as syphilis and chlamydia, have few or no symptoms and, if untreated, can have serious consequences. Routine screenings—approximately twice a year—are highly recommended for sexually active people generally and for couples about to become sexually active. Many public health centers provide such services for free or at a low cost.

### See Also the Following Articles

Homosexuality ■ Sexual Behavior ■ Sexual Behavior and Culture

### Further Reading

- Buss, D. M. (1989). Sex differences in human mate preferences. Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, 12, 1–49.
- Centers for Disease Control and Prevention. (2000, June 9). Youth risk behavior surveillance—United States, 1999. *Morbidity and Mortality Weekly Report*, 49(SS05), 1–96.
- Cooper, M. L., Shapiro, C. M., & Powers, A. M. (1998). Motivations for sex and risky sexual behavior among adolescents and young adults: A functional perspective. *Journal of Personality and Social Psychology*, 75(6), 1528–1558.
- Fine, M. (1988). Sexuality, schooling, and adolescent females: The missing discourse of desire. *Harvard Educational Review*, 58, 29–53.
- Fraser, A. M., Brockert, J. E., & Ward, R. H. (1995). Association of young maternal age with adverse reproductive outcomes. *New England Journal of Medicine*, 332, 1113–1117.
- Kaiser Foundation. (2002). *HIV/AIDS policy fact sheet*. Washington, DC: Henry J. Kaiser Family Foundation.
- Kelly, G. F. (2004). *Sexuality today: The human perspective* (7th ed.). Boston: McGraw-Hill.
- McAnulty, R. D., & Burnette, M. M. (2004). *Exploring human sexuality: Making healthy decisions* (2nd ed.). Boston: Allyn & Bacon.
- Rathus, S. A., Nevid, J. S., & Fichner Rathus, L. (2002). *Human sexuality in a world of diversity*. Boston: Allyn & Bacon.
- Tolman, D. (2002). *Dilemmas of desire*. Cambridge, MA: Harvard University Press.



# Robotherapy

Elena V. Libin

*Institute of Robotic Psychology and Robotherapy, Cyber-Anthropology  
Research Incorporated, Chevy Chase, Maryland, USA*

Alexander V. Libin

*Georgetown University, Washington, DC, USA*

1. Robotherapy: A New Era of Interactive, Nonpharmacological Interventions
2. A Concept of Artificial Partners
3. Robots as Therapeutic Tools: Early Experiments in the Field of Robotherapy
4. Person–Robot Communication as Complex Interactive System  
Further Reading

## GLOSSARY

*artificial partner* An interactive robotic or digital creature that provides positive sensorimotor, cognitive, or emotional stimulation.

*nonpharmacological therapy* The use of psychologically based or activity-based techniques for treating a person's problematic behaviors and negative emotional states.

*robotherapy* A framework of human–robotic creature interactions aimed at the reconstruction of a person's negative experiences through the development of coping strategies, mediated by technological tools, in order to provide a platform for building new positive life skills.

The use of artificial interactive creatures as therapeutic tools is a practical application that resulted from a newly developed theoretical and methodological field called robotic psychology and robotherapy. The innovative concept of robotherapy offers methodological and experimental justification for the use

of nonpharmacological interventions based on stimulation, assistance, and rehabilitation techniques for people with cognitive impairments, physical disability, special needs, and/or psychological problems. Robotherapy has been defined as a framework of human–robot interactions aimed at the reconstruction of a person's negative experiences through the development of coping strategies mediated by technological tools in order to provide an emotional and cognitive platform for building new positive life skills. Field experiments have shown the effectiveness of robots as human companions and therapeutic agents.

## 1. ROBOTHERAPY: A NEW ERA OF INTERACTIVE, NONPHARMACOLOGICAL INTERVENTIONS

A great variety of therapeutic techniques and theories traditionally related to mental health can be reduced to a basic dichotomy, defined as pharmacological vs nonpharmacological approaches to treatment. The main effects of pharmacological interventions are based on drugs' impact on human physiology associated with biochemical processes. Nonpharmacological approaches for treating problems that a person experiences during the course of his or her life are aimed at psychological mechanisms that underlie behaviors that are disturbing to the person or others. Psychological effects produced

by nonpharmacological interventions influence biochemical and neurological patterns that form individual behaviors, states, and traits. Psychological therapy achieves the same effects as pharmacological methods, affecting the wide range of mechanisms that underlie psychophysiological manifestations of emotions and complex social problematic behaviors. Among psychologically oriented techniques, a new group of technology-mediated therapy is rapidly developing. The invention of the microchip brought new technological tools to the therapist: telemedicine, virtual reality, computer games, and robotic creatures. Four major approaches are distinguished in technology-mediated therapy:

1. Virtual reality techniques for treating anxiety, attention deficit disorder, fear of flying, and a variety of phobias based on the phenomenon of presence and immersion
2. Internet-based communication, including telehypnosis, distant psychotherapy, and therapy-oriented knowledge acquisition from electronic databases
3. Electronic games tailored for clinical application
4. Robotherapy based on a person's interactions with embodied agents, such as social and entertainment robots, with different levels of artificial intelligence and integrated sensory feedback

The use of artificial interactive creatures as therapeutic tools for medical and psychological practice is the subject of a new field of research called *robotherapy*. The first definition of *robotherapy* by Libin and Libin in 2002 describes it as a framework of human-robotic creature interactions aimed at the reconstruction of a person's negative experiences through the development of coping strategies that are mediated by technological tools in order to provide a platform for building new, positive life skills. *Robotherapy* deals with the analysis of person-robot communication, viewed as a complex interactive system, while emphasizing psychological evaluation, diagnosis, and prognosis of problem behaviors and principles of nonpharmacological treatment.

## 2. A CONCEPT OF ARTIFICIAL PARTNERS

A paradigmatic shift in human-computer interaction studies a range of areas, from human factors to personal robot assistance, and outlines a new class of robots with therapeutic potential. In 2001, Libin defined artificial partners as interactive robotic or digital creatures

that provide positive sensorimotor, cognitive, or emotional stimulation for a human partner. The concept of artificial partner (ART) places the relationships between humans and robots into a psychological, rather than technological, context.

The class of interactive stimulation robots as human companions is distinguished by two characteristics that make it potentially valuable for sociopsychologically oriented research and practice. First, the design and behavioral configurations of interactive artificial creatures are tailored to the purpose of communicating with a human being on a "personal" level. The interactions between a person and a robot include tactile-kinesthetic, sensory, emotional, cognitive, and social-behavioral patterns. Second, ARTs imitate real-life (human- or animal-like) behavior and model motor, emotional, and cognitive behaviors normally experienced by living beings. These features form the artificial partner's functional status, which reflects a structure of a living world, making interactive agent an interesting and engaging communicating, gaming, or therapeutic partner for people of all ages, cultures, and life experiences. Numerous artificial creatures that are equipped with touch, audio, visual sensors, and different levels of robo-IQ exist on the market and in the laboratory. Anthropomorphic robot platforms Cog and Kismet, Doc Beardsley and Nursebot Pearl, humanoid AMI and ASIMO, emotional robotic cat NeCoRo, robotic dog AIBO and robot seal Paro, and automated dolls Amazing Amy and My Real Baby are the most vivid examples.

## 3. ROBOTS AS THERAPEUTIC TOOLS: EARLY EXPERIMENTS IN THE FIELD OF ROBOTHERAPY

The function of artificial partners is changing from a role as entertainment agents to include what were originally exclusively human occupations, such as education and therapy. Artificial creatures with high robo-IQs and integrated sensory, evaluating, and controlling systems perform certain functions in the therapy process, such as the following:

1. A diagnostic method for studying individual coping and defense mechanisms
2. A tool for guided physical and mental stimulation
3. An interactive device for training and development of certain individual and group skills
4. A mediator in person-to-person communication

### **3.1. Coping with Negative Behavioral and Emotional States via Technology: Behavioral Therapy**

Research with the lifelike robotic seal Paro, developed by Takanori Shibata in 1998, was aimed at studying the psychological, physiological, and social effects of robotic pet interactions with both children in the hospital setting and elderly persons in nursing homes. Paro is covered with synthetic fur, which makes interactions based on tactile–kinesthetic senses possible. Results showed that physical interactions mediated by the robot’s tactile sensors positively influenced the subjective evaluation of the artificial creature’s appearance. Analysis of attitudes toward Paro revealed two general factors that characterized the individual experience of communicating with the robot companion: The first factor was related to the unique tactile properties of the artificial creature, and the second factor was responsible for its communicative value. Analysis of psychophysiological and psychological indicators in the context of therapeutic effects produced by the human–robot communication showed that the texture and softness of Paro’s fur, in combination with its empathic appeal, had a positive therapeutic effect on the mood and associated state of comfort in interacting with persons of different ages, genders, and medical and cognitive status. It is also promising for future research that the 2003 *Guinness Book of World Records* considered the Paro to be “the most therapeutic robot” ever designed.

### **3.2. Robots for Personal Assistance: Psychological and Physical Rehabilitation**

Robotherapy studies have been conducted on diverse clinical populations, such as children and elderly with special needs, physically and cognitively impaired persons, and people with a wide range of psychological problems. In 1967, the founder of cybernetics, Norbert Wiener, noted that a major function of sophisticated communicative machines (as well as higher living organisms) is the ability to resist entropy by modifying their behavioral configurations on the basis of past experiences and by using sensitive feedback. Robotic creatures as therapeutic agents adjust to individual preferences and needs by collecting information associated with various modalities of the person involved with the treatment.

Allison Druin’s team at the University of Maryland’s Human Computer Interaction Laboratory developed robotic PETS or Personal Electronic Teller of Stories. Children customize the robot’s appearance using different animal parts (dog paws, wings, horns, etc.) so that the robot has a personal representation. Children then create stories to express how they feel and program the PETS robot to express these emotions through movement and sound. This method of communication gives children with special challenges a sense of control and fulfillment. Through interactions with the robot, children overcome their own limitations of emotional and verbal expression. A storytelling robot has the potential to provide effective therapy for emotionally challenged children (e.g., those with autism).

Gestural interface technology is based on recognition of human nonverbal behavior and associated emotions, and it is the core of the tactical mobile robot called JesterBot. When attached to a person in a wheelchair, JesterBot collects signals from tiny muscle movements and transforms them into meaningful verbal and physical interactive patterns. The robot modifies its own behavior by monitoring the vital signals and retrieving records about the person’s sensorimotor profile.

CosmoBot, invented by Corinna Lathan’s team at AnthroTronix, Incorporated, is designed to be a companion for children working on developmental goals associated with physical, occupational, or speech/language therapy or on special education goals. An interactive rehabilitation robot detects the child’s movements and/or voice commands, transforming them into command signals to make the robot move and/or talk. This rehabilitation aid is programmable, allowing the child to interact with CosmoBot in the context of new games and stories. Furthermore, the robot’s data-collection capabilities allow therapists and educators to objectively track the child’s progress with regard to developmental goals.

Robots assist mentally and physically challenged persons by facilitating support for their daily living activities through the enhancement of their sense of control. In 1993, Lancioni *et al.* developed a robot-assisted occupational program that provides choice opportunities for people with severe retardation and multiple handicaps. Another approach employed a group of robots with different design and functionality for a group of neuropsychologically impaired persons in order to compensate for their limited cognitive and physical resources. Individuals’ comprehension abilities, physical needs, and communicational styles were matched to the robots’ capacity to provide visual, auditory, and motor support and communicative stimulation. In 2002–2004, Libin

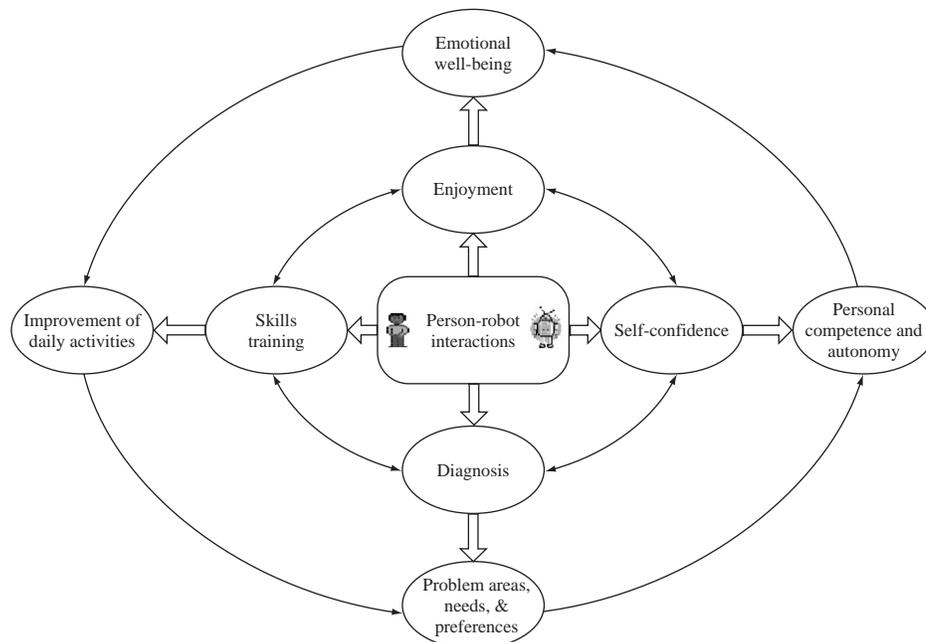
and Libin studied the impact of robotic-based entertainment systems on positive, emotional experiences and disruptive behaviors in persons with limited physical, sensorimotor, and cognitive resources. Libin and Cohen-Mansfield showed that for elderly people with dementia, interactions with the robotic pets resulted in decreased agitation and increases in positive emotions.

Interactive robots are capable of promoting sensorimotor activities, triggering cognitive and emotional stimulation, and providing an individual with psychological benefits (i.e., a sense of control, independence, and self-efficacy). The effectiveness of the robotic devices implemented during the course of therapy depends on a user-friendly design, a software program that provides choice-making behavior, and the feedback system with verbal and nonverbal reinforcement.

#### 4. PERSON-ROBOT COMMUNICATION AS COMPLEX INTERACTIVE SYSTEM

The theoretical framework of robotherapy, in which an artificial creature communicates with a person for assistance

or therapeutic purposes, stems from the concept of complex interactive systems, which is based on three principles. First, the structure of person-robot communication can be viewed as a constellation of associated exchanges between patterns related to different modalities. The robot as an embodied therapeutic function coexists with an interacting person in the same given space-time continuum, becoming an interactive agent. This quality of an artificial creature is vitally important for people with limited resources (i.e., physical, cognitive, or emotional impairment) or persons who strive toward further development through mastery of certain skills. Second, because an interactive system behaves equifinally (“the same goal can be reached from different initial conditions or in different ways,” as von Bertalanffy noted in 1966), robots with various levels of artificial intelligence and types of sensory feedback can benefit people with different combinations of skills and abilities as an effective tool in a very broad context of individual situations. Lastly, a systematic approach concentrates on the issue of matching a robot’s behavioral configuration with individual profiles. The components of this interactive system are summarized in Fig. 1. The model was developed using available data derived from studies on social robotics, computational



**FIGURE 1** Multidimensional model of a person-robot communication as a complex interactive system. The following are the correlations between system parameters and types of interactions: skills training—sensorimotor stimulation, attention, and memory; diagnosis—identification of problem areas, individual needs, and preferences; self-confidence—increasing personal competence and autonomy; and enjoyment—enhancing emotional well-being.

entertainment, modern information technology, and cybertherapy.

Two loops underlying the person–robot communication describe elements of a complex interactive system. The internal loop is formed by the patterns that underlie skills training, problem diagnosis based on monitoring vital signals, enjoyment resulting from emotional stimulation, and self-confidence developed through the sense of control. The external loop embraces the results of the internal loop functioning, such as improvement of professional or daily activities due to sensorimotor and cognitive training; positive well-being; personal competence and autonomy; and mental and physical profile highlights tailored to one’s problem areas, needs, and preferences. The internal loop produces such integrative system effects as compensation and optimization in a person’s functioning, whereas the external loop yields effectiveness and adequacy of interactions.

The effectiveness of robotherapy is influenced by (i) one’s past experiences, current needs, and individual preferences and (ii) an artificial partner’s nontransitive physical features and behavioral configurations defined through the intensity of stimulations and responses. At the same time, the effectiveness of robotherapy is associated with the productive compatibility between human beings and their artificial partners, which is achieved through systematic study of individual needs and preferences, user-friendly design and likeable appearance of the robotic creatures, adjustable behavioral configuration of a robot, and multilevel (tactile, sensorimotor, cognitive, emotional, and social) person–robot interactions based on the open-loop principle.

As a new way of helping people, robotherapy is aimed at the improvement of personal well-being through the development of various coping skills mediated by technological tools. The use of robots in therapy poses a basic moral dilemma regarding whether robotic assisted care should be a substitute for human care. The implementation of smart interactive technology for the treatment of people raises questions about the responsibility of the therapist for the quality of care. Robotherapy employs a humanistic approach in psychological consulting and therapy, according to which a psychological culture of the therapist defines the therapeutic value and the proper use of technological innovations.

### See Also the Following Articles

Coping ■ Robotic Psychology

### Further Reading

- Fukuyama, F. (2003). *Our posthuman future*. New York: Picador.
- Libin, A., & Libin, E. (2004). Person–robot interactions from the robopsychologist’s point of view: The robotic psychology and robotherapy approach. In *Person–Robot Interactions for Psychological Enrichment*, Special Issue. Los Alamitos: IEEE.
- Ludwig, L. (2001). *Robo sapiens*. Cambridge, MA: MIT Press.
- Mahony, M. (2000). *Cyborg: The man machine*. Bradford: Bradford Press.
- von Bertalanffy, L. (1967). *Robots, men, and minds: Psychology in the modern world*. New York: George Braziller.



# Robotic Psychology

**Alexander V. Libin**

*Georgetown University, Washington, DC, USA*

**Elena V. Libin**

*Institute of Robotic Psychology and Robototherapy,  
Cyber-Anthropology Research Incorporated,  
Chevy Chase, Maryland, USA*

1. Robotic Psychology: On the Coexistence of Humans and Robots
  2. Psychology of Person–Robot Interactions
  3. Companionship between People and Robots
- Further Reading

## GLOSSARY

**robo-IQ** The complexity of the software-based artificial intelligence related to an adequate performance of an uncertain or multiple-choice task.

**robot** An automat with programmed behavior designed for the purposes of substituting for certain human activity or to satisfy a person's physical, psychological, or social needs.

**robotic psychology** A study of compatibility between people and artificial creatures on different levels: sensorimotor, emotional, cognitive, and social.

Robotic psychology is defined as an interdisciplinary field of research and practice that focuses on the compatibility between people and artificial creatures on different levels: sensorimotor, emotional, cognitive, and social. Robotic psychology studies the psychological significance of robots' behavior and its intertwining with elements of physical and social environments. Different classes of robots are aimed at fulfilling different human

needs. Artificial creatures have their own distinct individuality, which manifests itself in the robot's design and behavioral configuration, in the same way that people and other living beings differ from each other in various parameters, such as weight and height, behavioral reactions and character, emotions and cognition, abilities, and coping strategies. Analyses of person–robot interactions account for the essentials of both humans' behavior and that of their artificial partners.

## 1. ROBOTIC PSYCHOLOGY: ON THE COEXISTENCE OF HUMANS AND ROBOTS

Interactions with robotic creatures of different kinds are rapidly becoming part of people's daily lives. Satisfying various human needs, robots with artificial intelligence and sensitive feedback perform industrial and military functions, research and medical procedures, and assist in educational and therapeutic tasks. A study on human–robot coexistence based on innovative robotic psychology concept now encompasses cybernetics and humanities, engineering and biology, and information sciences and aesthetics. An analysis of robotic beings and their behavior from the psychological standpoint opens new perspectives for theoretical and practical applications in both technological and social sciences. Robotic psychology as a new

research area employs a systematic approach in studying psychophysiological, psychological, and social aspects of person–robot communication.

The search for an artificial being as a human companion dates back to ancient mythology and creatures such as Golem in medieval literature. However, the contemporary understanding of an automatic creature began when Czech writer Karel Capek coined the term “robot” (meaning a worker or a servant) in 1921 and Russian-born U.S. author Isaac Asimov developed the concept of robotics in the 1940s to describe the field of designing and using automats for human purposes. The term robot can be defined nowadays as an automat with programmed behavior aimed at substituting for certain human activities or to satisfy a person’s physical, psychological, or social needs.

### 1.1. Classification of Artificial Creatures in Robotic Psychology

Due to the diversity of human activities, a variety of robotic creatures are required to help in the performance of these activities. From the perspective of human needs, all kinds of interactions between people and robots deserve psychological attention. Contemporary primary classification of artificial creatures is based on a robot’s relation to human activity, a robot’s behavioral configuration defined by degrees of freedom, robo-IQ (complexity of software-based artificial intelligence related to an adequate performance of an uncertain or multiple-choice task), and the robot’s physical properties manifested in the hardware architecture and appearance (Table I).

Robotic psychology emphasizes three major modes of a robot’s relation to human needs: substitute for human activity (e.g., rescue robots), extension of human activity (e.g., medical robots), and stimulation of human activity (e.g., social robots and therapeutic robots).

## 2. PSYCHOLOGY OF PERSON–ROBOT INTERACTIONS

Rodney Brooks, a scientist from the Massachusetts Institute of Technology (MIT), developed a concept of robotic beings. He argued that robots’ behavior must be based on the principle of dynamic interactions with changing physical and social environments. A robotic being is perceived as an embedded part of human relations with the world, as one of the components of the interactive process but not the separate system. To

be relevant to the unpredictable psychological reality of its human partner, the organization of a robotic creature’s behavior needs to be built with the motor and sensory loops running in parallel rather than employ hierarchically structured functioning.

### 2.1. Modeling the Human Mind

Robots’ capacity to adjust to the requirements of the human world stems from the concept of artificial intelligence, or a set of programs that control the automatic actions of machines. John McCarthy, who coined the term artificial intelligence (AI) in 1956, was an enthusiast of the psychologization of computers. This position reflects an idea that there is no difference between human reasoning and computer programming. A different approach expressed by philosopher John Searl states that although the human mind has the ability to compute, it is completely different in its nature from AI. Whereas robotic actions result from embedded programs, human behavior is determined by subjective needs and preferences, likes and dislikes, and intentionality. That is why it is such a challenge to build an artificial creature capable of interactions relevant to the complexity of human relationships. It is of no coincidence that the competence of a robot’s AI is tested by the human expert implementing the Turing test, based on the robot’s ability to answer random questions in the same manner as people do.

On the other hand, it is considered an honor to challenge an artificial partner in fair competition, such as in a game of chess. Grossmeister Gary Kasparov, along with the rest of the world, regarded IBM’s Big Blue robot as a serious gaming partner in a person–robot chess competition. Sophisticated artificial intellect requires an adequate vehicle. In 2003, Hugo de Garis developed an artificial brain with 37.7 million artificial neurons that will allow a robot-kitten designed by Michael Korkin to perform an animal-like intelligent behavior based on retrieving sensory experience from the different parts of the environment.

### 2.2. Modeling Emotions

The idea of modeling intellectual functions of the human mind has existed throughout the history of robots. The reproduction of human emotions did not attract investors’ attention as much as the built-in cognitive component of artificial creatures. Nevertheless, the new millennium was marked by the paradigmatic shift in robot design from mechanocentric to human-oriented

**TABLE I**  
**Classification of Artificial Creatures in Robotic Psychology**

	<i>Human need</i>	<i>Behavioral configuration</i>	<i>Physical appearance</i>
Industrial robots	Perform hard labor and hazardous work		
Research robots	Provide extension of human sensorimotor capacities	Defined by a situated activity that substitutes for human behavior	Machine-like appearance with emphasis on peripheral devices
Military and rescue robots	Activity in life-threatening situations		
Medical robots	Performance of fine motor operations on the human body		
Recreational robots	Trigger positive behaviors through entertainment	Aimed at human feelings and emotions	Animated form of existing and non existing beings or objects
Interactive stimulation robots			
Social robots	Communication	Based on human facial expressions and complex gestures with social meaning	Anthropomorphized (humanoid) appearance
Robots with therapeutic potential	Therapy and treatment of negative emotional states and behaviors	Models basic emotional states and social behaviors embodied in lifelike form	Animal-like or human-like appearance

principles, which transform the very purpose of artificial creatures' existence, their architecture, and their physical appearance.

A new concept of an artificial emotional creature proposed an original lifelike design and behavioral configuration relevant for natural human communication. The robotic cats Tama, designed in 1998 by Toshihiro Tashima, and NeCoRo, created by Takanori Shibata and colleagues, employ fuzzy logic in the organization of AI, which makes the robotic creatures' behavior "unpredictably" engaging and emotionally attractive for humans. A distinct feature of emotional robot pets is the use of tactile sensors covered with "skin," a synthetic fur. The combination of lifelike behavior and tactile stimulation creates an appeal that is missing in many other robotic models. In 2002, Libin and colleagues studied the influence of past experiences with real animals and technology on a person's emotional communication with a robotic creature, the Omron Corporation's robotic cat NeCoRo. The intensity and direction of emotional responses triggered in a human partner by the robot were

associated with the human's relevant past experiences. People perceived the robots as part of their living world.

### 2.3. Modeling Social Behavior

Regardless of shape (human or animal), the robot's most difficult task is to model people's social behavior with such cues as eye contact, gestures, and emotional answers recognizable in the social context. Social robots model the natural process of communication and are able to learn through social interactions. Braezeal and colleagues applied a developmental psychology approach to the design of the autonomous robot Kismet, build at the MIT media lab in 1998. The learning processes of robots are similar to those of children and are described by Lev Vygotsky's concept of the proximity zone; that is, the complexity of the stimuli increases when a learning agent reaches a certain level of competency in performing the task. In the robot's process of social learning, this condition is achieved by maintaining a level of interactivity that is neither overwhelming nor understimulating.

## 2.4. Modeling Teaching Behavior

During the education and learning process, both the levels of robo-IQ or communicational ability and the robot's anthropomorphic characteristics are important. Children ages 3–5 years are more motivated to learn from a personified mobile robot with a smiling face and verbal positive reinforcement than from an immobile and nonverbal machine-like device.

Analysis of children's needs and preferences resulted in the creation of a realistic robotic doll called My Real Baby, which enhances a child's communication experience via 15 human-like emotions of different intensity: facial expressiveness, soft and flexible skin, and a rich vocabulary of sounds and words. This doll, which is in the class of interactive stimulation robots, allows a child to develop educational interactive play experiences triggered by his or her imagination and creativity through natural responses expressed in emotionally appropriate ways.

A variety of humanoid robots, such as Kismet and Cog, Doc Beardsley and Nursebot, and AMI and PINO, are able to provide warmth and friendliness for their human partners. Another option considers animals, especially pets, as prototypes for the robotic creature design. Among the most interesting examples are the robot seal Paro and horse Karakuri, dog AIBO and cat NeCoRo, and robo-puppy Techno and robo-kitten Robokonenko. Robotic pets' resemblance to real animals and associated human experience allowed Peter Kahn and colleagues in 2002 to study the influence of communication with robotic dog AIBO on children's cognitive, social, and moral developmental processes. Children's conceptualization of the robotic creatures' behavior was similar to their perception and reasoning based on experience with real pets.

## 3. COMPANIONSHIP BETWEEN PEOPLE AND ROBOTS

A core aspect of the concept of companionship between people and robots is their compatibility. Compatibility in the robotic psychology context depends on how humans perceive their robotic companions and how they feel about them emotionally. In a study conducted by the Sakamoto Laboratory of Ochanomizu University in collaboration with the NEC company, three robotic creatures—dog AIBO and humanoids ASIMO and PaPeRo—were compared with each other as well as with animals, humans, and inanimate objects. People's

perceptions of the robots differed from their perceptions of animals, other humans, and inanimate objects. The appearance and shape of the robots influenced the level of comfort people experience around technological devices, especially robotic creatures.

In 2002, Libin, Libin, and colleagues performed a cross-cultural study of the United States and Japan that investigated biological, psychological, and cultural determinants of people's communication with the robotic cat NeCoRo. A culture was defined as a system of created social meanings and artifacts that have symbolic value and reflect the logic of human mentality derived from norms shared by a given community. A specially designed scale evaluated verbal and nonverbal, emotional, and tactile-kinesthetic patterns of communication. For age and gender, preferences regarding the robot's features and interactive patterns were similar in both cultures. Older people liked the robotic pet more, whereas men enjoyed the cat's triggering behavior more than women. The influence of a person's past experiences with live pets and modern technology on the type of person-robot communication was universal. Differences between cultures were found for the participant's evaluation of the robotic creature's appearance and character and the intensity of tactile interactions.

Robotic psychology research incorporates the elements of training, education, entertainment, and therapy into the person-robot communicative process, which leads to the development of an enhanced constructive model of interactive technology.

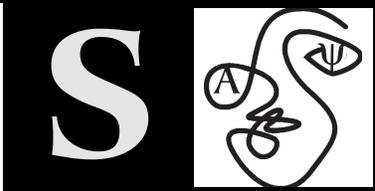
### See Also the Following Articles

Emotion ■ Learning ■ Robototherapy ■ Social Skills Training  
■ Teaching Effectiveness

### Further Reading

- Arkin, R. (1997). *Behavior-based robotics*. Cambridge, MA: MIT Press.
- Brazeal, C. (2003). *Designing sociable robots*. Cambridge, MA: MIT Press.
- Brooks, R. (2002). *Flesh and machines: How robots will change us*. Cambridge, MA: MIT Press.
- Libin, A., & Libin, E. (2003). Robototherapy and robotic psychology: New diagnostic tool for measuring person-robot communication. *Cyberpsychology and Behavior*, 6(4), 369–374.
- Moravec, H. (2001). *Robot: Mere machine to transcendent mind*. Cambridge, MA: MIT Press.

---



# Safety and Risks: Errors and Accidents in Different Occupations

Doris Fay and Patrick Tissington

Aston University, Birmingham, United Kingdom

1. Introduction
  2. Errors: From the “Freudian Slip” to Taxonomies of Errors
  3. Strategies of Error Prevention
  4. Antecedents of Accidents: The Search for Individual Difference Variables
  5. From Engineering Control Solutions to Safety Behavior
  6. Recent Approaches to Safety: The Role of Management and Organizational Climate
  7. Safety-Enhancing Interventions
  8. Summary
- Further Reading

**system design** The deliberate application of science to the planning and specification of the operation of machines by humans.

**unsafe behaviors** Behaviors that are not in accordance with safety rules and procedures; they enhance the likelihood of accidents.

## GLOSSARY

**accident** An unexpected event, damaging person(s) and/or a system, involving the potential to impair the goal accomplishment.

**error** Intentional action that does not result in the attainment of the goal; prevention of the error must have been under the control of the person(s), thereby excluding an “act of God” as the reason for the nonattainment.

**human factors** The study of humans and their interaction with machines and the environment.

**safety climate** Employees’ perception of the organization’s stance on compliance with safety-based rules and procedures and their perception of the priority that management accords to occupational safety.

Errors and accidents are a severe threat to the well-being and performance of the individual worker, of the social and technical work system, and of the wider public and environment. Therefore, understanding factors and interventions that reduce errors and enhance safety is of interest for multiple stakeholders. This article reviews the diverse theoretical and practical approaches to errors and accidents. It introduces prominent error taxonomies and strategies of error reduction such as appropriate system design. Contemporary models of accidents acknowledge that they are the result of a multifactorial process in which errors can play one part. The article highlights factors such as quality of system design, individuals’ safety behaviors, behaviors of management, and the organization’s safety climate. It discusses interventions that relate to the various factors involved.

## 1. INTRODUCTION

“A man who makes no mistakes does not usually make anything.” This quotation, from Bishop William Conor Magee, makes the case that errors are a natural and unavoidable component of human activity, and some theories of individual and organizational development take a positive stance to errors. Such theories take the view that errors contribute to the development of human knowledge and skills in that they represent opportunities for organizational learning. However, errors can be seen as a double-edged sword. In occupational settings, errors predominantly have negative connotations because they are associated with reduced performance and quality. For example, in manufacturing, errors lead to higher costs through time spent on rectification or product recalls. In health care, errors can involve severe threats to the safety of patients and care providers. Furthermore, errors are frequently causal factors in accidents. For example, errors in power plants or chemical production can cause accidents that have significant consequences to the health and well-being of the people immediately involved, to the wider public, and to the environment.

On March 6, 1987, the ferry *Herald of Free Enterprise* capsized and sank outside the Belgian port of Zeebrugge, resulting in the loss of 188 lives. The cause of the accident became clear almost immediately. The ship had set sail with its large bow car-loading doors open, allowing water to cascade into the car deck as the ferry increased speed. The subsequent inquiry found that the root cause of the accident was a series of errors by individual operators, designers, and the organization's senior management. Essentially, the accident was caused by a junior member of the crew failing to operate a series of levers that would have closed the watertight doors. However, there was some doubt as to whose responsibility this was, given that the captain did not know that the doors were open. Many requests had been made by officers to add a warning light on the bridge to indicate when the doors were open, but these requests had gone unheeded by the company's managers. Policies on status reporting were confusing but were founded on a principle of “negative reporting only,” so there was no emphasis on checking. Clearly, this was a tragedy on a large scale and was of great interest to researchers in accidents and errors due to the multiply stranded antecedents of the accident. This one incident graphically demonstrates the potentially disastrous end result of errors and shows that the study

of errors in organizational contexts is a vital area of study and must include both rigor and practical application. The scientific approach to accidents and their causes is to develop models of accident causation.

Contemporary models of accidents acknowledge that accidents are the result of a multifactorial process in which errors play a part. Furthermore, the display of unsafe behaviors that, in turn, influence the likelihood of an accident is given a pivotal role. Models suggest that diverse factors such as managerial strategies, design of the system, equipment, design of the entire job, characteristics of the worker(s), and social environment influence unsafe behaviors. Figure 1 illustrates a number of factors that are relevant to accidents. The costs involved in errors and accidents are social, economic, and psychological ones. They include individuals' pain and suffering, insurance settlements, indirect costs such as investigation time and downtime, and replacement of worker(s) and/or equipment. Therefore, understanding factors that assist in the prevention of errors and accidents is critical. In what follows, the article describes current approaches to errors, discusses factors contributing to safety, and then shows how they are used for safety interventions.

## 2. ERRORS: FROM THE “FREUDIAN SLIP” TO TAXONOMIES OF ERRORS

The study of error, in terms of its prevention, has been the subject of research by psychologists for more than 100 years. An early attempt to explain error in psychological terms was made by Freud, who viewed error as being indicative of an underlying need or some other purpose—the so called “Freudian slip.” Consider the case of a train driver passing a signal that was set at “danger”—an error worthy of study for obvious reasons. Tracing this error to some kind of psychosexual origin in the driver has little currency today. Many other approaches to error have focused on the person making the error and attributed error to carelessness or inattention of the individual concerned—in other words, attributing blame to the driver of the train in this example, or to the operator of the watertight bow door in the example of the Zeebrugge ferry, and possibly reprimanding him or her. Modern researchers question the value of this approach because it tends to correct a very specific error made by an individual on a specific occasion and, therefore, is unlikely to aid in the prevention of future errors. Current thinking favors the development of

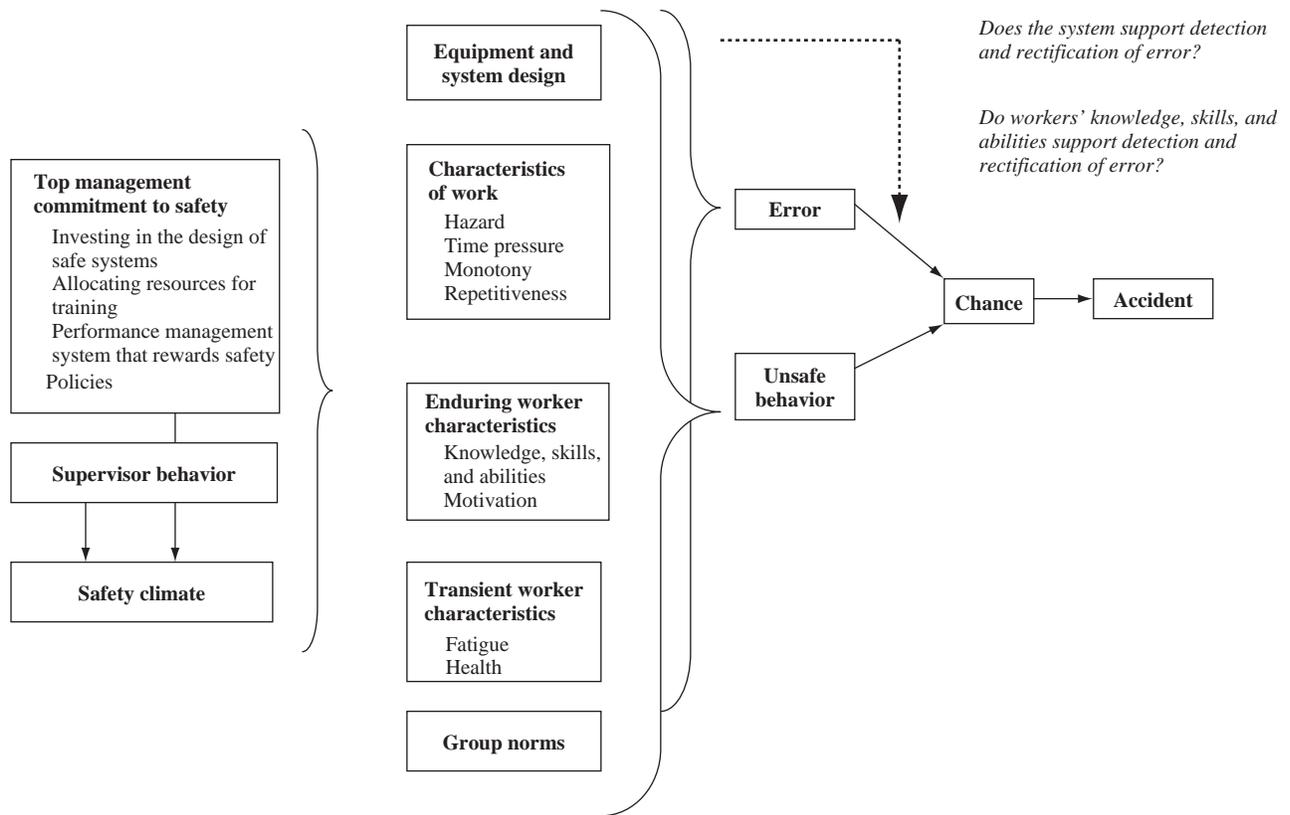


FIGURE 1 Factors contributing to errors and accidents. Adapted from Sanders and McCormick (1993).

domain-specific taxonomies that may be used as frameworks to investigate the errors preceding accidents.

In 1990, Reason presented a widely used definition of error that conceptualizes error as “a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome and when these failures cannot be attributed to the intervention of some chance agency.” The study of error has evolved to consist largely of highly domain-specific research that leads to highly domain-specific taxonomies. A large number of taxonomies with both overlapping and domain-specific features are available. This section reviews the taxonomies and approaches of Norman, Rasmussen, and Reason as well as the work of Zapf and colleagues. These models have been selected not only on the grounds of their prominence in the literature and their perceived application in a number of workplace situations but also because they are relatively distinct.

In 1981, Norman generated a three-category taxonomy of errors: slips, mistakes, and violations. A slip is a good plan that is poorly implemented, a mistake is an

incorrect intention, and a violation is deliberate (for whatever reason). In the case of the train driver, a slip would be the driver attempting to stop the train but pressing the wrong lever, a mistake would be the driver not recognizing the signal as being set at danger, and a violation would be the driver deciding to ignore the signal (perhaps thinking the signal was faulty). It can be seen immediately that the framework assists in tracing the source of the error and helps the organization to construct a remedy.

In 1983, Rasmussen developed a framework that approaches errors by looking at the cognitive processes behind task performance and, therefore, the source of error. His cognitive control taxonomy describes three levels of cognition involved in task performance: skill, rule, and knowledge. Task performance ranges from sensorimotor, automatically performed actions to actions that are conscious and based on directed deliberation. Behavior is skill based when stored patterns of information are automatically recalled from past experience and enacted. An experienced typist moving from a German-language keyboard to an English-language

keyboard—where the “y” and “z” keys are interchanged—is likely to make skill-based typos. Rule-based actions are directed by rules copied from procedures that are stored in memory in the “if (condition) then (action)” form. A typical error that would be considered rule based is when an individual misinterprets the situation and chooses the wrong rule. The key difference between skill-based and rule-based behaviors is the extent to which conscious attention is directed at the task; skills are automatic, whereas rules are applied mindfully. In entirely unfamiliar situations where skills and rules do not apply, cognition moves to the knowledge level at which operation is based on one’s study of the situation and the goal of the individual.

The cognitive control model has its origins in field research carried out in many domains, including the control rooms of nuclear power plants, and is a general model of cognition applicable in these domains. Therefore, the train driver not pushing the brakes hard enough would be a skill error, whereas the train driver not realizing that the signal (e.g., a red light) indicated danger would be a rule error. In the case where the driver had insufficient or inappropriate training, and so was uncertain whether or not to stop, a knowledge-based error would have occurred. In the example of the Zeebrugge ferry disaster, it would appear that errors occurred at the rule level (i.e., the rules were ambiguous). But when it was noticed that the bow doors were open, the personnel concerned also made an error at the knowledge level (i.e., they did not understand the gravity of the situation and take action to close the doors or alert the bridge).

In a further development, Reason combined his model with that of Rasmussen, proposing that an error classified as a mistake could be either rule based (i.e., using the wrong rule) or knowledge based (i.e., not knowing what the rule was) and providing a finer definition of the source of error.

In 1992, Zapf and colleagues developed an error taxonomy that also considers the different steps of the action process. An action encompasses the steps of developing a goal, collecting information on the environment (e.g., via attending signals), developing a plan and executing it, monitoring its execution, and processing feedback on the progress toward or achievement of the goal. Different levels of regulations control the action process. Similar to Rasmussen’s model, levels of regulation range from skill-based regulation producing automatized behaviors to intellectual, knowledge-based actions. A fourth level of regulation represents the use of heuristics. This taxonomy crosses the steps of the action sequence with the four levels of regulation. This results in error categories that help to

track the source of an error in terms of failure of memory, wrong movements, poor judgments, or other factors.

### **3. STRATEGIES OF ERROR PREVENTION**

Selection, training, and design are the generic strategies used to prevent errors. Although selecting individuals with the relevant skills and abilities is certainly important, it is limited where jobs change and, hence, workers are subjected to new demands on their skills. Therefore, training is used to complement the former strategy. Finally, system design strategies that recognize human factors (e.g., the limits of memory capacity and of perceptual speed) can reduce the likelihood of errors. System design approaches range from attempting to make it impossible to commit an error (e.g., the train driver could not override a red signal because electricity for the train would be automatically cut when a signal was set to red) to making the system fail-safe. The latter relates to reducing the consequences of an error. In the case of the train driver, the error consequence could be a crash with another train. This consequence could be reduced by brakes being automatically activated when another object is detected on the track as well as by installing huge bumpers to buffer the hit.

The idea of fail-safe systems was taken further by Reason, who took the pragmatic view that in practice it is impossible to eliminate error completely. Therefore, the role of the organization is to protect against the consequences of error. In 1990, Reason conceptualized protection against error as a series of defenses. When all defenses fail in a particular way, an accident occurs. This is termed the “Swiss cheese” model because it conceptualizes errors as being like holes in a piece of Swiss cheese, whereby an accident occurs when all of the holes line up. This model is not of particular value to the analysis of incidents but is certainly the way in which many practitioners would consider the analysis of systems protecting against error and checking whether the “holes” in the “cheese” are not generally aligned. In the case of the train driver, one might consider whether the introduction of a system in which the driver could not override the signal might provide additional protection. One also might consider whether additional protection could be provided by adding further signals in particular stretches of track so that if one signal was passed, other signals might be used as a further indication to the driver.

Because it is impossible to design systems in which errors are entirely impossible, it is important that systems be transparent and allow the operator or user to intervene. System transparency helps to quickly detect an error so that it can be corrected before the negative error consequences unfold fully. Furthermore, the system needs to allow the operator to intervene so as to make rectification possible.

The development of computer programs for office work during the past two decades provides good examples of how the strategies have been implemented. Many programs that are typically used for office work follow the “what you see is what you get” strategy, meaning that whatever the user does is reflected in changes on the monitor (cf. Fig. 1). The systems have become more fail-safe. When the user does a potentially critical step (e.g., deleting a file), the program will ask for verification of the step. The program will also allow errors to be rectified given that “undo” options are designed into the systems.

After an accident, the organization frequently seeks to trace an error to a human source and then punish the individual. The use of the models described in this section will assist in addressing the underlying organizational causes of accidents by informing the choice of prevention strategy. This potentially leads to more lasting solutions.

#### **4. ANTECEDENTS OF ACCIDENTS: THE SEARCH FOR INDIVIDUAL DIFFERENCE VARIABLES**

An accident is defined as an unexpected event that results in injury to one or several persons and/or damage to a system and that potentially impairs the accomplishment of the system's goals. One influential early approach to occupational safety goes back to the presumed discovery of the accident-prone person. The observation that a small minority of workers seemed to suffer the majority of accidents led to the conclusion that some individuals are more likely to have accident than are others. This thinking, which located the cause of an accident within the individual, was very popular for some time. Apart from its intuitive appeal, it was attractive to organizations because it justified a shift away from designing safety into the work environment to selecting the right people for the job—a less costly accident prevention strategy. Currently, however, most researchers agree that there is no convincing evidence of accident proneness. Data showing that some individuals have more accidents

than do others are explained in terms of differences in the degree to which individuals are exposed to hazards in the first place. Furthermore, wherever the law of probability operates, some individuals will have more than one accident by chance.

Related research has tried to identify personality variables that would predict the occurrence of accidents. Two recent pieces of work, a 1998 review by Lawton and Parker and a 2002 meta-analysis by Salgado, integrated the numerous individual pieces of research but could not produce a conclusive picture. Furthermore, this field is dominated by cross-sectional approaches in which many studies use companies' accident records and relate these to current expression of individual difference variables. Causal conclusions are likely to be flawed because the magnitude of some individual difference variables will reflect having had accidents rather than predicting accidents.

The only result on person variables that emerges consistently across different industries is the association between accidents and age. Younger people appear to have more accidents (an effect well known from vehicle driver accidents), and this proneness decreases with age. However, some findings indicate another increase in accidents after 50 years of age. In addition, accident severity, as reflected in degree of permanent disability or fatal accidents, seems to increase with age.

Despite the inconclusive result regarding the role of personality in accidents, this approach should not be dismissed entirely. First, it is possible that individual difference variables that contribute to successful coping with a workplace hazard differ from job to job. Whereas a high resistance to monotony might be important for some occupations, a high reaction time might be critical for others. Second, focusing more on personality variables that are related to the behaviors relevant for safety might be fruitful. For example, a study in the chemical industry found a relationship between psychological maladjustment and accidents. The extent to which individuals adhere to safety rules might be the link between the personality variable (maladjustment) and accidents. Finally, patterns or combinations of specific individual difference variables could deliver new insights.

#### **5. FROM ENGINEERING CONTROL SOLUTIONS TO SAFETY BEHAVIOR**

The engineering approach to safety represents an alternative to the previously described person-centered

perspective. The engineering approach seeks to design safety into the work system by acknowledging human factor variables. Numerous technical improvements have increased the safety of the physical working environment, tools, equipment, and machines. The effectiveness of the engineering approach is limited where it depends on whether the individual decides to make use of it. This is related to the three levels of prevention employed in this approach. The first level of prevention is to design a system such that the individual cannot engage in an action that could lead to an accident. For example, for some operations that involve the use of hazardous chemicals in the chemical industry, it is technically possible to design the system such that the worker cannot ever come into direct contact with the chemicals. In other words, the hazard has been designed out of the system. If this is not possible, engineering safeguards target the second level of prevention. This is done, for example, by installing a safety screen in front of a galvanic bath.

Safeguards often do not deliver the expected result. In 1982, Wilde proposed one potential reason for this in the risk homeostasis theory. This theory holds that whatever a person does, each individual holds a specific level of risk that he or she finds subjectively acceptable. While engaged in a given task, individuals monitor the level of risk they feel exposed to and regulate their behavior so as to keep it at an acceptable level. If the perceived level of risk is lowered through adding a safety mechanism, more risky behaviors will be displayed until the targeted risk level is reached (provided that the risky behavior is attractive). Originally developed in the context of traffic accidents, the theory tried to account for risky driving. In the framework of this theory, driving at a specific speed is associated with a specific probability (i.e., risk) of injury in case of an accident. Reducing this risk level through fitting the car with air bags will increase the speed of driving. In an industrial context, this theory explains why fitting a safety mechanism (e.g., installing a safety screen in front of a galvanic bath) possibly increases hazardous behavior (e.g., being less careful to avoid splashes, not wearing protective goggles). Finally, the third level of prevention is used when the means of design and safeguards are exhausted. Then, warning systems such as signs and symbols are installed.

As pointed out previously, whenever systems use the second and third levels of prevention, safety depends on the extent to which workers work with or alongside the safeguards and safety procedures. This

understanding broadens the focus of safety measures from a pure engineering perspective to look at worker behaviors. The majority of behavior-oriented research has focused on compliance to safety rules such as whether workers use the proper work strategies, and use tools properly. In 2002, Burke and colleagues developed a broader taxonomy of generic safety performance with the goal of capturing categories of safety behaviors that cut across different jobs. This research, carried out in the nuclear waste industry, defined general safety performance as “the actions or behaviors that individuals exhibit in almost all jobs to promote the health and safety of workers, clients, the public, and the environment.” Four categories of safety performance were identified: (a) using personal protective equipment to shield individuals from hazards, (b) engaging in work practices to reduce risk (e.g., adhering to safety procedures in terms of risk-reducing work practices), (c) communicating health and safety information (e.g., a hazard, an accident, a related incident) to the appropriate personnel, and (d) exercising employee rights and responsibilities. Although previous research that looked at worker behaviors typically emphasized worker safety compliance (as captured in the first two categories), the third and fourth categories identified in this research clearly went beyond this approach. A fruitful addition to this taxonomy, developed by Griffin and Neal in 2000, is the extent to which workers participate voluntarily in safety activities or attend safety meetings. It is assumed that these behaviors, called safety participation, contribute indirectly to workplace safety by supporting the development of a safe environment.

## **6. RECENT APPROACHES TO SAFETY: THE ROLE OF MANAGEMENT AND ORGANIZATIONAL CLIMATE**

In many industrial contexts, unsafe behaviors are inherently more rewarding than safe behaviors. They frequently allow faster and more comfortable accomplishment of work, and they are often more socially acceptable (e.g., the display of safe behavior may be regarded as “unmanly”). Accidents or injuries occur too rarely to represent a sufficient reinforcement for displaying safe behaviors. So, what induces individuals to act in a safe way despite the attractiveness of unsafe behaviors? Characteristics of individuals (e.g., their

safety knowledge) will enhance these behaviors. Of similar importance is a factor that pertains to the environment in which workers operate, that is, the organization's safety climate.

Organizations develop policies, procedures, and practices concerning occupational safety. Organizational climate has been defined as employees' perception of these policies, procedures, and practices. For example, safety climate captures the level of priority accorded to occupational safety, and this is assumed to influence individuals' safety behavior. It is presumed that safety climate is shaped by supervisors as well as by top management (cf. Fig. 1).

Supervisors, especially line managers, influence an organization's safety climate through several mechanisms. First, line managers serve as role models for their staff. In 2000, Maierhofer and colleagues found line managers' use of personal protection equipment to be related to the frequency with which employees use it. Second, also in 2000, Zohar demonstrated that supervisors' tendency to respond to staff safety violations, to react positively to safe behaviors, and to show that they take safety seriously is important to bringing rules and regulations to life. This supervisor behavior also proved to be related to microaccidents. In 1999, Hofmann and Morgeson revealed that the more comfortable employees felt in raising safety concerns and discussing safety issues with their supervisor, the lower the accident rate. The wider social environment existing with coworkers also has some potential to control unsafe behaviors and its consequences. For example, in 1996, Hofmann and Stetzer found, at the group level, that the more inclined individuals in a group were to approach someone about unsafe behaviors, the lower the accident rate.

The Zeebrugge ferry incident showed how apparently minor examples of management behavior (e.g., refusal to install a warning light on the bridge of the ship) can combine with other policies (e.g., negative reporting only) in the antecedents of an accident. These illustrations are indicative of poor safety climates.

The meaning of supervisor behavior becomes more significant when workers are confronted with high performance pressure. Under time pressure, shortcut work methods are more readily adopted to save some time—often at the expense of safety—and supervisor behavior plays a critical role in the alignment of these competing goals of working safely and reaching performance targets. Zohar's study on microaccidents showed that although there was a general trend for

microaccidents to increase with growing time pressure, this effect was reduced when the supervisor reinforced safe behaviors and sanctioned unsafe behaviors.

In addition to direct supervisors, top management influences organizational safety by shaping the safety climate and, in turn, workers' safety behaviors. Top management's commitment to safety issues, the priority top management places on safety that shows through allocation of resources and investments (for tools, health, and safety training), and the support that top management gives to direct supervisors all shape the safety climate. Shannon and colleagues' 1997 overview of studies on injury rates showed that injury rates were consistently lower in organizations where top management played an active role in health and safety issues.

## 7. SAFETY-ENHANCING INTERVENTIONS

Interventions aim at enhancing safety using one of three pathways: changing characteristics of the workforce (e.g., increasing safety behavior), changing the work itself (e.g., improving the design of tools), or altering elements of the work context (e.g., changing reporting principles).

The interventions employed most frequently seek to influence characteristics of workers, mainly to enhance safety behaviors. In 2003, Burke and Sarpy suggested that safety behavior could be seen as a function of knowledge, skills, and abilities (KSAs) as well as motivation and values (cf. Fig. 1). Safety training, based on diverse training methods such as lectures, demonstrations, and hands-on simulations, have been employed successfully in providing staff with the relevant KSAs and in motivating staff to use the skills. Research on training effectiveness outside the domain of safety has identified numerous factors that influence the effectiveness of training, and many of these can be applied to safety training as well.

Other interventions seek to specifically enhance the motivation to display the trained behavior using behavioral principles. One element used by a variety of intervention strategies is the use of feedback that informs workers about their level of safety behavior or number of accidents. An intervention study in the vehicle maintenance division by Komaki and colleagues in 1980 showed that training was somewhat useful in preventing accidents but that a much stronger effect, in terms of reduced accident rates, was achieved

when the training was accompanied by regular feedback to workers on safety-related behavior. Related interventions have successfully merged training with goal setting or used incentives and competition between units—always combined with feedback.

The second type of intervention, changing work itself, is employed whenever engineering control mechanisms are used. Furthermore, changing the pace of work to make safe behaviors less “costly” for workers may be important. Finally, changing the work context can improve safety. For example, in 2001, Kaminski showed that rewarding principles are related to injuries; that is, the use of performance-related pay, an important and successful principle of performance management, was significantly related to higher levels of injuries.

Finally, a more encompassing intervention method would target higher levels of an organization’s hierarchy. Changing line, middle, or top managers’ approaches to safety can improve the organization’s safety climate, and this in turn can enhance employees’ safety behaviors. In 2002, Zohar reported such an intervention. The intervention sought to increase the relative frequency with which line managers spoke to their staff about safety behaviors versus production-oriented issues. Increasing the proportion of safety-oriented interaction with a feedback method made safety an explicit performance goal. As a result, employees’ safety behaviors improved, their perception of the safety climate increased, and the injury rate was reduced successfully.

## 8. SUMMARY

This article has traced accidents as having antecedents in the twin areas of error and safety. By understanding these, it is possible to trace causes of accidents and take judicious steps to avoid their recurrence. Combining the engineering approach with the various strategies to increase the display of safety behaviors would seem to be fruitful.

### See Also the Following Articles

Environmental Versus Individual Risk Taking: Perception, Decision, Behavior ■ Work Safety

### Further Reading

Barling, J., & Frone, M. (2004). *The psychology of workplace safety*. Washington, DC: American Psychological Association.

- Burke, M. J., & Sarpy, S. A. (2003). Improving worker safety and health through interventions. In D. A. Hofmann, & L. E. Tetrick (Eds.), *Health and safety in organizations* (pp. 56–90). San Francisco: Jossey-Bass.
- Burke, M. J., Sarpy, S. A., Tesluk, P. E., & Smith-Crowe, K. (2002). General safety performance: A test of a grounded theoretical model. *Personnel Psychology*, 55, 429–457.
- Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach. In H. C. Triandis, M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 4, pp. 271–340). Palo Alto, CA: Consulting Psychologists Press.
- Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology*, 5, 347–358.
- Hofmann, D. A., & Morgeson, F. P. (1999). Safety-related behavior as a social exchange: The role of perceived organizational support and leader–member exchange. *Journal of Applied Psychology*, 84, 286–296.
- Hofmann, D. A., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviors and accidents. *Personnel Psychology*, 49, 307–339.
- Kaminski, M. (2001). Unintended consequences: Organizational practices and their impact on workplace safety and productivity. *Journal of Occupational Health Psychology*, 6, 127–138.
- Komaki, J. L. (1998). *Leadership from an operant perspective*. New York: Routledge.
- Komaki, J., Heinzmann, A. T., & Lawson, L. (1980). Effect of training and feedback: Component analysis of a behavioral safety program. *Journal of Applied Psychology*, 65, 261–270.
- Laflamme, L., & Menckel, E. (1995). Aging and occupational accidents: A review of the literature of the last three decades. *Safety Science*, 21, 145–161.
- Ludwig, T. D., & Geller, E. S. (1997). Assigned versus participative goal setting and response generalization: Managing injury control among professional pizza deliverers. *Journal of Applied Psychology*, 82, 253–261.
- Maierhofer, N. I., Griffin, M. A., & Sheehan, M. (2000). Linking manager values and behavior with employee values and behavior: A study of values and safety in the hairdressing industry. *Journal of Occupational Health Psychology*, 5, 417–427.
- Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, 88, 1–55.
- Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols; and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 257–266.
- Reason, J. (1990). *Human error*. Cambridge, UK: Cambridge University Press.
- Reason, J. A. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate.

- Reber, R. A., Wallin, J. A., & Chhokar, J. S. (1990). Improving safety performance with goal setting and feedback. *Human Performance*, 3, 51–61.
- Salgado, J. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10, 117–125.
- Salvendy, G. (1997). *Handbook of human factors and ergonomics*. New York: John Wiley.
- Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering and design* (7th ed.). New York: McGraw-Hill.
- Shannon, H. S., Mayr, J., & Haines, T. (1997). Overview of the relationship between organizational and workplace factors and injury rates. *Safety Science*, 26, 201–217.
- Wilde, G. J. S. (1982). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis*, 2, 209–258.
- Zapf, D., Brodbeck, F. C., Frese, M., Peters, H., & Pruemper, J. (1992). Errors in working with office computers: A first validation of a taxonomy for observed errors in a field setting. *International Journal of Human-Computer Interaction*, 4, 311–339.
- Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology*, 85, 587–596.
- Zohar, D. (2002). Modifying supervisory practices to improve subunit safety: A leadership-based intervention model. *Journal of Applied Psychology*, 87, 156–163.



# School–Community Partnerships

Richard J. Cowan

Kent State University, Kent, Ohio, USA

Susan M. Swearer

University of Nebraska, Lincoln, Nebraska, USA

1. School–Community Partnerships Defined
  2. Domains of School–Community Partnerships
  3. Approaches to Fostering School–Community Partnerships
  4. Conclusion
- Further Reading

## GLOSSARY

**collaboration** Working together toward a common goal or set of goals.

**community-based organization** Any agency, organization, or system (other than a school) that exists in the community and has the potential to partner with a school (e.g., public library, local law enforcement agency, community mental health clinic, local family medical clinic) to meet the needs of children, families, and the community as a whole.

**educators** General education instructors, special education instructors, teaching aides, classroom volunteers, administrators, support staff, school psychologists, and any other persons within a school setting that contributes to the education of students within that environment (e.g., lunchroom aides).

**parent** The adult who is responsible for the care and well-being of a child (e.g., biological parent, caregiver, legal guardian).

**partnership** A coequal interdependent relationship that is established and developed over time with the primary purpose of working together toward a mutually determined set of goals and objectives.

**problem solving** Process of identifying a priority need or problem, determining outcome goals and objectives, assessing what might be causing and/or maintaining the problem (or blocking access to a desired resource), determining a

solution to target potential causal mechanisms, implementing the intervention or strategy to meet the need or solve the problem, evaluating outcomes with regard to predetermined goals and objectives, and determining the need for continued intervention; this process may be recycled as needed based on the assessment of coparticipants.

**system** A particular environment with an emphasis on specific components of the environment, including people, physical objects, placement of people and objects, physical arrangement, and interactions among these components (e.g., home system, school system).

Partnership may be defined as a coequal interdependent relationship that is established and developed over time with the primary purpose of working together toward a mutually determined set of goals and objectives. In the case of school–community partnerships, the targeted goals and objectives are related to the needs of students and the community as a whole. Such partnerships use individuals and resources from the school, together with those from the community, to serve both the educational system and the community. School–community partnerships are capable of addressing multiple domains that include, but are not limited to, family and student health, social and mental health programs, violence prevention, tolerance and diversity, government legislation/regulation and law enforcement, community volunteers and the faith community, business–community partnerships, and school/academic academic readiness and educational support.

## **1. SCHOOL-COMMUNITY PARTNERSHIPS DEFINED**

### **1.1. Partnerships Defined**

A common definition of the term “partnership” is that of a coequal interdependent relationship that is established and developed over time with the primary purpose of working together toward a mutually determined set of goals and objectives. As applied to relationships between schools and communities, a partnership generally refers to a relationship occurring between educators (e.g., teachers, school psychologists, support staff) and parties from noneducational settings (e.g., parent and advocacy groups, mental health care workers from community wellness centers, doctors from local clinics and hospitals, law enforcement officers), with an emphasis on goals and objectives related to the educational and social development and the well-being of students for whom all members of a particular community share interest and responsibility. In addition, school-community partnerships are formed to address the needs of the community as a whole. One could easily argue that because students ultimately grow up and become members of the community, school-community partnerships that target the needs of students within the school serve the interests of both developing students and the future of the community. Establishing and developing a partnership in which participants share ideas, information, resources, skills, and expertise is a process that requires care and cultivation. Partnerships do not result from one or two encounters between educators and non-educators within the community; rather, they develop over time through multiple collaborative interactions. The process of building and developing partnerships requires a guiding philosophy. Approaches to fostering school-community partnerships are discussed later in this article, following a discussion of characteristics of, parties served by, and domains for activities related to school-community partnerships.

### **1.2. Characteristics of School-Community Partnerships**

In 2002, Sheridan and colleagues offered the following characteristics of meaningful school-community partnerships:

1. Interactions among partners are collaborative.
2. Maintenance of a positive relationship is a priority.
3. Relationships are balanced.

4. Relationships are cooperative and interdependent.
5. Relationships occur within a context with students at the center.
6. Actions between schools and other community agencies are flexible, responsive, and proactive.
7. Differences in viewpoints and perspectives are seen as strengths rather than hindrances.
8. There is a commitment to cultural diversity.
9. There is an emphasis on outcomes and goal attainment.

The first characteristic, collaborative interactions, may take place through collaborative problem solving whereby there is mutual determination of needs and goals in addition to joint ownership of solutions and outcomes. Positive relationships may be defined as relationships that are balanced, cooperative, interdependent, and equal in terms of decision-making power. Because schools and other community agencies serve multiple unique functions in a given community, it is important for partners to remain flexible, responsive, and proactive. It is nearly certain that differences in opinion regarding priorities and strategies will result during interactions among various individuals representing different agencies. However, if the partners embrace a coequal partnership approach, whereby each participant’s relative expertise is valued and equal decision-making power is granted, these differences are not viewed as hindrances; rather, they are viewed as opportunities to work through differences in a joint effort to meet the needs of students and families served by the community. School-community partners should consider the needs of all community members, including those individuals and families who come from culturally diverse backgrounds. Cultural diversity extends beyond race and ethnicity to include family attitudes, beliefs, values, and experiences with various schools and community-based organizations. Shared ownership and responsibility for the education, development, and well-being of all children allows for the whole to be greater than the sum of its parts and is derived from the work of individuals who are able to put aside their personal needs and goals in an effort to meet the needs and goals of the entire community.

### **1.3. Foci of School-Community Partnerships**

In 2002, Epstein and colleagues, who are leaders in the area of family-school-community partnerships within the field of school psychology, maintained that school-

community partnerships generally center on the needs of four identifiable targets: (a) student-centered services, (b) family-centered services, (c) school-centered services, and (d) community-centered services. Student-centered partnership activities may focus on the development of student awards, student incentives, scholarships, student-centered extracurricular activities (including travel), and scholastic and career mentoring. Family-centered partnership activities may focus on parent training and education, family-based recreational activities, adult educational courses (e.g., GED courses), parent rewards, parent counseling and consultation, and other contributions for parents. School-centered partnership activities may focus on gaining access to equipment and materials, beautification and repair of the school and/or community, teacher-related rewards and incentive programs, funds for school activities and events, and office and classroom supports. Finally, community-centered activities implemented through school-community partnerships may focus on community improvement and beautification, student-generated community-supported exhibits and performances, volunteer work, and other outreach activities.

## 2. DOMAINS OF SCHOOL-COMMUNITY PARTNERSHIPS

### 2.1. Ecological Theory

The fields of education and psychology have come to embrace a systems-based approach to understanding those variables that influence how an individual behaves, develops, and learns over time and across multiple settings. This approach is grounded in ecological theory. Based on the work of Bronfenbrenner, ecological theory is concerned with the interaction between an individual and various contextual systems. Systems include the microsystem, mesosystem, exosystem, and macrosystem. The microsystem is concerned with the individual and his or her immediate environment (e.g., a child interacting within a particular classroom), the mesosystem involves the interrelation between major systems in the individual's life (e.g., between home and school), the exosystem is concerned with environments not directly related to the individual yet influencing his or her life (e.g., a parent's workplace), and the macrosystem includes overall cultural or subcultural patterns and influences (e.g.,

policies, federal and state legislation, national and global economic factors). This ecobehavioral paradigm illustrates how systems influence and are influenced by one another, stressing that contextual systems do not exist in isolation; rather, interactions across systems occur frequently and are multifaceted and multidetermined over time. All system levels influence school-community partnerships, with multiple mesosystemic intersections occurring between the school and various community settings. For example, partnerships may exist between a school and a business within the community or between the school and a community mental health agency. These intersections between the school and various community settings involve multiple individuals, are influenced by multiple factors, and represent opportunities for both educational and noneducational partners to intervene with regard to multiple domains. Some of these domains are discussed next.

### 2.2. Specific Domains

Specific domains to consider for potential partnership activities include family and student health, social and mental health programs, violence prevention, tolerance and diversity, government legislation/regulation and law enforcement, community volunteers and the faith community, business-community partnerships, and school/academic readiness and educational support (Table I). These domains of school-community partnerships are not mutually exclusive.

#### 2.2.1. Family and Student Health

Because good health has implications for education and development, and because learning about wellness and a healthy lifestyle is a critical curriculum in and of itself, schools and other community-based

TABLE I  
Domains of School-Community Partnerships

- 
- Family and student health
  - Social and mental health programs
  - Violence prevention
  - Tolerance and diversity
  - Government legislation/regulation, and law enforcement
  - Community volunteers and the faith community
  - Business-community partnerships
  - School/academic readiness and educational support
-

organizations (e.g., health care clinics) are identifying opportunities for partnerships with the primary goal of wellness education (often as a preventive mechanism) in addition to medical intervention for prevention (e.g., flu shots, immunizations) and remediation (e.g., medical intervention) of medical problems. Bringing wellness education and medical services to the school makes sense given that making appointments is a challenge for many families and given that the school setting is where most children spend the majority of their waking hours. School-based nurses, counselors, and/or psychologists are in an ideal position to collaborate with medical health care partners to ensure treatment fidelity and monitor behaviors related to medical intervention that may be indicators of effectiveness (e.g., monitoring the effects of medication on the on-task performance of an individual diagnosed with attention deficit/hyperactivity disorder).

### **2.2.2. Social and Mental Health Programs**

The field of school psychology has experienced an enormous shift in the provision of services, expanding beyond the role of conducting psychoeducational assessment to include roles such as consulting with parents and educators, being involved in program development and evaluation, providing small group interventions for students with special needs (e.g., social skills training), crisis intervention, and working with community-based organizations to form partnerships in the best interest of individuals and families served by the school and community. School psychologists strive to expand their role as practitioners in schools and communities to provide a wider range of mental health services to individuals and families within the school. This could include direct services (e.g., therapy, one-on-one counseling), indirect services (e.g., behavioral consultation with parents and educators, family-based consultation services), and group-based services (e.g., anger management groups, social skills training groups, mental wellness training programs).

School psychologists, school counselors, and social workers may be in an ideal position to develop, implement, and evaluate mental health services. As with the provision of medical services within the school setting, providing mental health services within the school has implications for the education and development of students and makes common sense given that school-aged children spend a significant proportion of their waking hours within the school setting.

Providing a wider array of mental health services within the school may require additional training of some school-based personnel. This may be promoted in part through partnerships with professionals from the community (e.g., physicians, licensed psychologists, mental health care providers) whereby experts provide training and professional development in specific domains of mental health services, including wellness training and preventive mental health care practices.

### **2.2.3. Violence Prevention**

Given the recent national focus on violence prevention and intervention among students, schools have been encouraged, or even mandated, to implement programs that tackle the problem of violence in schools. Violence occurs on a continuum ranging from bullying to more extreme behaviors such as assault and homicide. Although most schools have adopted a “zero tolerance” policy toward more severe forms of violence, schools are also extending this hardline approach to bullying. In fact, as Limber and Small noted in 2003, there are currently 15 states that have passed laws to address bullying in schools. These laws have set the stage for parents to pursue litigation if they believe that their children are being bullied at school.

School-community partnerships are vital for reducing violence among and toward students. In fact, many schools have law enforcement officers who help to ensure the safety of students in and around the school buildings. Partnering with local law enforcement is an important component in helping to reduce violence in schools. In addition, it is clear from the research that violence prevention is a whole school and community responsibility. Violence prevention programs must have buy-in from students, teachers, administrators, parents, and adults in the community. When school-community partnerships are forged, violence prevention efforts can be successful.

### **2.2.4. Tolerance and Diversity**

The educational system in the United States continues to experience an increase in culturally diverse students and families. Some schools that have experienced an increase in culturally diverse students and families have found it helpful to seek partnerships with community-based organizations to help families acclimate to the school system and community that they are entering.

For example, there are community-based organizations whose primary mission is to provide English as a Second Language (ESL) or English as a Learned Language (ELL) programs for students and families as well as to provide transition supports to families and schools as they learn to mutually accommodate one another's attitudes, beliefs, and values.

In addition to seeking partnerships with community-based organizations that help families in transition become a part of the educational and community systems, schools may seek interactions with organizations that serve particular minority groups in an effort to expose students to new cultures and diversity. This latter type of partnership affords school-based learning opportunities for students to interface with individuals with cultural traditions that are different from their own in hopes of instilling greater acceptance and tolerance. Such partnerships may also allow educators to gain insight into a family's attitudes, beliefs, and values regarding education and citizenship.

### **2.2.5. Government Legislation/Regulation and Law Enforcement**

The U.S. educational system is replete with state and federal legislation that guides the provision and implementation of educational services and programming. One implication of such legislation is that the school is responsible for providing a free and appropriate public education (FAPE) to all individuals, including those with specific learning, medical, and/or psychological disabilities. Once an individual is identified with a specific disability, the educational team must determine specific learning goals and objectives related to the needs of that individual. Once these goals and objectives are determined and specific accommodations are specified, the school must provide a meaningful placement for the student to learn. The school is legally required to provide services in the least restrictive environment (LRE) possible, meaning that whenever possible and reasonable, the student should be included in the general education setting with peer models to support his or her learning. In addition, once a student meets eligibility for special education services, the school is responsible for providing services until the student is 21 years of age. Some high school-aged students will require accommodations and placements that foster the development of independent living and vocational skills. Because school settings might be limited in terms of vocational training opportunities, schools often seek partnerships with community-based organizations to

place students in industry-specific training programs. Ideally, the school determines each student's relative strengths and needs, conducts a thorough assessment of abilities and interests, and seeks to match these variables with specific job training sites. This type of partnership not only addresses the needs of individual students and families but also helps contribute to future workers who are needed in community-based employment programs.

There is a long-standing tradition of partnerships between schools and community-based law enforcement agencies whereby police officers come into schools to talk about crime and drug abuse prevention. In many cases, local law enforcement agencies provide curricular materials to supplement their training programs beyond 1- or 2-day presentations. These types of programs generally involve the use of multimedia presentations and/or live skits to demonstrate important safety-related tips. At the middle school and high school levels, it is common to have law enforcement officers placed within a school building to ensure that students abide by the law as well as to provide opportunities for police officers to teach classes geared toward recruiting high school seniors into law enforcement-related occupations.

### **2.2.6. Community Volunteers and the Faith Community**

Because most communities have unmet needs (e.g., streets and parks that need to be cleaned, elderly individuals who might not have many social opportunities, citizens who cannot afford clothing and/or food), and because many view the educational process as an opportunity to instill ethical and character-building beliefs and practices, fostering volunteering opportunities for students within local faith-based and/or charitable organizations represents a unique opportunity for students to learn the value of making contributions to the community while at the same time seeing that a need of the community is met, at least for a brief time. This type of partnership will vary greatly across schools and regions of the country, being heavily dependent on the predominant faith of a particular community and/or whether the school is private or public. However, a common theme of school-volunteer agency partnerships is that students make a contribution to a local organization that is based on volunteering and providing services to the community while at the same time benefiting from such an experience.

### **2.2.7. Business–Community Partnerships**

Because major companies need tax breaks, and because making contributions to schools allows for an investment in the future of the local community, school–business partnerships have the potential to be a “win–win” situation. The range of activities may vary greatly; however, there is a common theme whereby businesses “adopt-a-school” as an “educational partner.” At this point, the business may make financial contributions, offer training opportunities and/or paid jobs to select students, contribute materials and equipment that may be business related (e.g., computers, calculators), and/or offer to pay employees to volunteer within the school setting. This type of partnership provides materials, training, and support to the school while at the same time building a potential pool of future employees and/or customers for the business partner.

### **2.2.8. School/Academic Readiness and Educational Support**

An important aspect of academic readiness is the ability to read. In addition to partnering with parents to foster regular reading in the home learning environment, schools frequently seek support from local community agencies such as libraries and teacher training programs within local colleges and universities with an emphasis on securing materials and resources to help provide supplemental reading programs. There are many opportunities for schools to partner with other community-based organizations to help enhance and promote primary curricular skills and activities outside of the classroom. For example, schools may partner with local city and county library systems to encourage and foster summer reading programs, including incentives for students and families who participate in active reading throughout the out-of-school months. Another means of enhancing reading skills in students is to seek opportunities to include teachers-in-training from local and regional colleges and universities in reading enhancement programs. This form of partnership allows teachers-in-training to learn about classroom management and literacy curricula while at the same time providing an invaluable service to students in the community. Reading is not the only subject area that may be targeted in school–community partnerships emphasizing academic skills support. For example, schools may partner with local natural history museums and/or planetariums to provide science-based extracurricular activities with incentives for child and family participation outside of the school setting.

## **3. APPROACHES TO FOSTERING SCHOOL–COMMUNITY PARTNERSHIPS**

### **3.1. Guiding Principles**

In 2001, Christenson and Sheridan maintained that partnerships are derived from the careful consideration of four primary components: approach, attitudes, atmosphere, and actions. Although this particular partnership model was developed to describe home–school partnerships, the core components of these authors’ philosophy are easily translated into a philosophy that may guide school–community partnerships. Approach is a means of interacting with another party that embraces the value of both parties with regard to the educational and social development and well-being of students for whom all participants share interest and responsibility. This component is concerned with communicating to the other party that its members possess unique knowledge and skills that make them invaluable partners. Attitudes are the perceptions that participants in partnerships have of one another. This is concerned with identifying strengths in partners, focusing on these relative strengths, and assuming a joint responsibility for the education of students within a community. Atmosphere is the overall climate set for participants engaged in a partnership. Finally, actions are strategies—not activities—intended to build long-term relationships through actively involving all partners in the educational process. This philosophy offers specific components for partnership facilitators to consider when working with individuals across various systems who possess the intent of establishing and developing long-term collaborative partnerships.

### **3.2. Roles and Responsibilities**

Partnerships may include multiple individuals, including parents, educators (i.e., general education instructors, special education instructors, teaching aides, classroom volunteers, administrators, school psychologists, and other support staff), paraprofessionals, community partners (e.g., medical professionals, mental health professionals, business owners), and students (especially high school students). When partnerships are grounded in relationships that are positive, balanced, cooperative, and interdependent, and that maintain equal power in decision making, participants recognize, appreciate, and welcome the unique relative expertise and skills

of each participant. Each individual's contribution to the partnership-based change process will vary depending on his or her unique set of strengths, skills, and expertise. For example, parents and other family members offer unique information about the home system from which each student is derived. This is important in understanding the relative contribution of each setting to the behavior and development of students. Parents may intervene to assimilate community values in the home. In addition to providing information about the school setting, educators offer much knowledge in terms of general learning principles, specific academic areas, and various curricula. Their participation may be critical to determine means of implementing interventions within the educational system. Depending on their background, paraprofessionals and community partners offer unique expertise in specific domains that fall outside of parental roles and educational programming *per se*. Their involvement can provide critical connections to various community-based systems and organizations. In addition, they may take on roles directly related to their practice (e.g., medical doctors offering health and wellness checkups, mental health professionals offering student and/or family therapy). Finally, older students make invaluable partners in determining a valid needs assessment and in determining specific means of proactive student involvement and/or response to adult-generated priorities and strategies. If long-lasting systems level change is to occur, it is critical to involve key players from multiple interconnected systems to assist with implementing specific strategies that will alter the attitudes, values, and beliefs of multiple community members.

### 3.3. Strategies for Change

Establishing and developing school-community partnerships between one or more parties requires that an individual or a small team takes the initiative in organizing and promoting cross-setting, multipurpose relationships between schools and other community-based organizations. Such planned facilitation allows for the process to unfold in a predetermined organized manner. Those individuals who guide the organizational process must take precautions against being too directive, thereby resulting in an imbalance of power across partners. The initial step in the process is to develop a working relationship between educators and community-based non-educators. During this phase, it is critical to strive for relationships that are positive, balanced, interdependent,

and based on coequal status regarding decision-making power. Seeking partners in the community is an ideal role for a school psychologist or administrator who possesses multiple connections in the community and with families served by the school. Initiating a long-term collaborative relationship requires a guiding philosophy. The "four A's" (i.e., approach, attitudes, atmosphere, and actions) derived from Christenson and Sheridan's model is one such philosophy. This includes recruiting various individuals from community-based settings to participate in an inviting environment. If relative expertise is recognized and valued, each member will make a unique contribution to the process. The contribution of the facilitator is expertise regarding how to establish and develop collaborative interactions across individuals and settings (i.e., knowing how to establish partnerships). This person can explain the process, clarify roles, and provide a collaborative open means of communicating and interacting with all participants. Once relationships are formed and the process is clarified and agreed on by all participants, the team may move forward with the team-based problem-solving process.

The team-based problem-solving process generally occurs through six predetermined phases: (a) problem identification, including needs assessment; (b) determination of outcome goals and objectives; (c) ecological assessment, including identification of resources and limitations; (d) intervention development; (e) intervention implementation; and (f) evaluation of outcomes (Table II). Although these phases are sequential, the process is dynamic and may be recycled as needed to meet the mutually determined goals and objectives of the project. The first step in this problem-solving process is to identify a priority need or problem. This step is commonly referred to as a school-community needs assessment. The team should include as many relevant participants from across settings as possible so as to increase the likelihood that multiple individuals and settings will commit to change. A needs assessment may occur through interviews, surveys, and/or group meetings. This step requires that participants put aside their own desires and needs in an effort to focus on specific needs or problems that affect the school and community as a whole. Once a need or problem is targeted, the team must mutually determine specific, observable, measurable outcome goals and objectives. Because progress toward meeting these goals and objectives will serve as a means of assessing overall success, the team must strive to be specific and reasonable. Identifying specific goals allows the team to determine how to measure progress toward goal attainment using

**TABLE II**  
**Phases of the School-Community Partnership**  
**Problem-Solving Process**

- 
1. Problem identification, including needs assessment
  2. Determination of outcome goals and objectives
  3. Ecological assessment, including identification of resources and limitations
  4. Intervention development
  5. Intervention implementation
  6. Evaluation of outcomes
- 

objective concrete units of analysis. Identifying reasonable goals helps to foster initial success and often involves first focusing on short-term, readily attainable goals and then expanding toward long-term, "broader picture" goals.

The next step is to conduct an ecological assessment to determine what might be causing and/or maintaining a problem or blocking access to a desired resource. This assessment involves examining the characteristics of the school, community, individuals, and families to determine resources they may contribute to an intervention. Resources might include individual expertise, financial contributions, materials, access to research and training, and/or existing interventions. Once the team has (a) made hypotheses regarding what might be causing or maintaining a problem and (b) identified resources available in the school and community, the team should mutually determine the type of intervention to be implemented. As the intervention is developed, it is critical to clarify the intervention across all participants and settings and to assign specific roles to ensure that the plan is carried out as intended. In addition to clarifying roles, it is helpful to write out the plan and make sure that everyone has a copy of the written product. It is often beneficial to assign a person(s) to monitor treatment fidelity throughout the intervention. This person can ensure that the treatment is being implemented as intended as well as offer support as needed. The final step in the process is to evaluate outcomes with regard to predetermined goals and objectives and to determine the need for continued intervention. At this point, the team evaluates the need or problem with regard to change over time, with the goals and objectives serving as "bars" for measuring relative change and/or overall success. If the initial goal is met, the team may decide how to adjust the existing intervention such that a larger goal may be met. If the initial goal is not met, the team may

decide to change the intervention and/or recycle through the problem-solving sequence.

#### 4. CONCLUSION

By branching out and fostering long-term relationships with pivotal community partners, schools can create a team oriented toward student-, family-, school-, and/or community-centered services. Establishing and developing a partnership in which participants share ideas, information, resources, skills, and expertise is a process that requires care and cultivation. Such relationships are attainable when all partners value equal decision making, approach each other with mutual respect, work toward the development of attitudes that highlight the strengths of all participants, develop a collaborative inviting atmosphere, and engage in actions that promote meaningful involvement of all participants. Partnerships do not result from one or two encounters between educators and noneducators within the community; rather, they develop over time through multiple collaborative interactions (e.g., the problem-solving process). To meet long-term, mutually determined goals and objectives, it might be necessary for the team to recycle through this change process until all parties are satisfied that the need has been met or the problem has been solved with regard to the mutually predetermined set of goals and objectives. Long-term partners are likely to engage in this problem-solving process multiple times to meet the needs of the school and community. School-community partnerships are critical given the multiple and ever-evolving needs of the schools and communities in which individuals function on a regular basis.

#### See Also the Following Articles

- Community Psychology ■ Consultation Processes in Schools
- Family and Culture ■ Home-School Collaboration
- School Violence Prevention

#### Further Reading

- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Christenson, S. L., & Sheridan, S. M. (2001). *Schools and families: Creating essential connections for learning*. New York: Guilford.
- Epstein, J. L., Sanders, M. G., Simon, B. S., Salinas, K. C., Jansorn, N. R., & Van Voorhis, F. L. (2002). *School,*

- family, and community partnerships: Your handbook for action* (2nd ed.). Thousand Oaks, CA: Corwin.
- Henegger, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. New York: Guilford.
- Howard, K. A., Flora, J., & Griffin, M. (1999). Violence prevention programs in schools: State of the science and implications for future research. *Applied and Preventive Psychology*, 8, 197–215.
- Larson, J., Smith, D. C., & Furlong, M. J. (2002). Best practices in school violence prevention. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 1081–1097). Bethesda, MD: National Association of School Psychologists.
- Limber, S. P., & Small, M. A. (2003). State laws and policies to address bullying in schools. *School Psychology Review*, 32, 445–455.
- Nelson, G., Amio, J. L., Prilleltensky, I., & Nickels, P. (2000). Partnerships for implementing school and community prevention programs. *Journal of Educational and Psychological Consultation*, 11, 121–145.
- Sheridan, S. M., Napolitano, S. A., & Swearer, S. M. (2002). Best practices in school–community partnerships. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 322–336). Bethesda, MD: National Association of School Psychologists.
- Welch, M., & Sheridan, S. M. (1995). *Educational partnerships: Serving students at risk*. Fort Worth, TX: Harcourt Brace.



# School Discipline and Behavior Management

George Bear

University of Delaware, Newark, Delaware, USA

1. Introduction
2. Definitions and Components of School Discipline
3. Developing Self-Discipline
4. Using Discipline to Correct Misbehavior
5. Preventing Misbehavior with Effective Classroom Management
6. Remediating Serious Misbehavior  
Further Reading

right from wrong, and inhibiting socially inappropriate behavior.

**zero tolerance approach** An authoritarian approach to school discipline in which students are punished harshly, automatically, and irrespective of the circumstances involved.

## GLOSSARY

**alternative education setting** A setting in which students are placed as a consequence of serious or chronic misbehavior. The setting is an alternative to the regular classroom and to expulsion.

**authoritative discipline** A style of discipline that is neither overly harsh nor overly permissive. Authoritative teachers establish clear rules and enforce them in a fair and consistent fashion, hold high behavioral expectations, demonstrate warmth and support, and foster self-discipline.

**character education** Strategies and techniques that develop thinking, emotions, and behaviors associated with important values and virtues, including self-discipline.

**classroom management** Practices, procedures, and routines used to establish and maintain a learning environment that fosters instruction and learning.

**self-discipline** Internal regulation of one's behavior. It entails assuming responsibility for one's actions, understanding

School discipline and classroom management have always been primary concerns among educators. This article addresses the four major components of effective school discipline and classroom management: developing self-discipline, correcting misbehavior, preventing misbehavior with effective classroom management, and remediating serious and chronic behavior problems.

## 1. INTRODUCTION

Since the onset of schooling, educators have been concerned about issues of school discipline and classroom management. Educators in Colonial America faced problems of fighting, noncompliance, truancy, and dropping-out—problems common in schools today. To be sure, the frequency and severity of discipline problems have changed over the years. In general, discipline problems have become more common and violent in nature, particularly since the late 1960s. Indeed, nearly every year since an annual Phi

Delta Kappa/Gallup Poll of the public's attitudes toward public schools was first conducted in the 1970s, the poll has revealed that the "lack of discipline/more control" is either the greatest problem or among the greatest problems facing U.S. schools.

## 2. DEFINITIONS AND COMPONENTS OF SCHOOL DISCIPLINE

Among the many definitions of discipline, two are most applicable to the context of schooling: (i) teaching that develops self-control and character and (ii) treatment that corrects or punishes (*Webster's New World College Dictionary*, 1996, p. 391). Both of these definitions reflect two traditional goals of U.S. educators: (i) to teach, or develop student self-control, character, or self-discipline, and (ii) to use discipline to correct, punish, or manage misbehavior.

Closely related to school discipline is classroom management. Classroom management refers to practices, procedures, and routines used to establish and maintain a learning environment that fosters instruction and learning. This consists of such tasks as organizing the classroom and materials; managing academic assignments; planning and delivering effective instruction; monitoring student behavior; and establishing rules, procedures, and routines. As is true with the development of self-discipline, classroom management serves a preventive function. That is, both help prevent many discipline problems from occurring in the future. They differ in focus, however. Whereas developing self-discipline focuses primarily on the competencies of students, preventing discipline problems with effective classroom management focuses primarily on the competencies of teachers. Developing self-discipline, using discipline to correct misbehavior, and preventing discipline problems with effective classroom management are important components of a comprehensive school discipline plan. Each of these components should target all students. A fourth component of comprehensive school discipline, remediating serious misbehavior, primarily concerns schoolwide discipline and focuses on the needs of the small percentage of students who typically require removal from the regular classroom and placement in another educational setting. These four components are the focus of this article.

## 3. DEVELOPING SELF-DISCIPLINE

Until relatively recently, developing self-discipline was the primary aim of U.S. education. The aim was to develop self-governing, socially responsible citizens who require the minimum of external control. Self-governance, or self-discipline, was viewed as necessary for the survival of a democratic society. For several centuries, religious education and moral education served this aim. These two approaches to developing self-discipline were replaced by character education in the early 20th century. Although their methods differed, each of these educational approaches was characterized by systematic and deliberate attempts to teach self-discipline. In nearly all schools, students were required to read value-laden literature, memorize scripts or codes that endorsed religious and moral principles or character traits, discuss issues of morality and problems of democracy, and participate in activities that promoted cooperation, helping, and democratic thinking.

In the 1960s and 1970s, character education faded precipitously. Although student protests and violence in society increased markedly during this period, few schools continued to take deliberate and systematic actions to develop self-discipline. Instead of attempting to develop character, many schools adopted the perspective of cultural relativism, believing that no set of values should be taught in the schools. Instead, it was widely held that students were to develop their own values, with little or no direction from teachers. This perspective, combined with increased concerns about the legal rights of students and parents (especially freedom of speech and the right to due process) and a rapidly growing focus on the teaching of math and science, contributed to the decline of character education.

In many schools, renewed efforts are being made to develop self-discipline. These schools recognize that promoting self-discipline, or responsible citizenship, reduces discipline problems in the classroom, promotes positive relations with others, fosters academic achievement, and enhances self-perceptions of self-worth. These positive outcomes benefit not just individuals and classrooms but also society in general.

Popular programs that promote self-discipline are commonly referred to as character education or social and emotional learning programs. These programs promote self-discipline by targeting a combination of social cognitions, emotions, and social skills that research has shown to be related to prosocial and antisocial behavior. These include empathy, social

perspective taking, attributions of responsibility, social problem solving, interpersonal understanding, moral reasoning, conflict resolution, anger management, and impulse control. In teaching and developing self-discipline, these programs rely heavily on the techniques of verbal instruction (e.g., direct teaching of social problem-solving steps), role playing, modeling, positive practice, positive reinforcement, and classroom discussion of moral and social problems. Such programs teach students how to think, feel, and act—a strategy that most authorities in school discipline believe is much more effective in preventing behavior problems than are strategies designed to teach students what to think and do, such as lecturing, moral appeal, and fear arousal.

Most programs of character education or social and emotional learning are of one of two general types or a combination thereof: a curriculum package or curriculum infusion. A curriculum package, commonly produced by a commercial or nonprofit publishing company, consists of lessons that are typically taught by the classroom teacher at least once a week. Some also include lessons, activities, or support materials for parents. Unfortunately, few of these programs have been evaluated and shown to be effective, especially in reducing or preventing actual misbehavior. There are exceptions, however. Thus, before adopting any program it is strongly recommended that a school review evidence of the program's effectiveness [the Collaborative for Academic, Social, and Emotional Learning has an excellent Web site ([www.casel.org](http://www.casel.org)) that includes reviews and research on popular social and emotional learning programs].

The second general type of program for developing self-discipline, curriculum infusion, entails integrating lessons and activities for promoting self-discipline throughout the curriculum as well as other school activities. This includes using techniques, such as role playing and class discussion of social problems and moral dilemmas, to foster social and emotional learning when covering many issues in social studies (e.g., constitutional rights, issues of slavery, and the Holocaust), literature (e.g., the moral dilemmas and interpersonal problems of characters in literature), and health (e.g., issues of smoking, alcohol use, and peer pressure). It also includes the use of creative writing, plays, school assemblies, and club activities.

Perhaps the most comprehensive and research-supported program for fostering social and emotional learning is the Child Development Project. This project employs each of the techniques discussed previously while emphasizing the active participation of students in classroom and schoolwide activities that call for

social decision making, moral reasoning, and prosocial behavior. The project also emphasizes what it calls developmental discipline as a method of correcting misbehavior and preventing future misbehavior by developing self-discipline. Although developmental discipline includes punitive consequences for misbehavior, it focuses on teaching students to understand and respect the perspectives of others, especially how their own misbehavior impacts the feelings of others. It also focuses on teaching students skills of social problem solving and conflict resolution.

In using developmental discipline, misbehavior is viewed as providing teachers with opportunities to help develop self-discipline. Research shows that schools that participate in the Child Development Project, especially those that employ systematically and consistently a combination of the previously discussed techniques, are effective in promoting social decision making, social problem solving, empathy, respect for the opinions of others, democratic values, and a positive school climate. Moreover, in developing self-discipline these schools reduce future alcohol use and delinquent acts (i.e., threatening others, possessing a weapon, stealing vehicles, and skipping school).

#### **4. USING DISCIPLINE TO CORRECT MISBEHAVIOR**

Educators have always used discipline to manage, control, and punish misbehavior. Fortunately, today's methods tend to be less harsh, at least physically, and more sensitive to students' rights than in earlier centuries. Whereas disobedient students were commonly whipped or spanked in the past, approximately 30 states now ban the use of corporal punishment in schools. Unfortunately, many schools have replaced corporal punishment with perhaps an equally harsh "zero tolerance" approach to discipline. A pervasive zero tolerance approach differs from reasonable zero tolerance policies in that the latter applies only to the most serious offenses that threaten the welfare and safety of others. A zero tolerance approach requires that misbehaving students are suspended, expelled, and/or turned over to the police for common crimes and misbehaviors that were typically handled by schools in the past, including fighting, stealing, and vandalism—the three most common criminal offenses in schools. Indeed, the number of suspensions, expulsions, and police reports has increased dramatically

during the past decade, with a greater proportion of African American students being removed from school. Such disproportionality raises important ethical and legal issues. The increase in a zero tolerance approach to school discipline has occurred despite the lack of evidence that this approach, especially when used as the primary means of managing and correcting misbehavior, is effective in developing self-discipline and replacement behaviors or in reducing the future occurrence of the behavior problem. Not only are these practices ineffective but also they often reinforce the very behaviors that led to the student's removal from school. For example, this happens when students who dislike school, teachers, or peers are sent home and, in the absence of parental management and supervision, they spend their time watching television, playing video games, roaming the streets, etc. instead of acquiring academic, vocational, and interpersonal skills.

In responding to more frequent, and less serious, behavior problems, such as a student talking or getting out of his or her seat without permission, educators tend to use punitive, or reductive, techniques that are less harsh than those of suspension and corporal punishment. Most popular among these reductive techniques are redirection (e.g., "John, you should be working on your math now"), ignoring the misbehavior, using a visual or physical cue (e.g., making eye contact, moving near the student, giving a hand signal, and placing a hand on the student's shoulder), verbal warnings and reprimands, taking away privileges, time-out (e.g., requiring that the misbehaving student spend a brief amount of time in a setting away from peers), overcorrection (e.g., requiring a student who writes on his or her desk to wash not only his or her own desk but also those of others), and response cost (e.g., awarding a student 15 minutes of free time but then deducting 1 minute for every incidence of a given inappropriate behavior).

Although easy to use and often effective in stopping misbehavior, at least in the short term, these reductive techniques are subject to the same limitations that characterize all forms of punishment, including corporal punishment, suspension, and expulsion. That is, punishment (i) teaches students what not to do instead of how to behave appropriately; (ii) is nonlasting in effectiveness; (iii) teaches students to punish others; (iv) fails to address factors that underlie or contribute to the punished misbehavior; (v) harms the teacher-student relationship; (vi) is likely to produce undesirable side effects, including resentment, revenge, fear, social avoidance, and poor motivation to learn; and (vii) creates a negative classroom environment.

Despite these limitations, it is perhaps understandable why punishment remains the most commonly used technique for managing and correcting misbehavior: Supported by most parents and schools' code of conduct, nearly all schools require that misbehavior not go unpunished. Moreover, punishment is easy to use. Often, it is effective in immediately stopping misbehavior. Also, it does communicate to all students which behaviors are not acceptable. Although punishment often has these advantages, many authors of popular models of school discipline recommend strongly against the use of punishment. Based more on research, however, other authorities and researchers tend to recommend that mild forms of punishment (not including corporal punishment and school removal) should not be abandoned but be used wisely. That is, it is recommended that punishment be used only after other more positive techniques have already been used and have failed to improve appropriate behavior. Moreover, it is recommended that punishment be used only in combination with corrective techniques designed to teach students to make responsible choices and to exhibit appropriate behavior. Popular corrective techniques that have been shown to be effective in promoting appropriate behavior include individual and class discussions that encourage social problem solving and social decision making, modeling, and the use of various forms of positive reinforcement such as social reinforcement (e.g., teacher praise, peer recognition, and notes sent home to parents), privileges (e.g., recess and parties), and material rewards (stickers, prizes, and certificates) contingent upon good behavior. Not only do these techniques increase appropriate behaviors but also they often simultaneously decrease the frequency of inappropriate behavior. For example, when students are praised or otherwise reinforced for working on their assignments and considering the needs of others, they are less likely to misbehave.

## **5. PREVENTING MISBEHAVIOR WITH EFFECTIVE CLASSROOM MANAGEMENT**

What best differentiates the most effective teachers from the least effective teachers are not the techniques they use to correct behavior problems but, rather, the techniques they employ to prevent behavior problems from occurring in the first place. Effective classroom management not only prevents most behavior problems but also helps

promote the development of self-discipline and academic achievement. It does so by creating a classroom climate characterized by clear procedures, routines, and high expectations, as well as feelings among students of safety, pride, and interest in learning and of warmth, acceptance, and support from the teacher and classmates.

As described in classic studies by Jacob Kounin in the 1960s and 1970s, effective classroom managers demonstrate what Kounin called *withitness* (they closely monitor the behavior of every student in the class), *overlapping* (they deal effectively with multiple events simultaneously), *smoothness and momentum in lessons* (they provide instructions that motivate students), and *group alerting* (they elicit and maintain the attention of all students). Recent research has supported most of Kounin's findings and has identified a number of additional characteristics of effective classroom managers.

In addition to the characteristics identified by Kounin, effective classroom managers can be described as *authoritative* rather than *authoritarian* or *laissez-faire* in their approach. That is, instead of being overly punitive and controlling or overly permissive, they establish clear rules and high behavioral expectations; enforce rules in a firm, fair, and consistent fashion; and implement a variety of activities that promote student decision making, autonomy, responsibility, motivation, and positive self-concept. In the classroom, they work hard to develop positive student-teacher relations. The critical importance of warm and supportive teacher-student relations is supported by ample research showing that students tend to emulate, listen to, and internalize the values and standards of those adults whom they perceive to be warm, respectful, and caring, as well as research indicating that one of the most common reasons for dropping out of school is the perception that teachers do not care.

Effective classroom managers also understand the critical importance of establishing supportive relations among students, recognizing that students who feel alienated, socially rejected, or lonely are at great risk for a variety of negative social, emotional, and academic outcomes. Thus, effective classroom managers use cooperative learning activities, group decision making, and other techniques to foster feelings of belonging and a "sense of community" in the classroom. In addition to valuing the critical importance of teacher-student and student-student relations, effective classroom managers value the need to develop and maintain supportive teacher-parent relations; thus, they strive to communicate often, and in a positive manner, with the parents of all students.

## 6. REMEDIATING SERIOUS MISBEHAVIOR

The previously discussed techniques for developing self-discipline and for preventing and correcting misbehavior are sufficient for addressing the needs of the majority of students. To be sure, the number of techniques used, and their intensity and frequency, should vary depending on the needs of individual students as well groups of students. In particular, a combination of the techniques is needed for students considered at risk for future behavior problems. These students are at risk by nature of exhibiting frequent acts of noncompliance at an early age or by nature of certain biological, emotional, or social factors (e.g., attention deficit hyperactivity disorder, a high level of anger, and peer rejection). Research shows that a combination of the previously discussed techniques, especially when implemented in a systematic fashion and over an extended period of time, can be quite effective in reducing negative social and emotional outcomes among at-risk children. An excellent example is the Families and Schools Together program (FAST Track), an empirically supported, federally funded, school-based intervention program for promoting social and emotional competence and preventing future behavior problems. Whereas certain components of the program are provided to all children, other components are provided to those deemed at risk (i.e., those rated by their teachers as being in the top 10% with regard to conduct problems). Beginning in first grade, all children in FAST Track schools receive two or three lessons weekly in social and emotional learning, which are taught by the children's regular classroom teachers. In addition to receiving training in the teaching of lessons, teachers receive training in discipline and classroom management and are provided consultative assistance from support staff, such as a school psychologist or counselor, in managing children with discipline problems. At-risk children participate in a Saturday program that reinforces the skills taught in the social and emotional lessons during the week. Their parents also participate in the Saturday program and receive intensive training in effective parenting. Another important component of FAST Track is the provision of remedial reading instruction to those students reading below grade level, which includes many at-risk children. FAST Track has been found to be effective in preventing aggressive and disruptive behavior and in fostering a positive

school environment in which students are more likely to follow rules, express their feelings appropriately, show greater interest and enthusiasm, and stay more focused and on task. Substantial improvements in a variety of areas were also found among highly disruptive at-risk students. For example, at the end of first grade they were found to have improvement in social problem solving, emotion coping, positive peer interactions, and social acceptance. Perhaps more important, they were rated by parents, teachers, and unbiased observers as less aggressive/disruptive.

Whereas a combination of techniques for promoting self-discipline and for preventing and correcting misbehavior is appropriate for all children, including children viewed as being at risk for future negative outcomes, a small percentage of these students require more intensive or comprehensive services and supports, or what is referred to in this article as remediation of serious or chronic behavior problems. These students are of two general types: (i) those who exhibit serious behavior problems that are harmful to others or to oneself (e.g., violent acts of aggression, drug use, and self-injury) and (ii) those who exhibit less serious behavior problems that may not necessarily concern issues of harm and safety but nevertheless are chronic and interfere substantially with the learning of others and with the development of healthy interpersonal relations with others. These behaviors continue despite the implementation of multiple interventions for developing self-discipline and for preventing and correcting misbehavior. Many, but not all, of these students are the estimated 3–6% of students in need of special education and support services due to an emotional or behavioral disorder. For these students, and others identified as having a disability, federal law [i.e., the Individuals with Disabilities Educational Act (IDEA)] specifies that they can be removed from the regular classroom, without parent permission, if they have committed one of three behavior problems: possession of a weapon, possession of illegal drugs, and any behavior that is judged by a special “hearing officer” as “substantially likely” to injure self or others. IDEA requires that during their removal students with disabilities be provided a continuation of their individualized educational program, including remedial and support services (e.g., counseling).

The codes of conduct in many schools require removal from the regular classroom, or school, as a consequence of serious or chronic behavior problems, as well as for the purpose of protecting the safety of others. Thus, students are often expelled or placed in

an “alternative” educational setting—an alternative to the regular classroom and to expulsion. The number of alternative educational settings has increased greatly in recent years, reflecting an increased zero tolerance approach to discipline problems, recognition of the limitations to expulsion, and the fact that IDEA forbids the expulsion of students with disabilities without a continuation of special education services.

Although there is a lack of research on the effectiveness of alternative educational settings, there is considerable research on the effectiveness of various interventions for children with serious and chronic behavior problems. Such research indicates that effective intervention programs have many characteristics in common. They are comprehensive, addressing a combination of multiple risk and protective factors, and broad based, adopting a systems perspective that recognizes the critical importance of schools working closely with families, outside agencies, and communities. Moreover, the interventions are intensive and sustained over time, implemented early, guided by theory and research, and sensitive to developmental differences and the determinants of behavior.

The primary purpose of remedial services and supports is to address the child’s current behavior problems and prevent their reoccurrence. Occasionally, such as when serious acts of violence are involved (e.g., use of weapons or a homicide), a crisis exists, consisting of a stress-inducing event in which students, as well as others in the school, experience a breakdown in the ability to cope. In preparation for such unfortunate events, it is recommended that schools have a crisis response team as part of the remedial component to their comprehensive school discipline program. Such a team would consist of specially trained teachers, administrators, and support staff who respond immediately to the crisis and who provide psychological support and resources needed to alleviate existing stress, restore coping skills, and prevent future emotional problems related to the crisis.

Each of the previously discussed four components is a necessary, and critical, part of a comprehensive school discipline plan. Their key features are summarized in *Table I*. Unfortunately, even among the best schools, correction and remediation are necessary. However, these schools recognize that an emphasis on developing self-discipline and preventing misbehavior with effective classroom management is the most effective way to reduce the need to correct misbehavior and remediate serious behavior problems.

**TABLE I**  
**Components of Comprehensive School Discipline<sup>a</sup>**

**Developing self-discipline**

Provide multiple models of desired cognitions (e.g., moral reasoning, social problem solving, and responsible decision making), emotions (e.g., pride, assumption of responsibility over wrongdoing, empathy, and regulation of anger), and prosocial behavior.

Reinforce students, especially with sincere praise, when they exhibit the above qualities.

Implement a research-supported curriculum that directly teaches the above qualities.

Provide ample opportunities for students to apply social and moral problem-solving and decision-making skills to real-life problems and to demonstrate empathy and prosocial behavior.

**Correcting misbehavior**

Emphasize the use of positive techniques that teach students thoughts, emotions, and behaviors that should replace the given misbehavior.

Never use punishment alone, but always in combination with positive techniques designed to teach appropriate behavior.

If punishment is used, use the mildest technique necessary to improve behavior and a technique that is consistent with the severity of the offense.

Always be aware of the many limitations to the use of punishment.

**Preventing misbehavior with effective classroom management**

Be authoritative, not authoritarian.

Demonstrate warmth, respect, and caring toward all students.

Establish and maintain close communication with each student's parents or caregivers, and work hard to garner their support.

Provide academic instruction and activities that motivate learning.

Create a physical environment that is conducive to teaching and learning.

Establish clear behavioral expectations, fair rules, and consequences.

Establish predictable procedures and routines.

Frequently monitor student behavior, praising appropriate behavior and responding immediately to signs of misbehavior.

**Remediating serious or chronic misbehavior**

Provide a comprehensive program that is intensive, sustained over time, and offers a wide range of services, such as adaptive and remedial instruction, consultation to teachers, counseling, and parent education.

Adopt a systems perspective toward understanding, preventing, and treating behavior problems. Whenever possible, always involve parents or caregivers in interventions.

Intervene early: Provide, or arrange for, services when behavior problems first become evident. This may be before or after children enter school.

Be sensitive to developmental and cultural differences in behavior and responsive to the multiple determinants of behavior.

Target multiple risk and protective factors associated with serious and chronic misbehavior, including biological, cognitive, emotional, and environmental factors.

Be aware of special provisions in federal law pertaining to discipline and children with disabilities.

Implement a crisis response plan and team.

<sup>a</sup>Adapted from Bear (2005).

**See Also the Following Articles**

Emotional and Behavioral Problems, Students with  
■ Learning Disabilities ■ School Violence Prevention

**Further Reading**

Battistich, V., Schaps, E., Watson, M., Solomon, D., & Lewis, C. (2000). Effects of the Child Development

Project on students' drug use and other problem behaviors. *Journal of Primary Prevention*, 21, 75–99.

Bear, G. G. (2005). *Developing self-discipline and preventing and correcting misbehavior*. Boston: Allyn & Bacon.

Bear, G. G., Quinn, M. M., & Burkholder, S. (2001). *Interim alternative education settings for children with disabilities*. Bethesda, MD: National Association of School Psychologists.

Bear, G. G., Webster-Stratton, C., Furlong, M., & Rhee, S. (2000). Preventing aggression and violence.

- In K. M. Minke, & G. G. Bear (Eds.), *Preventing school problems—Promoting school success: Strategies and programs that work* (pp. 1–69). Bethesda, MD: National Association of School Psychologists.
- Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track Prevention Trial of Conduct Problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 631–647.
- Dwyer, K., & Osher, D. (2000). *Safeguarding our children: An action guide*. Washington, DC: U.S. Departments of Education and Justice. [Available at <http://www.ed.gov>]
- Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., Kessler, R., Schwab-Stone, M. E., & Shriver, T. P. (1997). *Promoting social and emotional learning: Guidelines for educators*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Jones, V., & Jones, L. S. (2004). *Comprehensive classroom management: Creating communities of support and solving problems*. Boston: Allyn & Bacon.
- Kauffman, J. M. (2001). *Characteristics of emotional and behavioral disorders of children and youth* (7th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
- McClellan, B. E. (1999). *Moral education in America: Schools and the shaping of character from colonial times to the present*. New York: Teachers College Press.



# School Environments

Rotraut Walden

University of Koblenz, Koblenz, Germany

1. Introduction
2. Empirical Evidence Regarding School Environments
3. General Models Describing Environment–Behavior Relations in School Environments
4. Schools of the Future  
Further Reading

## GLOSSARY

**acquisition wall in classrooms** Wall that should include the chalkboard and a bulletin board and be in the front of the room; maintenance walls should be along the sides of the room and should include material that helps students to review work; dynamic walls should be at the back of the room and contain student work, school notices, and holiday decorations.

**appropriation** A process by which the objective environment is transformed into a subjective and personally meaningful environment; constructive appropriations can be expressed through articles such as graffiti, works of art; personalization, changes for use, and the use of an environment. Negative appropriations are shown through vandalism.

**centralized schools** Schools that are made up of two or three large buildings; decentralized schools are composed of many smaller buildings that are scattered around the campus.

**crowding** The subjective perception of high density; a greater density of people is linked to negative effects, physiological arousal, (sometimes) an increase in illness, changes in social behavior, a decrease in willingness to help others, and an increase in aggression.

**density** An objective measurement for spatial demarcation; there is a difference between the amount of available space

per person (i.e., spatial density) and the number of people per unit of space (i.e., social density).

**open-plan classrooms** Large areas of undivided flexible space; the educational changes that accompanied them grew out of the 1960s and were a departure from the standard “egg carton” design used in traditional classrooms.

**smart classroom** Classroom that includes a computer at the teaching station, the availability of a television or video-cassette recorder, and the ability to project or send a computer signal to a large monitor, projector, or each student station; it is a communication-rich classroom that brings the outside world into the traditional classroom.

**sociofugal seating arrangements** Rows of seats that are turned away from one another, making communication difficult.

**soft classroom** Classroom with semicircular cushion-covered bench seating, adjustable lighting, a small carpet, and some mobiles; student participation increases markedly in this type of classroom.

**stress in schools** Conceptualized as an outcome resulting from the lack of congruence between individual needs and goals with the opportunities and constraints afforded by schools.

**traditional classroom** Classroom that consists of rows and columns of student desks, a horseshoe, or perhaps clusters of four that face the teacher’s desk and the blackboard at the front of the room.

The physical school environment can be considered a school’s “third teacher,” after other children in the class and teachers themselves. The goal of designing new buildings, as well as modifying existing educational facilities, should be the architectural stimulation of achievement, social cooperation, and well-being.

Achievement and social cooperation are also evaluated in schools by student grades. This article describes the results of studies on different influencing factors: the size of school buildings and classes, the ground layout and façade, the state of the buildings, and their age. It examines the effects of privacy, density, seating positions, and open-plan classrooms. It also reports the results from studies on interior architecture (e.g., access to windows). A number of factors have an immediate effect on the sensory organs: noise, climate, color, lighting, and smell. The article then discusses schoolyards and playgrounds as well as the danger posed by vandalism. It then concerns itself with two conceptual frameworks for analyzing the effects of school environments on the behavior and experiences of teachers and students. A review of recent develops details the criteria that can be used to measure a school's success and details how a school of this kind should be designed. Acquiring a sense of competence with one's environment is deemed to be fundamentally important for future generations, and the opportunities for realizing this are an important part of the physical conditions that a school of the future should offer in the end.

## 1. INTRODUCTION

After a study carried out by the Organization for Economic Cooperation and Development's (OECD) Program for International Student Assessment (PISA) in 2000, industrialized nations such as the United States and Germany discovered that the performance of their students in relation to international rankings was relatively poor. Since this so-called PISA shock, education is once again being examined closely in various countries. The focus of this has been the development of performance standards, greater control through testing, and a broadening of subjects offered during the afternoons. In relation to this, questions have been raised about the causes of poor performance and discussions have centered on learning conditions and the learning environment. In connection with this, environmental psychologists have also been consulted. The physical school environment can be considered a school's "third teacher," after other children in the class and teachers themselves. Therefore, formal and other architectural concerns of school design should be subordinate to educational demands.

Various studies have shown that the educational demands on schools are multifaceted. Schools should be places that facilitate conflict resolution and encourage learning and living, that is, places to meet for social

learning that facilitates conflict resolution and encourages both privacy and public interaction. Learning is defined as a relatively permanent change in behavior as a result of experience. Good teaching can be found in poorly constructed schools, just as bad teaching can exist in well-built schools. What is commonly accepted, however, is that a school building can either facilitate or hinder learning, as observed by Weinstein in 1979.

The effects of the physical setting are moderated by the social and instructional contexts. Schools should involve varied sensory experience and should empower children to interact with their environment in a competent way. They should encourage development of creativity and activities, individual diversity, and teamwork. At the same time, a school can be an effective architectural, aesthetic, ecological, economic, and sustainable role model. Well-designed schools are the prerequisites for the development of successful motivational learning methods that also contribute to enjoyment of learning and a willingness to excel for students and teachers alike. Ultimately, students are the bearers of future innovations for worldwide competition and for society's benefit. The demographic change brought about by an increase in life expectancy demands that fewer and fewer young people support an ever-increasing aging population in industrialized nations.

As early as 1978, Gump calculated that the average person spends approximately 14,000 hours in school from kindergarten through to the 12th grade. Since the results of the international PISA study in many countries were reported, many schools where students were previously taught for only a half-day became full-time schools in the hopes of increasing the achievement levels of students. In Germany, for example, 10,000 of more than 40,000 schools will be full-time by 2007.

As mentioned previously, school environments should help to improve achievement, social cooperation, and well-being. A school of the future should fulfill all of these functions for students, teachers, and parents on a daily basis.

## 2. EMPIRICAL EVIDENCE REGARDING SCHOOL ENVIRONMENTS

### 2.1. Special Influences in Schools and Classrooms

During the 1970s and 1980s, most towns in the United States and other countries built very big schools. The

rationale for constructing schools of this size was the expected economic benefits and advantages of offering students a wide curriculum. Today, there is a reversal of that trend, whereby large schools are divided into smaller, more personal units. The goal is, among other things, a greater sense of personalization in the schools that is shown through decorations; participation in school clubs, recesses, and artistic performances; and the expression of a feeling of "I like to be here." The ministries of education had been seeking a solution to poor student performance, problems caused by accidents and crime caused by low levels of personal support, difficulties with supervision, and too much anonymity in schools.

The advantages of smaller schools seem to outweigh those of their larger counterparts. According to Wasley and colleagues in 2000, the benefits of small schools can be found in intimate learning groups, where students are well known by their teachers and can be encouraged to perform better and where adults can look after the students in their care more adequately. Isolation, which manifests itself as alienation, vandalism, theft, and violence, can be avoided. Smaller schools enable children from minority or economically disadvantaged backgrounds to advance sooner, and teachers can be encouraged to draw on their own experiences for the good of the children. In addition, teaching activities can be coordinated more effectively. As early as 1964, Barker and Gump's classic study showed that students in smaller schools are more likely to take part in school activities, even though larger schools offer more opportunities for such activities.

The need for disciplinary measures clearly declined when smaller class groups were introduced in Granby High School in Norfolk, Virginia. In 293 secondary schools in New Jersey, school size was the second strongest predictor (after socioeconomic status) for student achievement in central tests carried out by Fowler and Walberg in 1991. Further positive aspects that accompanied this discovery were greater satisfaction from parents, teachers, students and members of local government as well as a greater responsibility for students on the part of teachers, greater cohesion within groups, and more contact between individuals and the school staff and faculty. Researchers for the Rural School and Community Trust found that smaller schools always outdo their larger counterparts in terms of achievement, particularly with regard to the performance of economically disadvantaged students.

In 1984, Goodlad recommended that elementary schools should not have more than 300 pupils. The

highest levels of negative effects were found in high schools with more than 2100 students. Centralized schools with between 500 and 700 students, where children take an active role in activities outside of lessons themselves, and schools with attractive façades appear to be ideal schools for children. In a school of this kind, the number of students is arranged so that there is as little anonymity as possible among students, thereby giving rise to less criminal behavior. Once enrollment exceeds the threshold of 700 students, further increases seem to have little effect. The layout of the school buildings should be centralized to give students as many opportunities as possible to communicate, even between classes. In this way, students can gain more personal recognition through actively taking part in extracurricular activities. In 1982, Glass and colleagues maintained that once the size of a class reaches 20 to 25 students, adding more students seems to make relatively little difference. They concluded that smaller classes lead to better learning environments in nearly every respect, including teacher and student attitudes, interaction, and achievement.

Admittedly, the number of students or the size of a school can only rarely be set because most school buildings are old. Therefore, the teaching staff is asked to make commitments to reduce anonymity among students as much as possible in larger schools by organizing school festivals, letting students share input into the design of classrooms and corridors, and giving students in decentralized schools as much time and as many opportunities as possible to communicate with each other. The question of whether the state of maintenance of a school may have an influence on students was pursued by Gifford in 2002. According to his research, the influence of a school's state on students could not be determined, but he assumed that happier teachers working in more attractive schools are also better teachers. In conclusion, it is clear that students who are taught in new or renovated schools are happier.

The complexity of these concepts renders it difficult to make clear assertions about an optimal room layout. However, some basic assertions should be summarized. Each student needs a certain amount of privacy. Students who show deviant behavior might need a higher level of privacy, and facilities such as "privacy booths" and/or quiet corners could be made available for these children. Whether one decides on measures such as these or similar ways of encouraging privacy, the facilities should be either accessible to all students or not acquired at all. When students sit close to friends, or

near the door, in a classroom with high walls, they feel safe and, therefore, do not need as much personal space among themselves, the teacher, and other students. This can be taken into account by the division of the seating plan in the classroom. In addition, by approaching the students and being closer to them, the teacher can shed some of the distance that his or her status affords, thereby building trust with the students.

An ideal classroom density is between 2.8 and 3.7 square meters of space per student. If the density is higher than this, it can have a negative effect on the students. In any case, investigations carried out on these influences often end up with conflicting results, so that the effects on students must be considered as highly differentiated. No single, all-encompassing assertion about the effects of classroom density on students' performance can be made.

Many investigative results agree that higher classroom density impairs performance when students are carrying out complex tasks, working in groups, or engaging in work that requires mobility. In addition, other influencing factors, such as high levels of motivation, lack of the opportunity to choose one's own seat, a feeling of crowding, and the ages and ethnicities of students, all can exert an influence on students' performance and should be considered. The well-being of students can also be impaired by a classroom density that is too high. Likewise, social behavior gets worse because students must compete for the available teaching materials and the teacher's attention, and this presents itself as an increased level of aggression. This is particularly the case with younger pupils, meaning that a correlation between the cognitive state of students and the reaction to increasing density can be determined.

Just as with classroom density, a student's seating position in a classroom can influence his or her performance, well-being, and social behavior. Therefore, according to results of some research, places in the middle and toward the front of the classroom offer an optimal position for increased achievement and participation in lessons. Furthermore, students connect certain positions within the classroom with emotions and choose their seats with this in mind. In 2001, Bell and colleagues mentioned that students who choose to sit in the middle front seats in a classroom have a higher level of self-confidence than do students who occupy other seats. According to sources in McAndrew's 1993 book, the rear of the classroom is associated with freedom to interact with peers and freedom from teacher control. Consequently, there is a self-selection of pupils in the different rows. However, in Griffith's classic 1921

study, the grades of students who had been assigned to seats alphabetically were carefully recorded over many different classes. Griffith detected that students in the center of the class tended to receive the highest grades, whereas a steady dropoff occurred as the seating position got farther away from the center. If generally quiet students were seated in the "action zone," their participation in classroom discussions did not increase. Communication among students can also be influenced by seating arrangements. Therefore, when regulating the division of seating in a room, as much care as possible should be taken to create a "sociopetal" arrangement that encourages communication, for example, by using group tables.

The open-plan classroom, which found its way into many schools during the 1970s, can have a negative effect on the performance of average students, whereas it encourages those students who are above average. Average students are more susceptible to diversions, whereas those who are above average make use of the diverse learning opportunities. Students' well-being can be affected by an increased level of noise and undesirable distractions. These negative influences are juxtaposed to increased concentration on schoolwork and partially increased communication. However, the open classroom has lessened in popularity over the past 15 years or so due to its negative effects. In 1993, McAndrew concluded that although it is true that most studies comparing the actual achievement of children in open-plan versus traditional classrooms show no major differences, studies that do report differences tend to favor traditional classrooms over open-plan ones. A combination of the traditional classroom and some components of the open-plan classroom could, however, be beneficial for students, and the open concept of group work on tables could be applied. What is important is that the classroom type must correspond to teaching methods, the students' and teachers' need for stimulation or privacy, the type of activity, and the length of time spent on the activity.

It is not only the exterior appearance of a school and its layout that have an influence on students. The interior layout in terms of the division of walls and furnishings, the position of windows, and the type of flooring, decorations, and equipment in the classroom also has an influence. There is much to consider in the interior layout of a school. Temporary walls, or walls that are too low, increase distractibility. The main exterior walls should be high if possible. The division of walls for "acquisition," "maintenance," and "dynamic walls" clearly have advantages if they are used in

classrooms. Empirical results also indicate that decorations and images with “happy” motives should be used.

Having a school that one finds attractive is associated with better achievement, but boys and girls might not find the same decor attractive. In 1990, Cohen and Trostle indicated that girls prefer more windows, color, texture, shapes, and lighting, whereas boys prefer larger settings. When learning occurs in a given classroom, that material is better recalled in the same setting than in a different setting. Some researchers believe that classrooms should have more stimuli and opportunities for environmental exploration that facilitate learning. Others worry that students will have difficulty in concentrating on schoolwork. Both ways depend on the purposes of learning. On the one hand, learning includes social responsibility; on the other hand, it includes acquisition of “dry” intellectual learning material. Bell and colleagues concluded in 2001 that seeking the right fit between person and learning environment is probably the most desirable approach.

Research has shown that the floors should be laid with carpet and that students should always have an unimpaired view of teachers so as to encourage communication between students and teachers. According to research quoted by Gifford in 2002, there is a more direct, close, and learning-oriented contact between students and teachers when the floors are carpeted than when they are not. A lack of dividing walls or screens in classrooms is beneficial so that teachers do not engage in administrative work but instead remain in constant interaction with students. Furthermore, social interaction can be improved using new equipment and facilities. An increase in social communication of this kind is also possible with old furniture if it is renovated.

Some results of research show that the absence of windows leads to increased achievement of students, whereas others demonstrate that it leads to decreased performance, depending on the task given. According to some investigations, a classroom with windows and natural light is a good precondition for high achievement and is especially influential for the positive well-being of students.

Increasing demands are being made on computer equipment and freedom from glare in “smart” classrooms. Furnishings should correspond with ergonomic demands and should be suitable for the size of children and adults.

Despite the fact that “good acoustics” have a strong impact on successful learning in school, the effects of noise are not considered to be important as they should be, according to a 2003 paper by Klatt and colleagues.

Deficits regarding room acoustics occur when classrooms are too reverberant, speech intelligibility is too low, and/or background noise levels are too high. Talking and listening to one another is still a major teaching and learning interaction, although there are increasing Internet-based and multimedia forms of learning. Acoustic conditions are often very poor, and verbal communication is only possible by screaming. Physical education teachers often must endure levels of noise between 90 and 100 decibels. The measurement of noise levels in elementary schools in Germany during the entire length of a class period show average levels of 70 to 77 decibels.

In comparison, German legislation concerning workplaces places a limit of 55 decibels on the average noise level of labor in areas primarily devoted to mental activities. If verbal communication is the goal, values of no more than 40 decibels are allowed. Noise is a fundamental stress factor. What is more, the speaker’s level should be approximately 10 to 15 decibels higher than the noise level. Teachers who have spent much time in their occupation suffer greatly from noise. The flow of instruction is interrupted by frequent repetitions of information and warnings to be quiet. Reluctance, anger, and fatigue, as well as throat and voice problems, often result. Communication suffers and is reduced. What is more, poor hearing conditions cause verbal information to be misinterpreted or completely misunderstood. This leads to a decrease in resources available for short-term memory and mental processing of aurally given information. A growing number of children suffer from hearing disorders. There are even effects caused by poor acoustic conditions that influence the processing of visually presented information. The students do not adapt to background noise. Despite apparent deterioration in performance with background noise, the noise is often judged to be hardly disturbing or even not disturbing. It can be assumed that learning to read and write is complicated in surroundings that are too noisy. Even though music is associated with being pleasant, the same piece of music played by the same instrument as staccato music instead of legato music distracts from memory performance. (Legato music is characterized by tones attached to each other, whereas staccato chains include short and distinct tones.) The conclusion to findings about noise is not a demand for absolute silence and discipline in classrooms. Free work, singing, and joyful events go along with an increase in noise levels in classrooms. Therefore, it is essential to design classrooms in such a manner that the

potentially negative effects accompanying such positive educational development are prevented. This can sometimes be achieved by simple measures such as renewing felt padding under tables and chairs, checking furniture for squeaky drawers and rattling desk pads, oiling door hinges, and putting up heavy curtains, wide cork boards, and tapestries.

As was shown earlier with noise, a perfect solution to the problem of climate for all students and teachers is hard to find. The guidance levels for the influence of climate on performance of employees in offices are comparable to those of students in classrooms (21 degrees Celsius, 50% humidity, 0.1 meter per second air movement). Air conditioning, often found in the United States, is often set too cold and does not take care of the problem of providing fresh air. Air supply is also dependent on the location of the appliance from which the air is taken in (car park or roof). Gifford concluded that even when care is taken in the control of indoor climates, there must inevitably be a compromise that considers individual differences in amount of clothing, type of activity, and indoor or outdoor movement. There may also be preferred temperature levels based on adaptation to one's home climate or cultural background.

There is no given single color that positively influences all students' achievement, well-being, or social behavior. However, one can determine that well-lit rooms have a more positive effect on students than do sad and dark ones (because well-lit rooms diminish feelings of crowding) and that blue is a better color than red (because blue is more relaxing). Whether or not orange and pink rooms positively influence the social behavior of students cannot be clearly confirmed.

According to Gifford, many individuals prefer incandescent lighting, but it is more expensive than fluorescent lighting. Fluorescent lighting has not been shown to have dramatic negative effects on the performance or health of most students.

From a variety of research on the influence of lighting on performance, well-being, and social behavior of people, it is clear how important daylight is for people, regardless of whether it is streaming into a room directly from a window or is simulated using special lamps. Performance is better, and positive social behavior is more frequent, when daylight is available. However, light that is too bright or strong, as is often found in offices and classrooms, has a negative effect on people's well-being. Cold light increases the negative aspects of women's mood, whereas it decreases those of men, according to a study conducted by

Knez in 1995. This gender difference is problematic because it is clear that in a society in which males and females share many spaces, cold or warm light is advantageous for only one of the genders. Therefore, a lighting level must be found that is as good as possible for both genders. A classroom with windows and additional daylight lamps whose brightness can be regulated could be a solution for many students.

Although the connection between fragrances and performance, well-being, and social behavior is described here on the basis of only a single investigation, this is based on results gleaned from previous research that were confirmed or strengthened in this investigation. Therefore, the results may be considered reliable. However, this work did not concern young pupils; rather, it concerned university students. However, the overall positive influence of pleasant fragrances could possibly be transferred to school pupils. A separate investigation would have to be undertaken to substantiate this. Overall, this investigation showed that fragrances can positively influence a person's performance, well-being, and social behavior (the latter in the form of voluntary help). When applied to schools, this means that teachers would use different fragrances according to their requirements. Results from pleasant fragrances are consistent with an interpretation suggesting that pleasant fragrances can enhance task performance by serving as a source of environmentally generated positive affect.

According to McAndrew, schoolyards and playgrounds can be viewed as a type of learning environment if learning about social skills and cognitive abilities takes place there. This is particularly the case with playgrounds because pupils who would otherwise interact in a relatively controlled space must get along with each other independently and with little supervision. What should playgrounds look like to give children the necessary stimulation for learning? They should be made up of different corners that stimulate all of the senses and that are linked to each other using pathways. In addition, playgrounds and schoolyards that are less structured (e.g., adventure playgrounds with mountains of tires piled up) are more stimulating to children's creativity and encourage them to use role-play. Children using playgrounds like these reported that they came up with ever more creative ways of playing with the tires and that they felt a sense of success when they achieved their goals. Children in traditional schoolyards, which are normally fenced in and paved with asphalt, usually play games involving equipment brought from home such as balls of various

sizes. They occupy themselves less with the toys and equipment available because these bore them and can be used in only a limited number of ways. A further article on schoolyards and playgrounds, carried out by Moore in 1989, involved the introduction of nature and natural materials. Through their steady growth and changes, these plants and materials spur the imagination. Children also exercise their muscles by climbing trees. They play with branches and stones and hide in bushes. In addition, these plants and materials bring children closer to nature and let them learn about the problems facing the environment and the ways of working with tools and natural materials. By working with and learning about natural materials, children become more open and competent in interacting with their environment. Further opportunities for improvement of schoolyards and playgrounds include the introduction of safe entrances and exits that can serve as meeting points for children and adults and to bundle social contacts that link people to these places. In this way, greater control over crime can be exercised. In addition, there should be pathways, some of which lead to toilets and telephones and others of which lead to the countryside or surrounding areas.

Goldstein's book on the psychology of vandalism, published in 1996, contained three studies on school vandalism. A study by Pablant and Baxter concerned environmental correlates and confirmed the relationship between vandalism rates and the aesthetic quality and level of preservation of school property. Schools with low vandalism rates were characterized by excellent upkeep of buildings and the surrounding playgrounds. Highly vandalized schools were characterized by neglected maintenance and building upkeep, and the school grounds showed little evidence of aesthetic appeal or efforts to make them pleasant. In general, vandalized schools offered an appearance of abandonment regardless of architectural material or design. The age of the schools was found to bear no relationship to the frequency of vandalism. It was found that the schools' overall quality was generally superior to that of the immediate neighborhood in the low-vandalism group and was generally inferior to that of the immediate neighborhood in the high-vandalism group. Beautification efforts may be more effective in deterring acts of vandalism than are steel fences, electronic sensors, and expensive fortress-like buildings. The latter features did not differentiate between schools with high rates of vandalism and those with low rates of vandalism. Neighborhood diversification of use appears to be an important environmental variable

affecting schools' safety. Schools with low rates of vandalism were generally located in areas characterized by multimodality of use such as residences, commercial buildings, and churches. Fully 81% of schools with high rates of vandalism faced uninhabited land such as parks and sports fields. It may be assumed that isolated schools, or schools located in low-density areas, have less neighborhood vigilance and are more vulnerable to vandalism. Schools that recorded lower rates of vandalism could usually be seen by neighbors, were better lit, and had fewer restrictions such as trees blocking the view of the property. Recommendations for decreasing vandalism included encouraging activity on the grounds by opening pathways through school property and by using the community, maximizing the illumination surrounding the school property, and maximizing the aesthetic appeal of the property.

## **2.2. Environmental Competence as an Educational Goal**

Environmental competence is understood to be the ability to cope with the effects of involving oneself in an environment and its appropriation and, thereby, in changing the environment. It involves learning about the environment. According to Gifford, it includes three kinds of learning: (a) personal style, attitudes, and awareness of physical setting; (b) knowledge of physical settings, including technical knowledge, how to unearth new information, knowledge about how social systems control space, and knowledge of person-environment relations; and (c) practical environmental skills such as scouting, matching, personalization, and creative custodianship. Programs in and out of school should teach many different facets of environmental competence, from basic environmental ethics, to environmental protection, to campfire building, to architectural design.

## **3. GENERAL MODELS DESCRIBING ENVIRONMENT-BEHAVIOR RELATIONS IN SCHOOL ENVIRONMENTS**

### **3.1. A First Taxonomy for the Study of Stress in Schools**

Acting on the basis of studies known to that point, Ahrentzen and colleagues in 1982 developed a

taxonomy for the study of stress in elementary school settings. At that time, avoiding stress was of primary concern, and less so was the encouraging performance and creativity using stimulating environmental conditions. Most perceptions were based on the idea that pressure to succeed created learning difficulties. Stress is conceptualized as an outcome resulting from the lack of congruence between individual needs and goals and the opportunities and constraints afforded by the school. Examples of stressful environmental characteristics are “sociofugal” seating arrangements that hinder interaction, whereas “sociopetal” seating arrangements are stressful for students who seek privacy from others during study. Windowless classrooms may produce a decrease in well-being. The amount of aural and visual stimulation in open-plan classrooms may hinder students from performing class activities efficiently and speedily over time. Sometimes, environmental constraints do not influence achievement directly but reduce satisfaction and comfort.

The taxonomy used by Ahrentzen and colleagues arose from the “matching” or “fit” between personal attributes and environmental attributes. This taxonomy includes (a) personal attributes, (b) physical/architectural environmental attributes, (c) social/organizational environmental attributes, and (d) indicators of person–environment fit—all important variables that have been established by the literature.

The personal attributes include past classroom environment experience and the locus of control in open and traditional classrooms. The need for achievement and the related concept of competition/cooperation are meant to interact with environmental characteristics. Children and adults differ in the amount of stimulation desired from the environment. Therefore, the sensation- or arousal-seeking tendency will interact with the environment. Stress will result when there is a mismatch.

Ahrentzen and colleagues differentiated between two kinds of environmental attributes: physical/architectural design and social/organizational elements. Each of these environmental elements can influence educational outcomes directly or indirectly through the other domain. The physical/architectural design features include class enrollment, density, furniture arrangement, seating position, decorations, and aural and visual stimulation. Furthermore, territoriality and boundary clarity, privacy areas, and soft/hard atmosphere of the classroom are mentioned. A soft classroom allows more control by users. Resources mean the accessibility of materials (or teachers). In addition, these authors included openness/containment. They

aligned themselves with Moos in their social/organizational elements of the taxonomy to characterize the social climate of a classroom: student involvement, affiliation, teacher support, task orientation, competition, order and organization, rule clarity, teacher control, and innovation. According to Moos’s 1979 book, these elements consist of three social climate dimensions: relationship, personal growth, and system maintenance and change. Ahrentzen and colleagues added the teacher–environment fit. The expectations, goals, and needs of a teacher should be supported by his or her environment.

The indicators of person–environment fit are attitudinal indicators such as satisfaction with the classroom setting, commitment to schoolwork, and perceived quality of school life. Behavioral indicators include attendance, disruptive behavior, questioning behaviors, participatory behaviors, persistence in completing tasks, fidgety behavior/nervousness, and creativity. Students will expend more effort on difficult tasks in settings that are supportive of their expectations and needs.

### **3.2. A Recent Framework for Conceptualizing Person–Environment Relations in Learning Settings**

In 2002, Gifford mentioned nearly the same relationships among the personal characteristics of learners, the physical features of the learning setting, and the social/organizational climate of learning-related attitudes and learning behaviors. He drew a model with four boxes and their relationships to each other and called it a framework for conceptualizing person–environment relations in learning settings. When Gifford mentioned (a) the personal characteristics, he also started with school experience and then continued with attitude toward school, motivation to learn, age, gender, and personality. Then, he specified (b) physical features such as size, noise level, climate, population density, and design. The literature offers evidence that supports adding fragrances/odors to the list. By (c) the social/organizational climate rules, he meant curriculum and teaching styles, regardless of whether they are progressive or traditional in orientation. These all influence (d) reactions such as learning-related attitudes (e.g., satisfaction with the school or classroom, the desire to learn) and reactions such as learning behaviors (e.g., on-task time, class participation, questioning, appropriate

or inappropriate activity, persistence, creativity, learning, performance).

### 3.3. Newer Methods for Assessing School Buildings and Criteria for Improvements in the School Environment

In 2002, Henry Sanoff advocated the participation of the local community, students, parents, and teachers in the design and construction process, and in 2001 he developed methods for assessing school buildings. One of these is the six-factor school assessment scale. Furthermore, Sanoff proposed an observation form and rating scales for school buildings, encompassing everything from indoor and outdoor learning space, to informal social areas, to the dining space, to classrooms. He also used open-ended questions such as "I wish my school . . ." and used photographs of the school to assess it.

The school environment should contribute to the overall success of schools. According to Duke and Trautvetter in 2001, partially measurable objectives for school design include, among other things, how well they contribute to a more caring and supportive school culture and to improved achievement in standardized tests and better grades overall. The appropriate social, administrative, educational, or disciplinary measures can bring about reduced disciplinary referrals, suspensions, and exclusions from lessons. These measures can also facilitate faster intervention for children with learning difficulties and greater participation in different kinds of lessons and other instructional activities. Finally, parental involvement and better communication between the school and home can be increased. In the long term, successful schools will see fewer teachers taking sick leave or being lost through early retirement.

## 4. SCHOOLS OF THE FUTURE

How should schools of the future be built from the point of view of environmental psychology? In a qualitative study conducted by Walden and Borrelbach in 2002, the architects of six innovative schools, chosen by the editorial journalists of professional trade journals, answered this question in interviews. Intelligently designed schools should offer opportunities for users to make their own adjustments to the environment in response to their own levels of arousal, stress, overload, fatigue, and attitudes. Such adjustments include

physical conditions such as lighting, color, acoustics, noise, heating, air-conditioning, and the ergonomic composition of the buildings and their furnishings. Users' identification with school buildings is facilitated through participation and appropriation processes. This sense of control by the users should result in increased responsibility for damage and neglect as well as in a decline in vandalism. In Germany, Walden used data from supervisory school authorities in medium-sized cities to estimate the cost of vandalism to be approximately 50 million Euros per year for more than 40,000 schools. (In 1990, Stoner and colleagues estimated that the annual monetary cost of school vandalism in approximately 84,000 schools in the United States amounted to approximately \$600 million, according to Goldstein's book from 1996.)

The special needs of children and teenagers should be considered, for example, in the design of playgrounds, entrances, and circulation areas. Specific educational areas should be provided for learning, working, discovery, play, orientation, physical education, and recreation. Groups with special needs and disabilities would benefit not only from legally required entrances with wheelchair access, widened doors, walking frames, steps, handrails (for use by children as well), elevators, and rest rooms but also from specially designed play items.

What are the worst mistakes that one can make when building a school?

- Building too big (results in anonymity)
- Building too small (e.g., few provisions for clubs and activities, lack of specialized rooms)
- Poor connections between entrance and classrooms
- Waste of energy (due to drafts and bad heating insulation)
- Classrooms too small
- Poor arrangement of facilities for faculty, staff, and students as well as for the community
- Bad noise insulation
- Insufficient facilities for the disabled
- Poor upkeep and maintenance (leads to an increase in vandalism and uncleanliness)

The school of the future should offer a quality of experience, even at the planning stage, in consideration of the wishes of students and teachers. Also, it should provide a livable design and exude a pleasant atmosphere. To achieve a feeling of "being at home" in a school of this kind, or one where students and teachers spend a lot of time, it is crucial that children, teachers, and parents are able to participate in its

design. A further effect of “nice” or “pleasant” self-design of schools is an increased feeling of responsibility for one’s environment, an increase in creativity and environment intelligence, and a decrease in vandalism (Figs. 1–6).



**FIGURE 1** Provisions for rich sensory experiences and privacy (school by Peter Busmann). Reprinted courtesy of Asanger Press, Germany (Walden & Borrelbach, 2002).



**FIGURE 2** Sound Stone, a stone that buzzes. Reprinted courtesy of Asanger Press, Germany (Walden & Borrelbach, 2002).



**FIGURE 3** Classroom with decorations. Reprinted courtesy of Asanger Press, Germany (Walden & Borrelbach, 2002).



**FIGURE 4** Wood as warm material in schools (by Peter Hübner). Reprinted courtesy of Asanger Press, Germany (Walden & Borrelbach, 2002).

Schools looking to the future should take the following into account:

- The right of all users to participate in decision making
- The use of environmentally friendly construction materials
- The opportunity to regulate environmental conditions (e.g., lighting, heating) oneself while at the same time using sensory technology
- The importance of a clear orientation, beginning with the entrance
- The importance of natural lighting
- The flexibility and multifunctionality of rooms
- The arrangement of classrooms and hallways that offer opportunities for privacy
- The importance of providing opportunities for teachers’ privacy
- The value and importance of acoustics and noise prevention



**FIGURE 5** Oasis (with plants) for meeting opportunities (by Peter Hübner). Reprinted courtesy of Asanger Press, Germany (Walden & Borrelbach, 2002).



**FIGURE 6** Welcoming entrance to a school (by Friedensreich Hundertwasser). Reprinted courtesy of Asanger Press, Germany (Walden & Borrelbach, 2002).

- The importance of providing safe, accident-proof toys and equipment in the schoolyard
- The importance of providing rich “sensory experiences”

But who will pay for such ideal schools? This objection can be met, for example, with reference to advanced methods of heat insulation, energy conservation, and cost-effective building construction. Above all, however, a building that is perceived by its occupants as being “theirs” is more likely to be protected against vandalism and disrepair and, therefore, may become more affordable in the long run. As a consequence, a “well-being school” does not necessarily have to be expensive.

### See Also the Following Articles

Environmental Assessment ■ Environmental Design and Planning, Public Participation in ■ Environmental Stress ■ Full-Service Schools ■ Person–Environment Fit

### Further Reading

- Ahrentzen, S., Jue, G. M., Skorpanich, M. A., & Evans, G. W. (1982). School environments and stress. In G. W. Evans (Ed.), *Environmental stress* (pp. 224–255). New York: Cambridge University Press.
- Barker, R. G., & Gump, P. V. (1964). *Big school, small school*. Stanford, CA: Stanford University Press.
- Bell, P. A., Greene, T. C., Fisher, J. D., & Baum, A. (2001). *Environmental psychology* (5th ed.). Fort Worth, TX: Harcourt Brace.
- Duke, D. L., & Trautvetter, S. (2001). *Reducing the negative effects of large schools*. Washington, DC: National Clearinghouse for Educational Facilities.
- Gifford, R. (2002). Educational environmental psychology. In *Environmental psychology: Principles and practice* (pp. 296–336). Colville, WA: Optimal Environments.
- Goldstein, A. P. (Ed.). (1996). *The psychology of vandalism*. New York: Plenum.
- Gump, P. V. (1991). School and classroom environments. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 691–732). Malabar, FL: Krieger.

- Klatte, M., Meis, M., Nocke, C., & Schick, A. (2003). Classroom acoustics: Acoustical conditions in schools and their impact on pupils. In M. Cudina (Ed.), *Proceedings of the First Congress of Alps Adria Acoustics Association and Third Congress of Slovenian Acoustical Society* (pp. 9–16). Ljubljana, Slovenia: Slovenian Acoustical Society.
- McAndrew, F. T. (1993). *Environmental psychology*. Pacific Grove, CA: Brooks/Cole.
- Moore, R. C. (1989). Playgrounds at the crossroads: Policy and action research needed to ensure a viable future for public playgrounds in the United States. In I. Altman, & E. H. Zube (Eds.), *Public places and spaces* (Vol. 10, pp. 83–120). New York: Plenum.
- Moos, R. F. (1979). *Evaluating education environments*. San Francisco: Jossey-Bass.
- Pablant, P., & Baxter, J. C. (1996). Environmental correlates of school vandalism. In A. P. Goldstein (Ed.), *The psychology of vandalism* (pp. 213–233). New York: Plenum. (Original work published 1975)
- Sanoff, H. (2001). *School building assessment methods*. Washington, DC: National Clearinghouse for Educational Facilities.
- Sommer, R., & Olsen, H. (1980). The soft classroom. *Environment and Behavior*, 12, 3–16.
- Walden, R., & Borrelbach, S. (2002). *Schulen der Zukunft: Gestaltungsvorschläge der Architekturpsychologie* [Schools of the future: Design proposals from architectural psychology]. Heidelberg, Germany: Asanger.
- Weinstein, C. S. (1979). The physical environment of the school: A review of the research. *Review of Educational Research*, 49, 577–610.



# School Psychology, Overview

Howard M. Knoff

Arkansas Department of Education, Little Rock, Arkansas, USA

1. Qualifying School Psychology as a Profession
  2. The Elements of an Effective School and the Contributions of Applied Psychology to Effectiveness
  3. What School Psychologists Do to Be Effective
  4. This Section of the Encyclopedia
  5. Conclusion
- Further Reading

## GLOSSARY

**credentialing and licensure standards** Standards that typically are the state-specific training, supervised practice, and continuing education criteria that result in a school psychologist being certified for independent practice either in the schools (credentialing) or in a community-based or private practice setting (licensure).

**functional assessment** Assessment method that focuses on either increasing or establishing appropriate behavior or decreasing or eliminating inappropriate behavior through a process that evaluates students' strengths and assets along with their weaknesses and deficits. The process involves using "authentic" or curriculum-based assessments that directly evaluate the skill sets or domains of interest that determine why students are or are not making progress in various areas, and then linking the assessment results directly to strategic interventions to address targeted problems or concerns.

**"No Child Left Behind" Act (NCLB) and Individuals with Disabilities Education Act (IDEA) legislations** These two federal laws guide much of general education and special education practices in schools and school districts. The NCLB is the current elementary and secondary education

law for the United States that guides expectations for students' academic progress, teacher quality, positive school climate and safety, and other annually evaluated outcomes. The IDEA guides and monitors the delivery of services that are provided to special education students who are identified as having one of 13 specific disabilities, requiring that they receive a free and appropriate public education (FAPE), an individualized education program, and other due process and educational rights.

**professional practice and ethical standards in school psychology** Standards that govern the professional practices and interactions of school psychologists as they (a) provide assessment, consultation, intervention, and evaluation services to students and parents, staff and administrators, and schools and other organizations; (b) interact with colleagues and other professionals; and (c) work in service delivery units in the field and/or school psychology training programs at the university level. The ultimate goal of these standards is the protection of the rights and integrity of all individuals and parties who are directly or indirectly impacted by provided services.

**school psychology** One of the four original specialty areas of professional psychology recognized by the American Psychological Association—specifically, organizational psychology, clinical psychology, counseling psychology, and school psychology. While there is a common science of psychology undergirding all four of these professional areas, school psychology focuses on the application of social and organizational, educational and learning, child and adolescent, normal and abnormal, and biological and ecological psychology to the school and schooling process.

**scope of school psychological services** School psychology services are provided to all students along a continuum, from (a) those who benefit from effective instruction without

the need for additional support services, to (b) those who need strategic interventions in order to make appropriate academic, social, emotional, and behavioral progress, to (c) those who need intensive or crisis management services in order to be successful.

School psychology is one of the four original specialty areas of professional psychology recognized by the American Psychological Association—specifically, organizational psychology, clinical psychology, counseling psychology, and school psychology. While there is a common science of psychology undergirding all four of these professional areas, school psychology focuses on the application of social and organizational, educational and learning, child and adolescent, normal and abnormal, and biological and ecological psychology to the school and schooling process. More specifically, school psychology works from a systems level to an individual level. At the systems level, it emphasizes the importance of creating safe and positive school environments that facilitate effective instructional processes that result in students' academic, social, emotional, and behavioral progress and success. At the individual level, it emphasizes the importance of guiding, intervening with, and evaluating students' health and mental health, and ways to help them develop and respond to protective factors such that they become independent learners, effective behavioral self-managers, and resilient in the face of life's passages and crises.

## **1. QUALIFYING SCHOOL PSYCHOLOGY AS A PROFESSION**

School psychology is a profession. Guided by two national professional associations, the National Association of School Psychologists and the American Psychological Association's School Psychology Division, school psychology has training standards, professional practice and ethical standards, credentialing and licensure standards, and continuing education expectations.

School psychology is a post-baccalaureate profession. In most states, school psychologists are certified by their state departments of education or psychological licensure boards for practice in the schools at the educational specialist (Ed.S.) level. Typically, this involves 60 graduate hours of defined coursework and a school-based or school-related internship. Some school psychologists earn the doctorate degree (Ph.D.,

Ed.D., or Psy.D.). Typically, this involves a minimum of three years of practice, a full year internship, a specialization area, and research culminating in a dissertation. Currently, about 75% of the profession practices at the Ed.S. level, and 25% practices at the doctoral level.

Relative to course work, school psychologists receive university training, clinical practice, and supervision in the following domains: (1) data-based decision making and accountability; (2) interpersonal communication, collaboration, and consultation; (3) effective instruction and development of cognitive/academic skills; (4) socialization and development of life competencies; (5) student diversity in development and learning; (6) school structure, organization, and climate; (7) prevention, wellness promotion, and crisis intervention; (8) home/school/community collaboration; (9) research and program evaluation; and (10) legal, ethical practice, and professional development. Demonstrating their adherence to a published set of standards and criteria, university training programs are either approved (at the doctoral and Ed.S. levels by the National Association of School Psychologists) or accredited (at the doctoral level by the American Psychological Association), thus providing an independent and objective level of accountability and quality control.

The practice of school psychology in the field is guided by professional practice and ethical standards. Professional practice standards, largely generated by the two national professional associations and state certification and licensure entities, govern the professional practices and interactions of school psychologists as they provide assessment, consultation, intervention, and evaluation services to students and parents, staff and administrators, and schools and other organizations. These activities include professional practices in the 10 training and practice areas listed above, but they also involve the professional interactions of school psychologists with the clients, the consultees, and others that they serve, and how they carry out their various responsibilities. In a broader sense, professional practice standards also address how school psychology units or departments (often in larger school districts in which multiple psychologists are employed to provide a full continuum of comprehensive services) should function to ensure appropriate and responsible practice.

Critically, it should be noted that school psychologists practice in a number of different settings. As they largely focus on the academic and social, emotional, and behavioral development of children

(starting generally at the preschool level) and adolescents (sometimes, for students with disabilities, through age 21), school psychologists are employed in schools; community mental health and other agencies; alternative education, juvenile justice, and prison settings; hospital and psychiatric treatment facilities; private practice; universities; and other specialized settings. In many of these settings, they interact not only with psychologists from other specialty areas, but also with teachers and administrators, social workers and counselors, speech and occupational/physical therapists, physicians and psychiatrists, and other health and mental health professionals.

Ethical practice standards, also generated by national associations and state entities, govern the professional practices and interactions of school psychologists relative to protecting the rights and integrity of all individuals and parties who are directly or indirectly impacted by provided services. Involving morally and legally responsible behavior, ethical practice standards are the “internal” guidelines that define, for individual psychologists and individual practice, appropriate professional scope, boundaries, communications, representations, interactions, and services. While addressing personal behavior, ethical practice standards are reinforced, at the state and national levels, by colleagues appointed to boards of ethical practice. When allegations of ethical violations are forwarded to these boards, they objectively gather the data and information necessary to make a judgment on the behavior, either dismissing the allegations as unfounded or requiring restitution and/or sanctions when the allegations are validated.

In professional practice, the most important ethical practice standard involves the need for school (and other) psychologists to practice within the scope of their training and expertise. Acknowledging that no one professional can be expert in all possible areas of practice, this standard ensures that psychologists provide services only in areas in which they can provide appropriate and effective services—independently or under supervision. For school (and other) psychologists, this is the ultimate protection for the consumer, a protection that is essential given the sometimes tenuous nature and enormous responsibility for safeguarding the psychological health and development of children, adolescents, and their families.

Credentialing and licensure standards are typically state specific, and they are written and enforced by a state’s Department of Education Certification Board and Department of Professional Practice, respectively. Most often, these standards are codified in law and/or

regulation, as both of these departments must submit them to their state legislatures for approval, and they require the psychologist to pass some type of written national or state exam.

The school psychology credential is the Department of Education’s official certification that an individual has received the appropriate university course work, training, experience, and supervision to practice in the schools as an independent practitioner. Often, school psychologists receive an initial or provisional credential, which is then upgraded to a full or independent practice credential after two or three years of full-time supervised practice in the field.

Licensure, meanwhile, is the Board of Psychology’s (or the equivalent) certification that an individual has received the appropriate university course work, training, experience, and supervision to practice in the community as a private, independent practitioner. Although licensure is most often conferred only at the doctoral level, there are currently more than 10 states in which a school psychologist can be licensed at the non-doctoral level. Whereas eligibility for school psychology credentialing generally requires university-based supervision, eligibility for licensure typically requires both university-based supervision and post-graduation supervision of practice in the field by a licensed psychologist. Licensure ultimately allows school psychologists to practice in virtually all settings—public and private—while credentialing generally limits the scope of practice to the schools. Some states, however, issue only a school psychology license, and some states require school psychologists to have a credential (even if they also have a license) to do school-based work.

If this were not confusing enough, the National Association of School Psychologists has issued national certification in school psychology (NCSP) to qualified school psychologists since 1988. Through their state credentialing regulations and procedures, approximately 30 states now recognize the NCSP for immediate credentialing. The NCSP has not only established a more uniform set of training standards for entry into the field, but it has also facilitated a level of reciprocity so that school psychologists can more easily move into new states and begin to practice professionally.

School psychology’s professional practice and ethical standards, and most state credentials or licenses, require school psychologists to maintain their clinical skills through continuing education. Typically, this involves taking a certain number of courses, spending a specific number of hours at workshops or in-service, and/or

engaging in documented types of professional activities or experiences over a defined period of time. The goal is to ensure that practitioners continue to have the most up-to-date training, experience, and professional standing in a field that is constantly changing and upgrading. This is essential so that they have the knowledge, skill, and confidence to make effective and evidence-based professional decisions over time, and to implement effective interventions and efficient services.

## **2. THE ELEMENTS OF AN EFFECTIVE SCHOOL AND THE CONTRIBUTIONS OF APPLIED PSYCHOLOGY TO EFFECTIVENESS**

Over the past two decades, there have been increasing expectations that schools and school districts demonstrate that all students are successfully meeting specific academic standards and outcomes on an annual basis. These expectations have been federally legislated through such laws as the “No Child Left Behind” Act (NCLB), which is the current elementary and secondary education law for the country. The NCLB is reinforced by the federal IDEA (Individuals with Disabilities Education Act) legislation. IDEA guides and monitors the delivery of services that are provided to students who are identified with one of 13 specific disabilities (e.g., specific learning disabilities, cognitive or intellectual disabilities, emotional or behavioral disturbances, deaf/hard of hearing or blind/visual impairment disabilities) in a number of specific outcome areas: the number of students with disabilities graduating from high school with a regular diploma versus dropping out with no diploma at all; the disproportionate representation of minorities identified in certain disability categories or recommended for school suspension or expulsion due to alleged disciplinary actions; the participation of students with disabilities in their districts’ high-stakes or state standards assessments and their demonstrated adequate yearly academic progress on those tests; and the percent of students with disabilities who are taught in general education classrooms for the majority of the school day. Beyond these focus areas, schools are now also mandated to provide a free and appropriate public education to all students; to demonstrate that they have positive, safe, and supportive school and classroom climates; to employ “qualified teachers” in all classrooms; to notify parents annually as to the data-based effectiveness of

their children’s schools; and to track and maintain high standards for all students regardless of gender, socioeconomic background, race, culture, language of origin, or perceived potential.

To accomplish all of these requirements, schools must identify, organize, and implement evidence- and research-based procedures that facilitate school effectiveness, and then engage in a process of formative and summative evaluation that results in continuous improvement. Existing research can be merged into an “effective schools blueprint” that specifies the characteristics and variables that predict students’ academic success. In Arkansas, for example, this blueprint has been organized with the following 11 primary standards:

Standard 1: Effective schools have committees and/or teams that support their professional development, curriculum and instruction, and parent and community outreach activities and school goals. Effective schools also have committees and/or teams that support students’ successful attainment of specific academic and social, emotional, and behavioral goals.

Standard 2: Effective schools have an accessible and public document that outlines the scope and sequence of all goals and objectives in all curricular areas taught in the school. This scope and sequence document is cross-referenced with state standards and benchmarks, and it is used as a formative evaluation tool to track student progress. In academic areas, the classroom-based use of curriculum-based assessment and/or measurement is used to track students’ progress along these scope and sequence goals and objectives.

Standard 3: Effective schools have a plan and implement a school-wide positive behavioral support system that includes: (a) social skills instruction for all students, and (b) a developmentally successful, school-wide accountability system that includes behavioral expectations connected to incentives and positive responses, and levels of inappropriate behavior connected to negative responses, consequences, or strategic interventions. Included in this plan are prevention and intervention approaches that ensure positive student behavior and safety in common areas of the school (e.g., hallways, bathrooms, buses, cafeteria, playgrounds) and that address potential and actual incidents of teasing, taunting, bullying, harassment, and physical aggression. Also included in this plan are periodic evaluations of school climate and student, staff, and parent satisfaction with the school-wide system.

Standard 4: Staff in effective schools receive ongoing training, evaluation, feedback, and supervision (when needed) in the development and implementation of (a) effective instructional lessons, (b) approaches that effectively link students' acquisition of skills to curricula and instruction, (c) early academic interventions for students when needed, and (d) more intensive classroom-based interventions that are recommended through data-based problem solving and consultation processes. All of these processes result in demonstrated student progress and success.

Standard 5: Staff in effective schools receive ongoing training, guided discussions, supervision (when needed), and follow-up in effective classroom organization and behavior management approaches, and in other approaches that maximize students' time on task, their academic engagement, and their effective use of allocated academic learning time. Staff also use effective and flexible grouping patterns for their students to maximize learning outcomes, and they create and maintain positive, safe, and cooperative learning environments.

Standard 6: Staff in effective schools receive ongoing training, evaluation, feedback, and supervision (when needed) in data-based problem solving and the functional assessment of students' academic and behavioral skill and progress. This training includes the analysis of student records and cumulative folder information, and a linking process between problem analysis and intervention. This training also is especially focused on students who need strategic and/or intensive interventions due to their difficulties in acquiring or demonstrating such skills taught through more routine instruction.

Standard 7: After completing functional assessments, staff in effective schools have the skills (or have access to consultation that can provide the skills) to implement more intensive behavioral and academic/instructional interventions. These intensive interventions are specified for all staff, and they receive ongoing training, evaluation, feedback, and supervision in their implementation.

Standard 8: Effective schools have a written, systematic problem-solving and functional assessment process that involves grade, department, and building-level teams that provide pre-referral interventions for students who are not making sufficient academic and/or social, emotional, or behavioral progress. Related service and other professionals (e.g., school psychologists, speech pathologists, social workers) are represented on these teams, and they provide consultation and intervention support to general education teachers as a first step in the pre-referral process.

Standard 9: Effective schools have an organized, formal, and ongoing process to articulate (or transition) students, academically and behaviorally, from grade to grade and teacher to teacher at the end of the year and during other school year transitions. This articulation process includes strategic and intensive needs interventions that have been developed and implemented for students who are not making sufficient academic and/or social, emotional, or behavioral progress. It also includes students' receiving special education, vocational, and other services. Finally, it includes students' transitioning into the workplace or into alternative (or similar) settings.

Standard 10: Effective schools have a written and systematically implemented parent and community outreach program. The parent outreach program includes activities that encourage parent participation in school activities and help parents to understand the school's goals, objectives, programs, and desired student outcomes. The community outreach program includes collaboration with social service, mental health, law enforcement, and other relevant agencies such that there is direct and indirect support for all school and schooling goals and objectives.

Standard 11: Effective schools have a data-management system and evaluation process, which is supported by computer technology whenever possible, to formatively and summatively track the school's progress toward its stated school, staff, parent, and student goals. For students receiving strategic or intensive services, this data base should be able to track the implementation of student academic and/or behavioral intervention plans as written.

Clearly, the 10 domains that form the foundation of school psychology are well represented within these standards, as are many of the applied areas of psychology (i.e., social and organizational, educational and learning, child and adolescent, normal and abnormal, and biological and ecological psychology) that contribute to the effectiveness of a school. Cross-referencing the 10 domains and the 11 standards defining effective schools, however, suggests a prototype representing a "scope and sequence" of school psychological service delivery. Specifically, the scope of school psychological services involves all school-aged students—typical students, students who need strategic interventions, and students who need intensive services. The sequence of these services involves data-based problem solving that includes functional assessment, consultation, and interventions.

While there often appears to be a dichotomy between general and special education services, school psychology views students along a continuum, from (1) those who benefit from effective instruction without the need for additional support services to (2) those who need strategic interventions in order to make appropriate academic, social, emotional, and behavioral progress, to (3) those who need intensive or crisis management services in order to be successful. For the first group, there is an assumption that effective instruction is occurring in the classroom, and that students are acquiring new skills on the way toward becoming independent learners (academically) and effective self-managers (behaviorally). School psychologists focus predominantly on primary prevention—helping schools to develop and maintain positive behavioral support systems, helping teachers to match effective instructional approaches delivered using evidence-based curricula to individual students and student group learning styles, and helping students to adopt and use healthy behavioral styles and patterns.

The strategic interventions group of students does not acquire skills as quickly or as easily as those in the first group, but with the right assistance and interventions, they can be successful. In order to know how to strategically intervene on behalf of these students, it is critical that school psychologists and other educators determine why they are having difficulty. This involves data-based functional assessment, because there can be a number of different explanations for a student's lack of progress and, thus, a number of different interventions. In brief, school psychologists, working here at the secondary prevention level, often organize their functional assessments by considering six important domains: the student, the teacher or instruction, the curriculum, the classroom (which includes the peer group), the school and district, and the home and community. The first three of these domains generally explain students' difficulties in causally related ways, while the latter three domains provide explanations that are correlationally related. Causally related factors (e.g., poor student motivation, ineffective instruction, inappropriate student–curriculum match) directly explain what has caused an existing skill deficit and exactly what interventions need to be implemented. Correlationally related factors (e.g., a peer group that ridicules academic achievement, large class sizes, poverty at home) relate to the target problems, but do not directly cause them. For example, while poverty may contribute to a student's lack of reading experience and

working vocabulary, it is these two latter factors that actually cause the problem.

Functional assessments, to the greatest degree possible, occur in the settings in which the desired or problematic behaviors are occurring. While focused on either increasing or establishing appropriate behavior or decreasing or eliminating inappropriate behavior, functional assessments look at both student strengths and assets along with student weaknesses and deficits. In addition, they also involve “authentic” or curriculum-based assessments that directly evaluate the skill sets or domains of interest (e.g., improved performance in the area of mathematical computations are assessed using math problems that are sampled from the textbook being used in the classroom or the standards being assessed on a state standards test). Finally, functional assessments must provide results that are “psychometrically sound”—that is, that are both reliable and valid so that appropriate interventions can follow. Ultimately, functional assessment results need to link directly to strategic interventions. In this way, interventions with a high probability of success that address the root or source of the problem will hopefully be implemented with successful outcomes.

Finally, intensive needs or crisis management students are not successful in typical school settings and environments without significant levels of support and intervention. Although the same six functional assessment domains should be investigated to determine their contributions to the existing challenges, specific students' characteristics often are the most directly relevant. Indeed, many of these most-challenging students have disabilities that impair their academic and/or social, emotional, or behavioral progress. Others have mental illnesses that similarly contribute. For these students, school psychologists typically work at the tertiary prevention level as part of a school-based multidisciplinary team or community-based interdisciplinary team, completing ongoing functional assessments that lead to comprehensive, multidimensional “wrap-around” or multisetting, 24-hour interventions. Ultimately, these interventions serve to support the student's progress in the least restrictive environment possible and in the least intrusive way, while also helping the student to avoid future crisis situations or, at least, to minimize their impact and frequency over time. Intensive needs students require a great deal of time, effort, coordination, and resources when they are present in schools. Often, schools must accommodate for or help these students to compensate for their

challenges—sometimes in different settings or alternative programs. At other times, schools must recognize that these students will make progress only at their own pace and in their own way. They can learn, but in ways that are more idiosyncratic than their typical peers.

As noted, school psychological services involve data-based problem solving that includes functional assessment, consultation, and interventions. This approach contrasts with the more traditional “refer, test, diagnose, and place” mode of operation. In this more traditional approach, school psychologists worked primarily at the secondary and tertiary prevention levels, intervening with students with pre-existing or emerging difficulties. While doing this, their assessments focused predominantly on the students, working to diagnose and label their deficits or pathologies so that they could be placed in appropriate special or therapeutic settings. Some of the limitations of this approach included (1) the absence of multisetting, multisource, and multi-instrument ecological assessments; (2) the assumption that a diagnosis or label leads to functional, appropriate intervention strategies and plans; and (3) the belief that where an intervention occurred was more important than what actually occurred.

Functional assessment, as described earlier, exists within the data-based problem-solving context involving problem identification, problem analysis, intervention, and evaluation. Completed to determine why a problem situation exists, functional assessment links directly to interventions that have the highest probability of increasing or decreasing specific behaviors or interactional patterns. In the more contemporary approach to school psychological services, what needs to be implemented (i.e., the interventions) is determined first. Where the services need to be delivered (i.e., the setting) is identified next, representing the best setting both for effective service delivery and to obtain the fastest results with the greatest degree of change. This more contemporary approach also focuses its attention on student outcomes and not student diagnosis or placement; its interventions focus on behaviors and not labels.

The completion of a functional assessment, with its associated intervention recommendations, however, does not ensure effective intervention. This is where consultation comes in. Consultation is a collaborative process whereby school psychologists work directly with “consultees”—teachers, parents, counselors, and others who have the primary responsibility for implementing recommended interventions. Through the consultation process, school psychologists ensure that interventions are acceptable to consultees before they

are attempted, that they have a high probability of effecting meaningful change or outcomes, that they are implemented with integrity, and that they can be successfully generalized to different people, settings, times, and situations. In working with school-based consultees, school psychologists have two goals: (1) to help to “solve” or remediate the referred problem situations, and (2) to enhance the skills of the consultees such that they are able to implement the interventions developed more independently over time—both with the original students of concern, and, ideally, for other students with similar concerns in the future.

Critically, consultation is not limited to students with academic or behavioral challenges. Paralleling the foundational areas of psychology, there are at least seven different areas of psychological consultation: organizational consultation, (group) process consultation, instructional consultation, behavioral consultation, mental health consultation, advocacy consultation, and community-based strategic consultation. Recognizing that school psychologists deliver services at the home and community, district and systems, team and grade, and peer group and individual student levels, it is possible that any of these consultation areas might be practiced.

Finally, as previously noted, successful school psychological services should be defined by successfully implemented interventions that result in appropriate student progress and outcomes. At the same time, as with the different areas of consultation, some interventions may focus on the home and community, the school district and its schools, the staff and the curriculum, as well as the peer group and its individual students. Thus, interventions that result in effective schools are as important as, and related to, effective interventions for and with individual students.

### **3. WHAT SCHOOL PSYCHOLOGISTS DO TO BE EFFECTIVE**

While an understanding of successful school psychological services is important, these services require the training, supervision, and practice of effective school psychologists. Within the context of consultation, argued here to be school psychologists’ primary role in the schools, the characteristics of effective school psychological consultants have been empirically derived. In short, effective consultants possess four distinct sets (or factors) of related characteristics and/or skills: they have effective

interpersonal skills, sound problem-solving skills, and good consultation process skills, and they practice in ethically and professionally appropriate ways. While the importance of the latter three areas have been discussed, it is significant that the interpersonal skill area represent the strongest of these four factors, accounting for the greatest amount of the variance. This is not surprising from a psychological perspective—whether organizational, social psychological, or even therapeutic. Indeed, without people who develop relationships in which they communicate productively, work collaboratively, demonstrate care and empathy consistently, and show a commitment to desired outcomes, problem solving will not be successful, consultation will not be acceptable, interventions will not be implemented with integrity, and schools will not be effective. This means that school psychologists recognize and attend to the fact that effectiveness involves both process and content. The process involves, interpersonally, positive interactions with others; the content involves, professionally, appropriate information, strategies, and interventions that are collectively discussed and carefully implemented.

But school psychologists need more than consultation skills to be effective. They also need to consider their attitudes and attributions, and numerous other skill and competency areas. Thus, expanding beyond the 10 domains of school psychological practice discussed earlier, effective school psychologists also demonstrate many of the following attitudes and attributions:

- Accept responsibility for all children and adolescents in their schools
- Focus on the “whole” or entire student, looking at both strengths and weaknesses
- Deliver services based on student need and outcome, not on student label, disability, or diagnosis
- Identify the progress and successes accomplished in all initiatives, while seeking to expand and strengthen those successes
- Are committed to continuing education, professional development, and skill expansion
- Look at all challenges as learning and continued professional development opportunities

Effective school psychologists also demonstrate the following skills and competencies:

- Focus on comprehensive services at the systems, team, small group, and individual levels
- Recognize the psychological and education impact of gender, race, culture, socioeconomic status, and sexual orientation

- Recognize the psychological and educational impact of poverty, divorce, abuse and neglect, and other life crises
- Understand the characteristics of effective instructional environments and consult with staff to ensure that all students consistently experience such environments during their school careers
- Have good problem-solving skills with which they link functional assessment to strategic intervention
- Use acceptable assessment and evaluation procedures
- Can develop, implement, and evaluate effective behavioral and curricular interventions and programs
- Are guided by and effectively use data
- Are dedicated to using evidence- and research-based procedures and strategies

In summary, effective schools, effective school psychological services, and effective school psychologists are crucial in order to successfully impact all students. Guided by theory, research, application, and practice, psychology has and will continue to have a major impact on schools at the organizational, small group, and individual student and staff levels. This section of the encyclopedia reviews some of the most important and relevant areas of applied psychology as practiced by school psychologists. Critically, many of these areas are not unique to school psychology; they are similarly important to other specialties in applied psychology. This demonstrates the breadth and depth of psychology as a whole. As the science of human behavior, its many applications present myriad opportunities to positively affect people’s lives.

#### **4. THIS SECTION OF THE ENCYCLOPEDIA**

The articles in this section of the encyclopedia are broadly representative of the many important topics relevant to school psychology. The articles, nonetheless, certainly are not exhaustive. The articles prepared can be clustered logically into six broad areas of practice or application.

The first cluster of entries focuses on the importance of looking at schools within the context of the community. Effective schools recognize that they are a microcosm of their communities and that they are strongest when building partnerships that embrace and involve community constituencies. Halsell Miranda addresses the importance of dealing with

children and families from diverse cultures, a critical issue given our heterogeneous society and the need for cultural sensitivity and cultural equity as applied to both school and community. Cowan extends this discussion by emphasizing the need to bring home-school collaboration to a conscious level.

Parents are their children's first teachers, and their support of the school process significantly impacts students' academic achievement. Parent involvement must be nurtured by our schools. But parent involvement does not necessarily mean that parents must be physically involved in school activities on school grounds. Parent involvement, at its core, means that parents support the educational process of their children and prepare them for school. In fact, if parents only did the following five things with their children, academic achievement would significantly improve across our country: promote their children's health and nutrition, read to their children and support their involvement in literacy activities at home, talk to their children so that their vocabulary (and brains) can develop and so that they feel emotionally supported, decrease their children's use of the television along with other media-based technologies, and support the school's instructional, including homework, processes.

Finally, Cowan also talks about school-community partnerships. Many times, our schools are not collaborating in effective and efficient ways with community resources. Collaboration with health and mental health agencies can increase students' health and readiness for school. Collaboration with government and public safety officials can increase school safety. And collaboration with volunteer agencies can bring additional resources into our schools. School-community partnerships reinforce for students that many individuals and groups are committed to and support their education. All of this helps students to recognize that they are not alone in this process, and that others are ready to celebrate their successes.

The second cluster focuses on the place of assessment and student evaluation in effective schools. Clearly, students need to be evaluated on an ongoing basis to ensure that they are mastering the academic goals and objectives embedded in the curriculum. As students' ability to management their own behavior increases their academic engagement, which then increases their potential for academic achievement, student evaluation must include behavioral evaluation as well as academic evaluation. Student evaluation becomes more focused on student assessment when

students are not mastering their skills or when their behavior is interfering with academic progress. At this point, assessments should use functional assessment approaches to determine why the difficulty exists and, strategically, what interventions will most effectively remediate the situation. As reinforced by Ortiz, all assessment should be nondiscriminatory assessment. That is, all assessments should be devoid of bias (e.g., gender, racial, cultural, language, and geographic bias) and administered in objective and standardized ways. Nondiscriminatory assessment is essential to the validity of all assessment results. Without valid results, the potential for effective intervention will be compromised.

The other assessment areas that are investigated to identify the source of a student's difficulties reflect the specific outcomes that are expected of students. In general, these focus on students' academic, social, emotional, behavioral, and, sometimes, vocational outcomes. Thus, assessments can be done in the areas of intelligence and cognition, academics and achievement, behavior and emotions, and vocational interest and readiness. In this section, Hintze discusses behavioral assessment in schools; Evan, Williams, and Schultz discuss behavioral observation in schools; Miller discusses neuropsychological assessment in schools; and Levinson discusses vocational assessment in schools.

The third cluster focuses on the importance of characteristics of students with special needs. As noted, some students do not acquire academic and/or behavioral skills and require specific or strategic interventions. Some of these students have specific disabilities that can be accommodated for in their classrooms through minor curricular adaptations or specific supportive approaches (e.g., preferred seating, allowing oral instead of written responses, additional time for tests, lecture notes written by the teacher, tape-recorded books). Other students have specific disabilities that interfere with their school success so significantly that they are identified as special education students under the federal Individuals with Disabilities Education Act (IDEA) and receive specific interventions through an individualized education program (IEP). Under IDEA, 13 different disability areas are specified, including those that cover learning, behavior, cognition, health impairments, speech and language, hearing, vision, autism, and traumatic brain injuries. Some states also include students who are gifted and talented here, recognizing that these students do not have disabilities, but they do have exceptional needs. Additionally, some specific problem areas, such as attention deficit

disorders, may qualify for classroom accommodations and, if very serious, as a disability under the “other health impaired” category.

Regardless, effective schools are aware of the possible disability areas that interfere with a student’s progress, how to assess them, how to respond to them through interventions, and how to include those students, to the greatest degree possible, in general education classrooms and in the general education curriculum. In this section, a number of these disabilities and related areas are presented: learning disabilities by Tilly; students with emotional and behavioral problems by Huberty; attention deficit disorders—school-based interventions by DuPaul and Tresco; and the teaching of gifted students by Feldhusen.

The fourth cluster focuses on the importance of academic interventions. Under the federal No Child Left Behind legislation, students’ academic progress and achievement is the primary goal of the effective school. In order to meet this goal, teachers must be trained to be effective instructional coaches, the curriculum must be well matched to student styles and desired outcomes, and students must be fully engaged while being taught at their instructional levels. Academic success occurs because of effective planning and sound pedagogical instruction for all students. Academic success, however, also comes because specific interventions are implemented for those students who need more attention, practice, and alternative approaches to learning.

While effective schools focus on well-defined areas of instruction (e.g., reading and literacy, mathematics, science, social studies), they also focus on teaching students the skills they need to be independent thinkers and learners. Thus, metacognitive instruction is needed. This involves teaching students the “learning how to learn” process so that they can apply this process to novel learning situations. Metacognitive training can have a greater impact on students than the skill-based lessons of the traditional curriculum. As the proverb goes, it is often better to teach a person to fish than to simply give him or her fish on a daily basis.

In this section, Rafoth discusses the prevention of academic failure, emphasizing the proactive need to make all students successful, rather than waiting for students to fail and then reacting to their failure. Rathvon outlines a number of areas of academic interventions that should be considered when students are not mastering skills in timely ways. Joseph focuses more specifically on reading interventions, a major area of concern, especially during students’ elementary

school years. Finally, Gettinger addresses the importance of study skills instruction, part of the metacognitive foundation that helps students to become independent learners.

The fifth cluster focuses on the importance of discipline, behavior management, and school safety. Clearly, if students do not have the behavioral skills that support the academic process (e.g., listening, following directions, working cooperatively in groups, demonstrating task persistence), they will have more difficulty with academic engagement and achievement. Similarly, if students do not have good interpersonal and conflict resolution skills, and if their schools are unsafe and unsettled, their necessary concern with resolving their conflicts and protecting their own welfare will take precedence over school work. Thus, positive, safe, and predictable school environments are an essential component of effective schools.

To provide guidance in this important area, Bear talks about the component parts of an effective school discipline and behavior management approach. Larson addresses this topic in a different way, focusing on school violence prevention. Furlong and his colleagues then investigate a more specialized situation, bullying and abuse on school campuses, that schools must respond to both legally and ethically. Finally, Elliott and Lang discuss the importance of social skills training as the anchor to a positive and safe school. Expanding briefly, just as students need to learn skills to be academically successful, so too do they need to learn prosocial skills in order to be behaviorally and interpersonally successful.

The sixth, and final, cluster focuses on the importance of health and mental health prevention and intervention. Once again, when students are not physically and mentally healthy, their school success potentially suffers. In this section, Copeland and Crepeau-Hobson discuss the importance of health promotion in the schools. Although instruction in this area often occurs in many schools, the lessons still must be actively reinforced by students, parents, and staff in order for them to be consistently practiced. Poland then addresses suicide intervention and schools. Unfortunately, the United States has a rate of student suicide that is higher than that of any other industrialized country. This topic, then, when combined with the health promotion article, will hopefully alert psychologists and educators to the importance of attending to students’ health and mental health needs and to the action steps that are needed, in the same context, to prevent student suicide.

## 5. CONCLUSION

School psychology dates back to the early 1920s relative to training and practice, and to the mid-1950s at a professional organizational level. While integrating research and clinical practice from a number of other psychology specialty areas, as well as from education, school psychology has its own research base that it can distinctly call its own. Similarly, while other psychology specialty areas do indeed contribute to the application of psychology in the schools, school psychologists have a primary professional responsibility to understand the culture, norms, and contexts of education so that the best interventions and services are used. School psychologists can best serve their schools, staffs, and students when they are problem solvers, consultants, and intervention specialists. In the end, even in the face of so many needs and so many challenges, school psychologists contribute both to the science and practice of effective schools and successful students.

### See Also the Following Articles

Achievement Tests ■ Behavioral Assessment in Schools ■ Behavioral Observation in Schools ■ Consultation Processes in Schools ■ Diverse Cultures, Dealing with Children and Families from ■ Educational Achievement and Culture ■ Educational and Child Assessment ■ Effective Classroom Instruction ■ Exceptional Students ■ Full-Service Schools ■ Gender and Education ■ Gifted Students ■ Health Promotion in Schools ■ Home-School Collaboration ■ Intelligence Assessment ■ Learning ■ Learning Disabilities ■ Learning Styles and Approaches to Studying ■ Nondiscriminatory Assessment in Schools ■ Special Education ■ Teaching Effectiveness ■ Transfer of Learning

### Further Reading

Arkansas Department of Education–Special Education. (2003). *A portfolio approach to conducting focused monitoring in selected Arkansas school districts*. Little Rock, AR: Arkansas Department of Education.

Berliner, D. C. (1988). Effective classroom management and instruction: A knowledge base for consultation. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options*

*for all students* (pp. 309–326). Silver Spring, MD: National Association of School Psychologists.

Bickel, W. E. (1999). The implications of the effective schools literature for school restructuring. In Reynolds, C. R., & Gutkin, T. B. (Eds.), *The handbook of school psychology* (3rd ed., pp. 959–983). New York: John Wiley & Sons, Inc.

Cawalti, G. (Ed.) (1995). *Handbook of research on improving student achievement*. Arlington, VA: Educational Research Service.

Chall, J. S. (2000). *The academic achievement challenge: What really works in the classroom?* New York: Guilford Press.

Gettinger, M., & Stoiber, K. C. (1999). Excellence in teaching: Review of instructional and environmental variables. In Reynolds, C. R., & Gutkin, T. B. (Eds.), *The handbook of school psychology* (3rd ed., pp. 933–958). New York: John Wiley & Sons, Inc.

Knoff, H. M. (2004). Inside Project ACHIEVE: A comprehensive, research-proven whole school improvement process focused on student academic and behavioral outcomes. In K. Robinson (Ed.), *Advances in school-based mental health: Best practices and program models*. Kingston, NJ: Civic Research Institute, Inc.

Knoff, H. M. (2000). Organizational development and strategic planning for the millennium: A blueprint toward effective school discipline, school safety, and crisis prevention. *Psychology in the Schools*, 37, 17–32.

Knoff, H. M., Hines, C. V., & Kromrey, J. D. (1995). Finalizing the Consultant Effectiveness Scale: An analysis and validation of the characteristics of effective consultants. *School Psychology Review*, 24, 480–496.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84, 30–43.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993/1994). What helps students learn. *Educational Leadership*, December/January, 74–79.

Ysseldyke, J., Dawson, P., Lehr, C., Reschly, D., Reynolds, M., & Telzrow, C. (1997). *School psychology: A blueprint for training and practice II*. Bethesda, MD: National Association of School Psychologists.

Ysseldyke, J., & Elliott, J. (1999). Effective instructional practices: Implications for assessing educational environments. In Reynolds, C. R., & Gutkin, T. B. (Eds.), *The handbook of school psychology* (3rd ed., pp. 497–518). New York: John Wiley & Sons, Inc.

Ysseldyke, J. E., & Christenson, S. L. (2002). *Functional assessment of academic behavior: Creating successful learning environments*. Longmont, CO: Sopris West.



# School-to-Work Transition

David L. Blustein, Justin C. Perry, and David B. DeWine

*Boston College, Chestnut Hill, Massachusetts, USA*

1. Overview
  2. A Historical and Sociopolitical Perspective
  3. The Psychological Realm
  4. Conclusion
- Further Reading

historical overview of the STW movement and describe its underpinnings in vocational education and career education. The relationship between the STW transition and psychological theory and research is then discussed. Conclusions are given for scholars and practitioners in vocational psychology.

## GLOSSARY

**connecting activities** Psychological and educational interventions designed to foster the connection between school and work.

**school-based learning** Educational experiences rooted in the classroom or school setting.

**school-to-work** Focusing primarily, but not exclusively, on non-college-bound youth, school-to-work (STW) generally refers to the transition from high school into the workforce and/or post-secondary training. This area of inquiry, which has explored the economic, moral, sociological, and psychological factors associated the STW process, has yielded important research findings along with suggestions for the improvement of educational and psychological practice.

**work-based learning** Educational experiences rooted in an occupational setting.

**work-bound youth** Those individuals who do not attend a 4-year college immediately after graduating from high school.

This article summarizes the school-to-work (STW) transition movement with respect to its social and economic origins, current knowledge base, and implications for research and practice. The authors first provide a

## 1. OVERVIEW

Broadly speaking, the focus on the school-to-work (STW) transition, which dates to the late 1980s, refers to a movement of educational, social, political, and public policy reform that is designed to help students transition from high school to the world of work or post-secondary training. The initial emphasis of this movement was directed toward the needs of students who do not graduate from 4-year colleges. However, as the movement evolved in the United States, it was expanded to include youth who were transitioning to college. This transition is a highly challenging and critical task for adolescents, with significant relevance to both academic and nonacademic developmental trajectories. The STW movement has attracted much interest in several disciplines, most notably life course sociology, counseling psychology, and life span developmental psychology.

Compared to other countries throughout the world, the overall lack of a systematic approach to the STW transition in the United States has resulted in its diminished capacity to prepare work-bound youth for entry into the labor market. Youth in the United States,

particularly those who do not attend a 4-year college, are typically left to navigate a period of “floundering” in which individual agency and the availability of resources play a major role in determining the level of risk and potential for an adaptive transition. In comparison, many European countries, such as Finland and Germany, have highly formal, unified, and institutionalized STW systems. Although these STW initiatives have a number of assets, including the reduction of floundering after high school, some critics have noted that these highly structured transition processes risk reinforcing social and economic inequality.

Current programmatic and research efforts on the STW transition stem from a number of interrelated problems. The most salient sources of impetus for STW initiatives were (i) the emergence of declining wages, beginning in the early 1970s, among skilled and unskilled workers; (ii) the increasing level of automation and computer technology, reducing the need for semi-skilled and unskilled workers; (iii) the growing disparities in wealth between college-educated and non-college-educated individuals; and (iv) the increasing debates about the quality, relevance, and standards of education. These issues eventually led to awareness by policy analysts, educators, and social scientists that the U.S. economy is experiencing a labor crisis of low skills and high wages. Jobs with reasonably decent wages that were traditionally available for high school graduates in industrial, manufacturing, and agricultural fields are no longer realistic opportunities for work-bound youth due to lack of availability and security. For adolescents from inner cities and rural areas, this crisis has become especially pronounced; for instance, the unemployment rate for urban youth of color is approximately double that of their European American peers attending schools in suburban areas. Nonetheless, numerous STW policy statements during the past decade have been directed toward students from all socioeconomic backgrounds and school systems, including the college-bound.

## **2. A HISTORICAL AND SOCIOPOLITICAL PERSPECTIVE**

The historical and social antecedents of the STW movement can be traced to various educational and policy-based ideas and programs during the past several decades. Soon after the turn of the 20th century, the classical approach to secondary education that focused on the core areas of language, math, science, and history

(which had been available primarily to the affluent) began to be replaced by a philosophy that sought to train an increasing proportion of students for employment roles, largely in industry and agriculture. This shift in the purpose of U.S. education gave rise to the vocational education movement and later to career education. Such a fundamental change presaged the current debate pertaining to the objectives and pedagogical approaches of today’s school systems.

In seeking to provide high school students with practical, hands-on skills that could lead to direct employment, the vocational education movement divided academic training along a vocational education versus college-preparation track. This bifurcation occurred mainly because of the changing demands of a mass-producing and consumer-driven society, in which employers increasingly needed workers with technical skills who could perform tasks efficiently. As a result, vocational education was often viewed as a second-rate educational alternative designed to meet the demands of industries, covertly perpetuating class boundaries. According to many educators and policy analysts, however, vocational education has not been a successful system. In order to rectify the dualistic systems of learning established by this movement, career education emerged as a rubric for a wide array of programs that share two interrelated objectives: to enhance the occupational relevance of education and to reduce the social class distinctions between vocational and academic tracks. The major mechanism of career education, therefore, was the infusion of a career development and occupational focus into all levels of the traditional academic curriculum.

In the early 1970s, career education received considerable public attention and funding. Although much of the financial resources and intellectual energy within this movement have dissipated, many of the ideas were integrated into the schools and have formed part of the conceptual framework for the STW movement. Specifically, the discussion of the vast economic and moral implications of existing divisions between college-bound and non-college-bound, a dialogue initially set in motion by John Dewey in the early 1900s, was clearly articulated within the career education movement. This discussion helped set the stage for the debates that are currently taking place with regard to the best ways of reconciling the contradictory aims of secondary school education in the United States. Namely, this contradiction revolves around the notion of academic rigor versus academic relevance, or the disconnection between real-world applications of learning and the conventional “skills and drills” taught in math, English, and science.

### 3. THE PSYCHOLOGICAL REALM

The literature in psychology that seeks to understand how the STW transition process occurs represents a critical body of knowledge that has the potential to inform subsequent theory and practice. A broad examination of the literature suggests that a wide array of definitions of an adaptive STW transition have been used. In general, social scientists have tended to rely on global indices of employment and unemployment or level of occupational attainment achieved in the labor market. In contrast, psychological investigations have pointed to definitions encompassing internal or relatively proximal characteristics of one's posttransition behavior, including job satisfaction and occupational choice congruence, development of adaptive coping skills on the job, and development of self-concept. Based on a review of recent literature on the STW transition, an adaptive transition may be best understood as including some internal psychological index of the degree of individual satisfaction one experiences with respect to work and nonwork roles as well as more global and objective indices of how well the individual is faring in attaining access to opportunities for growth, advancement, and further training.

When considered collectively, the existing research has revealed that four broad factors are associated with an adaptive STW process:

- Competencies in basic academic skills and vocationally relevant domains
- Self-initiative, planfulness, flexibility, and purposefulness
- An active and supportive relational environment, including family, positive peers, teachers, counselors, and coworkers and supervisors
- An educational environment that offers a clear connection to the world of work

Post hoc observational studies have revealed several key psychological attributes associated with successful STW transitions that represent these characteristics. Furthermore, studies have found that families who exhibit characteristics such as cohesiveness, expressiveness, and democratic decision making tend to foster important career-oriented learning processes for adolescents.

Researchers have also shown that the sequence of occupational activities in which a young adult engages can affect his or her psychological well-being. For example, workers from European samples who move from employment to additional training exhibit higher levels of psychological well-being, whereas young

adults who move from school into the military, with no prior employment history, experience lower levels of psychological well-being.

Finally, longitudinal research data indicate that young adults who experience increased floundering are at risk for higher levels of psychological stress. Floundering refers to employment instability in which a young adult or adolescent changes jobs with little or no support or direction. Although a certain amount of occupational exploration can influence an adaptive transition by increasing person-environment awareness, too much instability has been linked to negative psychosocial outcomes.

As an integral aspect of gaining a richer understanding of transitioning from school to work, social scientists also have sought to understand how membership in a marginalized group might influence the STW process. Specifically, scholars have investigated how youth with disabling conditions, young women, and youth of color negotiate the complex transition in the context of less access to the educational and economic opportunity structure. Despite the sparseness of research in these areas, initial studies have suggested that disabling conditions, gender, and visible racial and ethnic minority status play a discernible and aversive, although not uniform, role in the STW transition. That is, the available evidence indicates that the transition from school to work offers a more pronounced set of challenges and barriers for these minority groups compared to members of the dominant culture.

### 4. CONCLUSION

One of the primary implications of the STW movement within applied psychology has been its explicit focus on the needs of poor and working-class students. Indeed, discourse and research initiatives stimulated by the STW movement have placed poor and working-class students at the forefront of thinking in vocational psychology. Moreover, the STW movement has highlighted the need for an active consideration of social class, race, and other social barriers in constructing educational policy, building effective theories, and designing research.

Furthermore, research concerning the STW transition has informed counseling practitioners of the need to examine their own implicit class biases in how they conceptualize work and in how they design interventions. Traditional office-based counseling practice may no longer be an effective way to engage work-bound youth in the process of negotiating adaptive transitions

to work and post-secondary training. The STW movement has highlighted the need for scholars and practitioners in vocational psychology to develop a broader understanding of issues pertaining to equity, structural aspects of the opportunity structure, and the social context of work, as well as the ability to communicate with larger systems, such as school and industry.

### See Also the Following Articles

Holland's Theory (Vocational Personality Types)  
 ■ Indecision, Vocational ■ Vocational Assessment in Schools ■ Vocational Interests ■ Vocational Psychology, Overview ■ Work Adjustment

### Further Reading

- Blustein, D. L., Juntunen, C. L., & Worthington, R. L. (2000). The school-to-work transition: Adjustment challenges of the forgotten half. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed., pp. 435–470). New York: Wiley.
- Charner, I., Fraser, B. S., Hubbard, S., Roger, A., & Horne, R. (1995). Reforms of the school-to-work transition: Findings, implications, and challenges. *Phi Delta Kappan*, 77, 58–59.
- Gray, K. C., & Herr, E. L. (1998). *Workforce education: The basics*. Needham Heights, MA: Allyn & Bacon.
- Marshall, R., & Tucker, M. (1992). *Thinking for a living: Education and the wealth of nations*. New York: Basic Books.
- Olson, L. (1997). *The school-to-work revolution*. Reading, MA: Addison-Wesley.
- O'Neil, H. F. (1997). *Workforce readiness: Competencies and assessment*. Mahwah, NJ: Erlbaum.
- Pautler, A. J. (Ed.). (1994). *High school to employment transition: Contemporary issues*. Ann Arbor, MI: Prakken.
- Pauly, E., Kopp, H., & Haimson, J. (1995). *Homegrown lessons: Innovative programs linking school and work*. San Francisco: Jossey-Bass.
- Resnick, L. B., & Wirt, J. B. (1996). *Linking school and work: Roles for standards and assessments*. San Francisco: Jossey-Bass.
- Savickas, M. L. (1999). The transition from school to work: A developmental perspective. *Career Development Quarterly*, 47, 326–336.
- Secretary's Commission on Achieving Necessary Skills. (1992). *Learning a living: A blueprint for high performance*. Washington, DC: U.S. Department of Labor.
- Wilson, W. J. (1996). *When work disappears: The world of the new urban poor*. New York: Random House.
- Worthington, R. L., & Juntunen, C. L. (1997). The vocational development of non-college-bound youth: Counseling psychology and the school-to-work transition movement. *The Counseling Psychologist*, 25, 323–363.



# School Violence Prevention

Jim Larson

University of Wisconsin, Whitewater, Wisconsin, USA

1. Introduction
  2. Definition of School Violence
  3. Historical Perspective
  4. Effects of School Violence
  5. Organizing Structures for School Violence Prevention
  6. An Integrated Approach to School Violence Prevention
  7. Conclusion
- Further Reading

## GLOSSARY

**bullying** Continual infliction of harm or threat of harm, physical or psychological, within the context of a clear imbalance of power.

**hostile excluding** A form of nonphysical aggression, more common in females, that involves efforts to cause psychological distress through directed exclusion from a social group or desirable social activities.

**protective factor** Characteristics of the individual or environment that serve to moderate the impact or improve people's resistance to human dysfunction.

**risk factor** Characteristics of the individual or the environment that serve to increase the probability of onset, degree of severity, or duration of human dysfunction.

**zero tolerance** A school discipline policy intended primarily as a method of sending a message that certain violence-related behaviors will not be tolerated by punishing both major and minor offenses severely.

School violence prevention involves the application of universal, selected, and indicated evidenced-based programs and procedures in the school setting in order to reduce the incidence of interpersonal violence in all its forms.

## 1. INTRODUCTION

Efforts to prevent school shootings such as those that occurred at Columbine High School and elsewhere are only one facet of a much larger continuum of school violence prevention. The statistical probability of being shot by a fellow student is extraordinarily remote, but the day-to-day likelihood of being threatened by a bully, punched by a classmate, or victimized by mean-spirited teasing is not. Understanding and appreciating the myriad contextual, intrapersonal, developmental, and educative aspects of maintaining a safe and supportive learning climate for both students and adults is the focus of this article.

## 2. DEFINITION OF SCHOOL VIOLENCE

School violence involves all the behaviors along the full continuum of interpersonal physical, verbal, and psychological aggression that interfere with the rights

of all children to learn and the rights of all adults to work in a safe educational environment. School violence is much more than just the rare headline-grabbing multiple homicides; it involves the quality of the day-to-day interpersonal relationships in the context of the school setting—relationships that contribute to or detract from everyone's sense of personal safety and well-being. Teasing, bullying, threatening, hostile excluding, drug dealing, and fighting are example behaviors of concern. This definition also recognizes that the continua of physical, verbal, and psychological aggression have intensity, frequency, and emotional and physical injury plot points. For example, mean-spirited teasing or bullying directed at a target student can be examined with regard to the intensity of the interaction, the frequency of occurrence, and the emotional and physical consequences to the victim.

Additionally, it is important to understand that school violence may also be viewed from a skills-deficit perspective. A child who is resorting to physical aggression to resolve a concern may be demonstrating a lack of skill in social problem solving and anger management. From this viewpoint, school violence can also be defined as an index measuring the deficiency of prosocial skills and nonviolent conflict resolution strategies among individuals in the context of a school setting. Framing the definition in this fashion has useful implications for prevention and intervention policies and procedures.

### 3. HISTORICAL PERSPECTIVE

For many educators responsible for student safety before the 1970s, preventing school violence meant little more than administratively ridding the school of the most potentially violent students. When efforts to intervene failed, children and youth with chronic aggressive anger problems, gang affiliations, or other problematic characteristics were summarily dismissed from school. Subsequent federal legislation, such as Public Law 94-142 and the Education of All Handicapped Children Act of 1975 (now Public Law 105-17, the Individuals with Disabilities in Education Act Amendments of 1997), and Supreme Court decisions such as *Goss v. Lopez* changed the way students with challenging behaviors needed to be addressed. The 14th Amendment to the Constitution holds that no state shall “drive any person of life, liberty, or property, without due process of law.” Ruling in

the 1975 *Goss v. Lopez* case, the Supreme Court held that to deprive an individual of an education is to deprive that individual of a property right. Barring or otherwise excluding any individual from a free and appropriate public education without due process was determined to be unconstitutional. The effect of Public Law 94-142 and the *Goss* decision was to place the obligation on the schools to address the needs of everyone who desired a public education, including those with challenging behaviors. The days of simply excluding aggressive students without due process were over.

In 1978, the federal government published the first large-scale report on school safety, *Violent Schools—Safe Schools: The Safe School Study Report to Congress*, and the nation's eyes were opened to a new and disquieting feature of public education. Among the findings were that in a given month at school, approximately 282,000 (1.3%) students were physically attacked in secondary schools, almost 8% of urban junior and senior high school students missed at least 1 day of classes because they were afraid to go to school, and nearly 5200 of the nation's million secondary school teachers were physically attacked. For most middle-class Americans, however, the issues of violence in school remained pretty much an urban concern: The Blackboard Jungle was a long way from the comparative safety of suburbia. The major urban school districts spent the 1970s and 1980s fighting the escalating street crime that spilled into their hallways using every method at their disposal, including locker searches, armed guards, and metal detectors. In 1984, a presidential directive established the National School Safety Center to serve as an information clearinghouse and resource center for beleaguered schools. In 1986, Boston public health physician Deborah Prothrow-Stith introduced the nation's first dedicated violence prevention curriculum, Violence Prevention Curriculum for Adolescents, into the Boston school system. In 1987, the Seattle-based company Committee for Children followed suit with its program for younger students, Second Step: A Violence Prevention Curriculum. In 1988, the Milwaukee Public Schools initiated the Violence Reduction Program, the nation's first districtwide, multisystemic effort to address the mounting problem of student violence through curricular, counseling, and parent training efforts. In suburban, small town, and rural school districts outside of the larger cities, however, schools had not yet even begun to lock their doors.

The relative complacency of the nonurban U.S. public school was changed forever on January 17, 1989,

when Patrick Purdy, 26, armed with an AK-47 assault rifle, opened fire on a playground at Cleveland Elementary School in Stockton, California. Five children died and 29 children and one teacher were wounded before Purdy killed himself. Whereas this assailant was an adult, it was the homicides perpetrated in the school setting by students in communities such as Jonesboro, Arkansas, Springfield, Oregon, and Littleton, Colorado, that provided the greatest impetus to the development of school violence prevention programs. Since the early 1990s, concerns both real and imagined about school shootings have fueled both research and product development in the area of school violence prevention.

Somewhat lost in all of the media coverage of these high-profile homicides were the historical facts regarding school violence. National surveys by the Bureau of Justice Statistics and the National Center for Educational Statistics indicated that there was a 46% decline in the nonfatal violent victimization rate (simple assault and more violent crimes, including rape, robbery, and aggravated assault) in schools between 1992 and 2000 among students ages 12–18. For all categories of crime victimization except theft, young people were safer at school than away from it. The number of teachers who were threatened with violence has also declined during the past decade, although the number of teachers who were actually assaulted has remained steady at approximately 4%. Data from the National School Safety Center's School-Associated Violent Deaths database showed that the incidence of single-victim homicide in schools also steadily declined during the past decade. Indeed, during the 1998–1999 school year that included the Columbine shootings, U.S. schools had the third lowest total of school-related homicides since 1992–1993. The statistic that increased during that same period was multiple homicide events in the school setting (Table I).

Although the number of homicides and serious criminal victimization in schools continues to decline, the same cannot be said for other forms of school violence. For instance, student reports of bullying are on the increase. It is estimated that during their school careers, one in three students will be involved in school bullying incidents as a victim, perpetrator, or both. Conversely, the odds of being killed in a school shooting have been estimated to be approximately 1 in 2 million, or approximately the same as being struck by lightning. This has implications for designers of violence prevention programs, who must account for base rates along the continuum of school violence if the

**TABLE I**  
**School Deaths Attributable to Single and Multiple Homicides**

<i>School year</i>	<i>Deaths by single homicide</i>	<i>Deaths by multiple homicide</i>
1992–1993	42	5
1993–1994	42	4
1994–1995	17	0
1995–1996	20	10
1996–1997	21	4
1997–1998	18	17
1998–1999	9	17
1999–2000	16	6
2000–2001	16	2
2001–2002	2	0

effort is to be meaningful in the lives of most students. Planning effectively for violence prevention takes into account that time and money spent preparing for a monumentally unlikely event at the expense of real day-to-day violence and victimization is both scientifically and fiscally indefensible.

#### 4. EFFECTS OF SCHOOL VIOLENCE

How can a student be expected to concentrate on academic learning when the potential for physical assault or the next round of mean-spirited teasing is just around the corner? What is it like to dread recess or suppress the urge to drink water for fear of the need to later have to venture into a dangerous restroom? How can a teacher possibly provide his or her full measure of skill when there is good reason to be frightened of some of the students?

Violence in the school setting has multiple physical, psychological, and educational effects on individuals and has a deleterious effect on the learning climate for everyone. Sociobiological theories of motivation posit that humans have four basic needs: drives to acquire, bond, learn, and defend themselves from harm. Students and teachers who must spend time and cognitive energy on concerns for their own safety cannot simultaneously devote their motivation and attention to the educational process in the most productive manner. Safe school environments allow students the freedom to attend to learning tasks with greater creativity, academic risk taking, and peer affiliation. Research from the Effective

Schools movement of the 1970s and, recently, from Project ACHIEVE has empirically demonstrated the positive effects on academic climate and student achievement through broad-scale measures to create safe, disciplined learning environments. Project ACHIEVE is a nationally known and evidence-based school reform and improvement program that has been implemented in more than 1500 school or districts throughout the United States. It has been designated as a Model Prevention Program by the Office of Juvenile Justice and Delinquency Prevention in the U.S. Department of Justice; the Collaborative for Academic, Social, Emotional Learning; and the Substance Abuse and Mental Health Services Administration of the Department of Health and Human Services.

## 5. ORGANIZING STRUCTURES FOR SCHOOL VIOLENCE PREVENTION

Effective school violence prevention programs are multisystemic endeavors that are integrated into the fabric of the educational process, not simply added on as time or budget permit. Planning and implementing procedures for the safety and emotional well-being of students are no less thoughtfully or rigorously studied than any other critical aspect of the educational process. Prevention research performed during the past three decades has emphasized the development of stakeholder partnerships. Individual classrooms exist in buildings, which exist in school systems located in communities. Consequently, the violence prevention process should have an identifiable presence and linkage across community, school district, and school building levels of policy and organization. Table II provides some of the basic roles and functions in school violence prevention at these levels.

## 6. AN INTEGRATED APPROACH TO SCHOOL VIOLENCE PREVENTION

School violence prevention measures aim to (i) prevent the onset of violent behavior in the general school population, (ii) identify and intervene with those students who are at risk for violent behavior, and (iii) reduce the frequency or intensity of violence in those students who are currently engaging in violent

**TABLE II**  
**Basic Levels, Roles, and Functions in School Violence Prevention**

---

<b>Community level</b>
Address community risk factors associated with youth violence
Enact and enforce juvenile criminal codes
Enact and enforce strong antidrug and firearm laws
Establish and coordinate community emergency response procedures
Provide emergency training expertise to school staff
Provide expertise and coordinate school security assessments
Provide safe after-school activities for all students
Provide expertise in research-supported prevention measures
<b>School district level</b>
Provide school board mandate for safe schools
Secure funding through budgetary process
Facilitate coordination with community resources
Ensure appropriate school mental health services
Develop policies for school security and emergency response
Facilitate staff training in:
Classroom discipline and problem-solving methods
Antibullying procedures
Instructional design and curricular modification
Emergency/security procedures
<b>School building level</b>
Establish representative school safety planning team
Conduct broadly based needs assessments
Establish primary, secondary, and tertiary prevention plans
Enact collaboratively developed school discipline plan
Practice emergency response procedures
Facilitate ongoing staff development

---

behavior. Compared to more reactive approaches, such as “zero tolerance” with its sanctions-based emphasis on suspension and expulsion, a prevention approach is primarily skills based and education oriented. Prevention approaches do not completely eschew the use of exclusionary procedures, but such reactive measures are viewed more as a last line of response and typically as a protective rather than intervention measure. Prevention approaches draw from numerous literature bases, including those from risk and protective factors, social context behavior, child development, and learning theory, to create a school

ecology conducive to healthy interactions and non-violent conflict resolution.

Because students learn and interact in several social contexts and with numerous individuals, these procedures are often multiply targeted. For instance, a child with anger control problems may receive anger management skills training, and simultaneously his or her parents may be engaged in parent training while his or her teacher is learning more effective classroom discipline strategies. Targets for school violence prevention include the student, teacher, family, classroom, school, school district, neighborhood, community, state, and national policy.

Fully integrated prevention programs target risk factors and skill development in multiple contexts and among multiple populations. In doing so, programs use universal, selected, and indicated prevention procedures.

### **6.1. Universal Prevention Measures**

When school violence prevention targets an entire population, such as a whole classroom or school, the endeavor is known as primary prevention and the measures used are referred to as universal prevention measures. Universal prevention measures are designed for all the individuals in the targeted population—students, teachers, or parents—regardless of each individual's risk status. At the classroom level, a teacher may provide the students with curricular instruction on the subjects of conflict resolution strategies, ways to manage anger, or safe handling of firearms. Because of its broad reach, universal classroom prevention curricula place primary emphasis on teaching the students about risk and prevention at a basic knowledge level rather than actually training them in complex behavioral skills. At the preschool to early primary levels, students may acquire problem-solving language and learn to name and identify feeling states in themselves and others. In the later elementary years, students may be taught to understand procedures for social problem solving, conflict resolution, and anger control.

With the whole school as a target for primary prevention, a school administration may offer in-service workshops to its teaching staff to bring them up to date on the latest techniques in classroom discipline skills or to help them learn to diversify instruction for behaviorally challenging students. A universal prevention measure that targets use of the physical environment may involve brightening the common areas with

new décor and additional lighting, modifying class changing times to reduce hallway crowding, or locking entranceways after the start of school.

### **6.2. Selected Prevention Measures**

Within school populations are individuals or groups that, because of personal characteristics or environmental factors, have elevated risk. Efforts to meet the needs of these individuals are known as secondary prevention, and the procedures used are referred to as selected prevention measures. One of the goals of secondary prevention in the school setting is to identify those individuals within a population who may be engaging in precursor behaviors that have predictive value for the onset of later, more serious concerns. Physically aggressive or bullying behavior, oppositional behavior, lying, or stealing, when observed in kindergarten or early primary-level children, are known to have a measure of predictive validity for later, more serious violent or antisocial behavior. In addition to receiving universal prevention measures, these children select themselves out for more intensive skill development.

Central to the successful application of selective prevention measures is the valid and reliable identification of students who will most likely benefit from the intervention. This is particularly important in situations in which there is significant environmentally based risk (e.g., a large population of low socioeconomic status children) and limited human resources to facilitate intervention. Multiple-gating methods allow classroom teachers and parents to use instruments such as risk-behavior checklists to narrow the population to those students or groups of students most likely to benefit from a particular selected measure.

Once identified, students receiving selected prevention measures focus on behavioral skill development, typically in an individual or small group setting under the direction of a school psychologist or guidance counselor. In these sessions, students learn to apply behavioral skills such as social problem solving, anger control, consequential thinking, or nonviolent conflict resolution to simulated scenarios in the training setting. In contrast to universal prevention, selected prevention measures go beyond merely knowing about violence prevention to actually learning and practicing critical behavioral skills, both in and out of the training setting.

Bullying prevention programs are examples of a selected measure applied at the whole school level. Longitudinal research indicates that students who

engage in bullying behavior in the elementary school are at high risk for continuing antisocial behavior in later life. In order to address the needs of both the bully and the victims of bullying, empirically supported bullying prevention programs train teachers, administrators, and students in role-appropriate identification, resistance, and effective intervention strategies.

### 6.3. Indicated Prevention Measures

The mission of the public schools includes providing an education to all students, regardless of their individual risk characteristics. Tertiary prevention services are designed to address the needs of students who display persistent, severely aggressive behaviors that have been unresponsive to universal or selected prevention procedures. These students are candidates for indicated prevention measures. Indicated measures are designed to reduce the frequency or severity of aggressive behaviors in the school setting—in effect, to keep the problem from worsening and protect the student and those around him or her. A tertiary prevention plan is typically formulated by a team of school and community professionals (e.g., school psychologists, mental health professionals, and agency representatives) working collaboratively with the student's parents. The student frequently receives a variety of services in and out of the school setting in a “wraparound” organizational structure. Multiple supports and services are “wrapped around” the student and, in many cases, the family in an effort to prevent problems from advancing to a more serious level. Indicated measures may involve skills training and individualized academic programming in the general or alternative school setting paired with family and community interventions.

## 7. CONCLUSION

Preventing school violence is a critical endeavor that requires training, organization, and ongoing commitment on the parts of those individuals responsible for its implementation. A fully integrated approach to school violence prevention brings together an understanding of childhood risk factors, skilled school and community professionals, and evidence-supported universal, selected, and indicated procedures to create a safe learning environment for students and staff members.

### See Also the Following Articles

Aggression and Culture ■ Bullying and Abuse on School Campuses ■ Emotional and Behavioral Problems, Students with ■ School Discipline and Behavior Management

### Further Reading

- Aber, J. L., Brown, J. L., & Jones, S. M. (2003). Developmental trajectories toward violence in middle childhood: Course, demographics, differences, and response to school-based intervention. *Developmental Psychology*, 39, 324–348.
- Brock, S. E., Lazarus, P. J., & Jimerson, S. R. (Eds.). (2002). *Best practices in school violence prevention and intervention*. Silver Spring, MD: National Association of School Psychologists.
- Brock, S. E., Sandoval, J., & Lewis, S. (2001). *Preparing for crises in the school: A manual for school crisis response teams* (2nd ed.). New York: Wiley.
- Elliott, D. S., Hamburg, B. A., & Williams, K. R. (Eds.). (1998). *Violence in American schools: A new perspective*. New York: Cambridge University Press.
- Hawkins, D. J., Herrenkohl, T., Farrington, D. P., Brewer, D., Catalano, R. F., & Harachi, T. W. (1998). A review of the predictors of youth violence. In R. Loeber, & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 106–146). Thousand Oaks, CA: Sage.
- Nohria, N., Lawrence, P., & Wilson, E. (2001). *Driven: How human nature shapes our choices*. San Francisco: Jossey-Bass.
- Olweus, D. (1994). *Bullying at school: What we know and what we can do*. Malden, MA: Blackwell.
- Pepler, D. J., & Slaby, R. G. (1994). Theoretical and developmental perspectives on youth and violence. In L. D. Eron, J. H. Gentry, & P. Schlegel (Eds.), *Reason to hope: A psychosocial perspective on violence & youth* (pp. 27–58). Washington, DC: American Psychological Association.
- Shafii, M., & Shafii, S. L. (Eds.). (2001). *School violence: Assessment, management, prevention*. Washington, DC: American Psychiatric Publishing.
- Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology*, 39, 189–200.
- Walker, H. M., & Epstein, M. H. (Eds.). (2001). *Making schools safer and violence free: Critical issues, solutions, and recommended practices*. Austin, TX: Pro-Ed.
- Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M. J. (1996). Integrated approaches to preventing antisocial behavior patterns in school-age children and youth. *Journal of Emotional and Behavioral Disorders*, 4, 194–209.



# Self-Confidence in Athletes

Robin S. Vealey

Miami University, Oxford, Ohio, USA

1. Definitions of Self-Confidence
  2. Importance of Self-Confidence to Athletes
  3. Sources of Self-Confidence in Athletes
  4. Self-Confidence within Sport Teams
  5. Measuring Self-Confidence in Athletes
  6. Strategies to Enhance Self-Confidence in Athletes
- Further Reading

*self-efficacy* Beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments.

*team confidence* Shared beliefs of team members about the ability of the team to perform successfully.

Self-confidence in athletes is an athlete's belief or degree of certainty that he or she has the ability to perform successfully in sport.

## GLOSSARY

*home advantage* Research finding that home teams win more than 50% of their competitions played under a balanced home-and-away schedule, often cited as a source of confidence for athletes.

*optimism* Tendency to expect the best possible outcome or dwell on the most hopeful aspects of a situation.

*psychological momentum* Athletes' perceptions that something has occurred that increases their probability of success, which typically creates a surge of confidence.

*SC-cognitive efficiency* An athlete's belief or degree of certainty that he or she can mentally focus, maintain concentration, and make effective decisions to perform successfully in sport.

*SC-physical skills and training* An athlete's belief or degree of certainty about his or her ability to execute the physical skills necessary to perform successfully in sport.

*SC-resilience* An athlete's belief or degree of certainty that he or she can regain focus after performance errors, bounce back from performing poorly, and overcome doubts, problems, and setbacks to perform successfully in sport.

## 1. DEFINITIONS OF SELF-CONFIDENCE

There are many terms used in the literature that relate to self-confidence. Self-efficacy is defined as beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments. Self-confidence is the belief that one can successfully execute a given activity, and the sport-specific term sport-confidence has been used to describe an athlete's belief or degree of certainty that he or she has the ability to perform successfully in sport.

Research indicates that sport-confidence is multidimensional, or that confidence in athletes may be subdivided into specific types of confidence. Sport-confidence in physical skills and training (termed SC-physical skills and training) is defined as an athlete's belief or degree of certainty about his or her ability to execute the physical skills necessary to perform

successfully in sport. SC-cognitive efficiency is defined as an athlete's belief or degree of certainty that he or she can mentally focus, maintain concentration, and make effective decisions to perform successfully in sport. SC-resilience is an athlete's belief or degree of certainty that he or she can regain focus after performance errors, bounce back from performing poorly, and overcome doubts, problems, and setbacks to perform successfully in sport. Thus, self-confidence in athletes is the summation and integration of not only athletes' beliefs in their abilities to execute physical sport skills but also their beliefs in their abilities to be mentally efficient and resilient in relation to the unique competitive demands of competitive sport.

Although self-confidence is typically defined as a belief system (beliefs about one's ability to succeed as an athlete), it also can be viewed and assessed as an emotion. Research with athletes shows that "confident" is a positive emotion described as optimal to enhance sport performance. Similarly, "uncertain," which denotes a lack of confidence, is a negative emotional feeling described by athletes as very dysfunctional for their performance. Thus, self-confidence in athletes is a belief as well as an emotional feeling that significantly influences how they think and feel and thus perform.

Overconfidence is a popular term often confused as meaning that athletes have too much confidence. However, true self-confidence involves believing one can be successful, with the realistic understanding of the effort and perspective needed to be successful. From this perspective, athletes can never have too much true self-confidence. Thus, overconfidence is not really a problem with confidence, but rather it is a sense of complacency about the need to train and exert effort to be successful. Overconfidence is an overestimation of one's abilities and need for expended effort and an underestimation of the situational demands that one will face. Feelings of complacency have been identified by athletes as extremely dysfunctional to their performance.

Certain terms are sometimes confused, or assumed to be synonymous with, self-confidence. Perceived competence and perceived ability refer to athletes' perceptions about how much ability or competence they possess in a certain achievement domain (e.g., soccer). However, perceived ability/competence focuses on the skills that athletes have, whereas self-confidence focuses on athletes' beliefs about what they can do (e.g., perform successfully) with the skills that they have. Self-confidence, therefore, is an athlete's belief about how he or she is going to do in competition, rather than just a perception of how good he or she is as an athlete.

Optimism has become a critical personality construct in relation to health and well-being but differs from self-confidence in that optimism is a tendency to expect the best possible outcome or dwell on the most hopeful aspects of a situation. Although different, optimism and self-confidence are related because confidence results from productive thinking habits that have become automatic and natural. Clearly, dispositional optimism is related to self-confidence due to the productive perceptions inherent in each. Other related terms represent similar "self" constructs, such as self-esteem and self-concept. Self-esteem pertains to an individual's personal perception of worthiness. Thus, athletes may have self-confidence but not self-esteem (and vice versa). Self-concept is more of an overall comprehensive view of oneself and incorporates both self-confidence and self-esteem.

In summary, although many constructs are similar and related to the notion of self-confidence, self-confidence is uniquely defined as a belief or self-perception about one's ability to perform successfully. For athletes, this involves believing one has the critical components for sport success, including physical skills and training, cognitive efficiency to focus and make effective decisions under competitive stress, and the ability to be resilient in responding to obstacles and poor performances.

## **2. IMPORTANCE OF SELF-CONFIDENCE TO ATHLETES**

The influence of self-confidence on athletes is one of the most intriguing topics in sport psychology. Athletes often attribute "choking" and performance slumps to "losing" their confidence. Research has strongly supported that higher levels of self-confidence are related to success in sport. The specific ways in which self-confidence helps athletes to achieve success are presented in the following sections.

### **2.1. Influence of Self-Confidence on Thoughts and Emotions**

Confident athletes are more cognitively efficient and emotionally adaptive than less confident athletes. Confident athletes have been shown to engage in more productive (personally controllable) attributions for their successes and failures, which serves to enhance their motivation. That is, when confident athletes fail,

they attribute their failure to unstable and personally controllable factors, which means their motivation remains stable and strong because they believe they can succeed in the future. Conversely, when athletes low in confidence fail, they attribute their failure to stable factors that are largely unchangeable (e.g., "I'm not any good"). Self-confidence also facilitates concentration or a task focus; research on basketball players showed that more confident players were more likely to narrow attention effectively and less likely to experience attentional overload and reduced attention. Athletes low in confidence tend to focus more on themselves and their perceived inadequacies and thus are unable to narrow attention to focus efficiently on performance demands. Higher self-confidence in athletes is also associated with more positive perceptions of arousal and anxiety because confidence provides a productive belief system in which negative emotions are reframed and viewed as necessary and facilitative. Athletes who lack confidence view their anxiety as negative and debilitating to their performance, and they lack the personal coping resources to deal productively with anxiety.

Confident athletes use more problem-focused coping strategies, as opposed to emotion-focused coping strategies. This again indicates an efficient and skillful use of cognitive resources in that coping should be problem focused to address important task issues as opposed to heightened emotionality, which does not directly address the problem and a solution. Confident athletes also engage in more mastery imagery and have better imagery ability than less confident athletes. Imagery has been shown to facilitate elite status in sport and is an important mental skill for athletes. Finally, self-confidence elicits adaptive emotional states for athletes, whereas a lack of confidence relates to anxiety, depression, and dissatisfaction.

## **2.2. Influence of Self-Confidence on Behavior and Performance**

Research has supported that self-confidence is positively related to performance in athletes. This has been shown in tennis, marathon running, baseball, basketball, field hockey, soccer, wrestling, track and field, gymnastics, and swimming. Research has also identified the behavioral mechanisms by which confidence enhances performance. Self-confidence increases athletes' effort and persistence in the face of obstacles, so these critical achievement behaviors along with enhanced cognitive and emotional efficiency serve to enhance the

performance of athletes. Interestingly, laboratory studies have shown that manipulating levels of confidence in individuals on motor performance tasks (arm wrestling, weight lifting, and muscular endurance tasks) actually affects the amount of effort and persistence engaged in by the participants. If participants in these studies were convinced that they were stronger or more fit than their opponents, they would increase their effort and persistence. Conversely, if they were convinced that they were weaker or less fit than their opponents, they demonstrated less effort and persistence. Thus, athletes who lack confidence tend to exert less effort and persistence in the face of obstacles and failures, which is clearly detrimental to their performance and development.

As one might expect, the effects of performance on self-confidence are even stronger than the effects of confidence on performance. That is, self-confidence helps athletes perform better, but successful performances are even stronger predictors of athletes' subsequent levels of self-confidence. This is discussed further in the next section.

## **3. SOURCES OF SELF-CONFIDENCE IN ATHLETES**

Most research in sport psychology on the sources of self-confidence has been conducted within the parameters of Bandura's self-efficacy theory. Bandura asserts that there are four sources of self-efficacy: (i) enactive mastery experiences, such as success; (ii) vicarious experiences, such as social comparison with others or watching successful models; (iii) verbal persuasion or the social influence of evaluative feedback, expectations of others, self-talk, and positive imagery; and (iv) physiological and affective states, such as being psyched up for competition or fatigued and listless, both of which influence subsequent levels of confidence in athletes. All four sources have been shown to be predictive of confidence in sport and motor activities, with enactive mastery experiences (or success) emerging as the strongest source of confidence, as predicted by theory.

A research line in sport-confidence has identified nine sources of confidence important to athletes in competitive sport (Table 1). The nine sources of self-confidence for athletes are divided into three broad domains: achievement, self-regulation, and social climate. Achievement sources of confidence include mastery and demonstration of ability. These sources are similar to Bandura's source of enactive mastery

**TABLE 1**  
**Sources of Sport Confidence in Athletes**

<i>Source</i>	<i>Confidence derived from . . .</i>
<b>Achievement</b>	
Mastery	Mastering or improving your personal skills
Demonstration of ability	Showing off your skills to others or demonstrating more ability than your opponent
<b>Self-regulation</b>	
Physical/mental preparation	Feeling physically and mentally prepared with an optimal focus for competition and performance
Physical self-presentation	Perceiving that you physically look good
<b>Social climate</b>	
Social support	Perceiving support and encouragement from significant others in sport, such as coaches, teammates, and family
Vicarious experience	Watching others, such as teammates, perform successfully
Coach's leadership	Believing your coach is skilled in decision making and leadership
Environmental comfort	Feeling comfortable in a competitive environment
Situational favorableness	Feeling that the breaks or the nature of a situation are in your favor

experiences. However, for athletes in sport, important enactive mastery experiences that enhance their self-confidence involve mastering or improving their skills as well as demonstrating greater ability than their opponent (which often translates into winning). Thus, although winning is an important achievement source of confidence for athletes, they may also develop confidence from other types of success (e.g., mastery or skill improvement).

The second domain of sources of self-confidence for athletes is self-regulation, which includes physical/mental preparation and physical self-presentation. Feeling physically and mentally prepared with an optimal focus for competitive performance is a critical source of confidence for athletes. Research and a great deal of experiential evidence from successful athletes indicate that precompetitive routines, rituals, and focus plans are critical predictors of how well athletes perform, especially in championship or pressure situations. Physical

conditioning has been rated by coaches as the most important way to develop self-confidence in athletes.

The third domain of sources of self-confidence for athletes is social climate, with five sources representing aspects of the physical and social environment that serve to enhance the confidence of athletes. Clearly, perceiving support and encouragement from others (social support) is an important source, as is vicarious experience, in which athletes (particularly young athletes) gain confidence to try skills after watching a teammate or friend perform them successfully. Very unique to the competitive sport environment is the source of coach's leadership, whereby athletes derive feelings of confidence from perceiving that their coaches are skilled leaders and effective decision makers. The final two sources in the social climate domain are also very unique to the competitive sport environment. Environmental comfort is often identified by athletes as a source of confidence gained from feeling comfortable in a particular competitive venue. The "home advantage," or research finding that home teams win more than 50% of their competitions played under a balanced home-and-away schedule, is often cited as a source of confidence for athletes. Situational favorableness involves gaining confidence by feeling that factors within the situation are in one's favor. For example, the popular notion of psychological "momentum" refers to athletes' perceptions that something has occurred that increases their probability of success, which typically creates a surge of confidence.

For individual sport collegiate athletes, the top five sources of self-confidence were physical/mental preparation, social support, mastery, demonstration of ability, and physical self-presentation. Physical self-presentation and social support have been shown to be more important sources of confidence for female athletes, indicating that body image along with social approval from others are learned to be important for females when competing in sport. Female athletes learn that maintaining socially defined feminine qualities affords them social acceptance and approval due to the emphasis placed on how women look in society. For team sport athletes, the top five sources of self-confidence were mastery, social support, physical/mental preparation, coach's leadership, and demonstration of ability. Demonstration of ability was a more important source of confidence for male athletes, indicating that male athletes typically experience extreme normative social comparison expectations with regard to sport prowess and participation.

Highly confident athletes tend to use personally controllable sources to nurture their confidence, including physical/mental preparation and mastery. Athletes with less confidence tend to base their confidence on less controllable sources, including environmental comfort and social support. Research on children and young athletes shows that frequency and intensity of negative feedback from adults and an emphasis on unrealistic goals and expectations can lower confidence levels. This is because young children base their confidence more on feedback from adults than do adolescents and adults. Overall, the sources of confidence that athletes develop and use are important determinants of their ultimate levels of self-confidence, and particularly the stability of their confidence over time. Research supports that athletes gain self-confidence when they achieve mastery goals; engage in effective self-regulation of cognitions, emotions, and behavior; and train and compete in a competitive environment that is supportive, challenging, comfortable, and motivating.

#### **4. SELF-CONFIDENCE WITHIN SPORT TEAMS**

People involved in sport understand that although self-confidence is an individual perception, there is also an important sense of team, or collective, confidence that is defined as the shared beliefs of team members in the ability of the team to perform successfully. Clearly, individual team members may be very confident about their individual abilities to perform successfully, but at the same time they may have little or no confidence in the team's ability to succeed. Team confidence is particularly important for highly interactive teams (e.g., volleyball, hockey, and basketball teams) as opposed to coaching teams (e.g., track and field or golf teams).

Research found that team confidence was a better predictor of team outcomes for ice hockey than the aggregate of individual levels of confidence of team members. Also, team confidence increased after winning and decreased after losing, whereas individual self-confidence was not influenced by team outcomes. Teams with high levels of confidence have also been shown to be more cohesive in terms of team members' feelings of involvement with the team's productivity. Finally, athletes' perceptions that there is strong leadership within the team was related to overall team confidence, particularly for teams that were unsuccessful the previous season. This finding indicates that the

confidence of teams coming off losing seasons may be positively enhanced if leadership abilities within the team are apparent to athletes on the team.

#### **5. MEASURING SELF-CONFIDENCE IN ATHLETES**

Self-confidence in athletes has been measured in many ways, depending on the conceptual perspective of the researcher. Several inventories have been developed to measure various aspects of confidence in athletes. The Trait Sport-Confidence Inventory and State Sport-Confidence Inventory have been used to measure dispositional sport-confidence and "right now" (often precompetitive) sport-confidence in athletes, respectively. The Sport-Confidence Inventory has been developed to assess not only total sport-confidence in athletes but also the various types of sport-confidence shown to be relevant to athletes in competitive situations (SC-physical skills and training, SC-cognitive efficiency, and SC-resilience). In addition, the Sources of Sport-Confidence Inventory is used to measure the nine sources of sport-confidence discussed previously.

Using a self-efficacy conceptual approach, researchers measure self-confidence in athletes specific to a particular task (e.g., volleyball serving as opposed to confidence in overall volleyball performance abilities). Self-efficacy measures can be hierarchical or nonhierarchical. Hierarchical self-efficacy measures are typically constructed by listing a series of tasks that vary in difficulty, complexity, or stressfulness. Athletes are asked to designate (yes or no) the tasks they believe they can perform (efficacy level). For each task designated as yes, athletes rate their degree of certainty (efficacy strength) that they can execute it, from total uncertainty to total certainty (scales range from 0 to 10 in 1-unit increments or 0 to 100 in 10-unit increments). Nonhierarchical self-efficacy measures provide a list of subskills needed to perform successfully in a specific sport (e.g., serving, passing, hitting, blocking, back row defense, and setting in volleyball). Athletes are asked how sure they are that they can successfully perform each specific task at a certain level (e.g., 1, 2, 3, . . . , 10 out of 10 times). This assesses the level of self-efficacy for each task. Athletes are then asked how certain they are that they can perform each task at each level (usually 0–100%). This assesses strength of self-efficacy, or how sure athletes are that they can attain each level of each task.

## **6. STRATEGIES TO ENHANCE SELF-CONFIDENCE IN ATHLETES**

Self-confidence may be enhanced in athletes by strengthening the sources of confidence that are important to athletes and success in sport.

### **6.1. Quality of Physical Training and Perceived Achievement**

The most obvious point about developing confidence in athletes is that self-confidence must be earned through the development of physical skill and fitness to serve as a basic prerequisite for performance success. Quality training to hone skills and fitness development are foundational achievements that serve as the basis for confidence building. Athletes have rated physical preparation as one of their top sources of confidence, and successful world-class athletes have particularly emphasized the importance of quality training for their confidence and performance success. Coaches and sport psychology consultants should identify training methods that enhance confidence, such as simulating pressure situations and creating unexpected obstacles to train adaptability in athletes.

Another basic foundation for athletes' confidence is their perception of their past achievement. The key here is the emphasis on perceptions, with the point being that athletes must engage in productive attributional assessments of their achievement attempts and perceptions of success based on controllable factors (e.g., appropriate goal setting and evaluation). Higher levels of confidence in athletes have been shown to be related to focusing on sources or strategies that athletes personally control, such as mastery of physical skills, whereas lower levels of confidence were related to uncontrollable socially comparative or environmental sources. Coaches and sport psychology consultants should help athletes define success in personally controllable ways, such as through the use of performance and process goals.

### **6.2. Self-Regulation**

Self-regulation has been identified as a significant source of confidence for athletes, and it is the central focus of interventions in sport psychology. To self-regulate in sport means that athletes must effectively manage their behaviors, thoughts, and feelings in relation to competitive challenges and demands. Goal mapping, imagery, and self-talk are the three basic

self-regulatory tools advocated by sport psychologists to enhance confidence.

The focused and persistent pursuit of goals serves as a basic regulator of human behavior; thus, it seems obvious that goal mapping is an important strategy to enhance the confidence and performance of athletes. A goal map is a personalized plan for an athlete that contains various types of goals and goal strategies as well as a systematic evaluation procedure to assess progress toward goals. Although athletes typically talk about the motivating qualities of outcome goals (winning or making the Olympic team), research indicates that performance and process goals are better sources of confidence. These types of goals seem to enhance confidence by allowing athletes to feel a sense of control over their performance and also by providing a specific task focus for productive concentration.

Imagery, or visualization, is assumed to enhance confidence in two ways. First, visualizing one's self successfully performing in sport may, in an individualized vicarious sense, provide a perfect mental model for performance. Imagery in this sense serves a cognitive function to improve or refine the physical reproduction of a motor skill. Second, imagery may enhance confidence through its motivational effect on behavior, creating a productive task focus of "what to do," and/or by creating positive emotion to energize confidence and performance. Researchers have shown that highly confident athletes tend to image the mastery and emotion associated with sport competition more so than less confident athletes. Thus, imagery interventions to enhance confidence should focus not only on the cognitive reproduction of skill execution but also on the emotional, attentional, and motivational aspects of performing in sport. Research has demonstrated that imagery training enhances sport-confidence in athletes.

Managing thoughts, or self-talk, is advocated by applied sport psychology sources as an important key for self-confidence. Athletes who effectively manage their self-talk are, in essence, controlling how they think as opposed to letting external events dictate how they think. This is essential due to the social cognitive nature of sport-confidence. That is, confidence is a belief system—a set of beliefs about one's ability to be successful in sport. Self-talk is representative of this belief system because what athletes say to themselves is what they believe. The difficulty in managing self-talk lies in the automaticity and invisibility of athletes' thoughts that make up their belief systems.

These belief systems are a product of past social interactions that have reinforced certain ideas about their abilities. These belief systems directly influence, and in fact define, one's self-confidence.

There are several variations of self-talk interventions that may serve to enhance confidence, including negative thought-stopping, countering, reframing, and the use of affirmations. Negative thought-stopping teaches athletes to recognize negative thoughts, self-monitor across time to learn individualized thought patterns in response to social stimuli, and habituate personal cues that allow them to stop negative thoughts from becoming internalized into their belief systems. Countering involves self-debate to understand how irrational and inappropriate beliefs create negative thinking, and this intervention has been shown to enhance confidence in athletes. Reframing involves choosing to view events in alternative ways, typically in a more productive way, and affirmations are a form of productive self-statements repeated over and over to internalize a belief or attitude. All these interventions focus on the development and internalization of a positive belief structure about one's self; thus, all serve to enhance confidence.

Overall, self-regulation is based on the idea of choice. Athletes have the volition, or will, to choose how they act, think, and feel. Social conditioning clouds this perception and seduces individuals into believing that their thoughts and feelings are controlled by their social and physical environments. The key to consistent, healthy confidence in sport is the exercise of personal volition through self-regulation. Although this is difficult, particularly in the organizational culture of sport, it is possible to learn and enhance self-regulatory skills to maintain optimal self-confidence in sport.

### 6.3. Social Climate

Self-regulation has been hailed as an important and successful strategy to enhance confidence, but coaches and sport psychology consultants must keep in mind that human behavior always takes place within a social climate. Intervention plans designed to enhance the motivational climate of educational settings have increased the motivation and confidence of learners. Social climate factors that seem to influence confidence are leadership style, types of goal recognition and evaluation, social feedback sources and types, and availability and characteristics of models.

Confidence is enhanced when athletes are part of a productive and congruent social support network that can be facilitated through team-building activities. The social influence of coaches' leadership styles and feedback patterns have been shown to be important influences on athletes' perceptions of competence and confidence. Based on the importance of vicarious experience as a source of confidence, athletes should be provided effective models for not only skill execution but also achievement behaviors such as effort and persistence. Although environmental comfort emerged as a significant source of confidence for athletes, teaching athletes to rely on such external and uncontrollable factors to enhance their confidence seems problematic. Athletes should learn to base their confidence more on self-regulatory control of perceptions, emotions, and behavior so that they are not subject to the inconsistent type of self-confidence that is dependent on social and environmental confidence builders.

### See Also the Following Articles

Arousal in Sport ■ Attention and Concentration Training in Sport ■ Intrinsic and Extrinsic Motivation in Sport ■ Psychological Skills Training in Sport ■ Successful Athletic Careers

### Further Reading

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Feltz, D. L., & Lirgg, C. D. (2001). Self-efficacy beliefs of athletes, teams, and coaches. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 550–565). New York: Wiley.
- George, T. R. (1994). Self-confidence and baseball performance: A causal examination of self-efficacy theory. *Journal of Sport & Exercise Psychology*, *16*, 381–399.
- Grove, J. R., & Heard, N. P. (1997). Optimism and sport confidence as correlates of slump-related coping among athletes. *The Sport Psychologist*, *11*, 400–410.
- Jones, G., & Swain, A. B. J. (1995). Predisposition to experience debilitating and facilitative anxiety in elite and nonelite performers. *The Sport Psychologist*, *9*, 201–211.
- Moritz, S. E., Hall, C. R., Martin, K. A., & Vadocz, E. (1996). What are confident athletes imaging? An examination of image content. *The Sport Psychologist*, *10*, 171–179.
- Vealey, R. S. (2001). Understanding and enhancing self-confidence in athletes. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 550–565). New York: Wiley.
- Vealey, R. S., Hayashi, S. W., Garner-Holman, M., & Giacobbi, P. (1998). Sources of sport-confidence:

- Conceptualization and instrument development. *Journal of Sport & Exercise Psychology*, 20, 54–80.
- Vealey, R. S., & Knight, B. J. (In press). Conceptualization and measurement of multidimensional sport-confidence. *Journal of Sport & Exercise Psychology*.
- Zinsser, N., Bunker, L., & Williams, J. M. (2001). Cognitive techniques for building confidence and enhancing performance. In J. M. Williams (Ed.), *Applied sport psychology: Personal growth to peak performance* (pp. 284–311). Mountain View, CA: Mayfield.



# Self-Control

Kathleen D. Vohs

University of British Columbia, Vancouver, British Columbia,  
Canada

Roy F. Baumeister

Florida State University, Tallahassee, Florida, USA

1. Definitions and Conceptual Considerations
  2. Research Paradigms and Findings
  3. Self-Control Process Models
  4. Domains of Self-Control
  5. Behavioral Change
  6. Individual Differences
  7. Summary
- Further Reading

## GLOSSARY

**delay of gratification** The idea that self-control is important in terms of being able to delay indulging in impulses or succumbing to temptation. Longer delays have greater benefits.

**ego depletion** A state of depletion of regulatory resources, typically associated with subsequent self-control failure.

**emotional control** Attempts to change the valence, intensity, or duration of one's emotional reactions. Most emotion control attempts are aimed at ameliorating negative affect.

**Low Self-Control Scale** A trait scale created to measure the effects of low self-control abilities on criminality.

**mental control** Attempts to change the content or frequency of one's thoughts.

**regulatory resource model** The idea that self-control is governed by a finite supply of energy that is shared among all varieties of self-control.

**self-control** The intentional, conscious process of changing one's behaviors or responses.

**Self-Control Scale by Tangney and Baumeister** A trait scale that measures generalized self-control abilities in the

domains of self-discipline, reliability, work ethic, healthy habits, and deliberate action.

**self-control subscale of the California Psychological Inventory** A trait scale created to measure ebullient and joyful relinquishments of self-control.

**self-discipline** People's actions to achieve deliberate goals by way of self-improvement.

**self-regulation** Modification, alteration, or guidance of one's responses (be they automatic or conscious) to achieve a desired goal or end state.

**TOTE model** shorthand term for the test–operate–test–exit cybernetic loop. Self-control is conceptualized as a cyclical set of processes in which one moves increasingly closer to the goal (“tests” and “operates”) and then finally “exits” the loop when the goal is reached.

The ability to regulate the self has been deemed “the hallmark of human functioning” because it constitutes a vital human skill. Mental and physical health rely heavily on the capacity to effectively and consistently exert control over incipient responses, often by inhibiting an undesirable response and then substituting a desirable response. Failures to self-regulate underlie health problems ranging from risky sexual practices to poor nutrition and improper diet, child abuse, overspending, and all manners of addictive problems. Accordingly, understanding how, why, and under which conditions people fail at self-control is a goal that has taken on paramount importance in recent decades.

## 1. DEFINITIONS AND CONCEPTUAL CONSIDERATIONS

The term self-control is used to indicate the intentional, conscious process of suppressing or inhibiting one's behaviors or responses. Self-discipline is a term that is subsumed under the concept of self-control. It is used to indicate people's actions to achieve deliberate goals that are often in the service of self-improvement. It is useful to distinguish the term self-control from the term self-regulation. Self-regulation is a broad term used to indicate the modification, alteration, or guidance of one's responses (be they automatic or conscious) to achieve a goal. The focus of this article is self-control, which means that all the findings discussed here could also be considered self-regulation research but only some of which are self-discipline research.

## 2. RESEARCH PARADIGMS AND FINDINGS

This section outlines the major processes and domains of self-control research. The aim of this section is to demonstrate the basic processes that are thought to underlie self-control attempts as well as to illustrate the breadth of responses that are governed by self-control.

## 3. SELF-CONTROL PROCESS MODELS

### 3.1. Cybernetic Model

Charles Carver and Michael Scheier developed the cybernetic model for self-control when they introduced the idea of self-control as occurring within a test-operate-test-exit (TOTE) loop. People enter the TOTE loop when they establish a goal. The first action, test, refers to the comparison between the current and goal state. Assuming there is a mismatch between current and desired state, the person engages the operate function of the model in which responses are enacted to close the gap. The person tests again and, depending on whether he or she deems the goal to be met or whether more work is needed, the person either returns to the operate phase or exits the loop.

One important idea from Carver and Scheier's TOTE model is that emotions reflect progress toward a goal: Positive affect is often a cue that a person is moving

toward the goal and negative affect signals movement away from the goal.

### 3.2. Regulatory Resource Model

The regulatory resource model (or ego depletion model) posits that the ability to control oneself is governed by a finite resource that is shared among all types of self-control faculties. (Readers may notice that the resource model specifies the operate function in a TOTE model.) Any one act of self-control leaves the person less successful for a subsequent act (within a circumscribed time period) because fewer resources are available to put toward the second self-control action. This is known as a state of ego depletion, which reflects an inability to muster enough regulatory resources to meet a specific self-control demand. Roy Baumeister, Kathleen Vohs, and colleagues posited that the responses involved in emotion regulation, mental control, impulse intervention, and behavioral guidance all use this resource and, consequently, levels of the resource diminish with each act of self-control.

Two examples of studies from the resource model are relevant to applied psychology because they involve dieting and impulsive spending. In these studies, people were instructed to control some aspect of their behavior (e.g., attentional or emotion responses). This was intended to deplete regulatory resources for these participants, relative to participants who were not instructed to engage in self-control. Subsequent behaviors in the realms of eating ice cream and unplanned purchasing patterns were assessed. People who had earlier controlled themselves in one domain were less able to control themselves with respect to eating or impulsive spending.

### 3.3. Delay of Gratification Paradigm

The third model of self-control centers on the ability to delay gratification. For more than 40 years, research by Walter Mischel and colleagues has illuminated the importance of curbing frustration, stifling unwanted responses, and overcoming temptation for goal achievement, psychological growth, and well-being. Typical studies testing the delay of gratification perspective focus on children approximately 3 or 4 years old, who are seated in front of a table on which sits an inviting treat. The child is told that if he or she does not eat the treat (e.g., one marshmallow) in front of him or her now then he or she will be given a larger

treat (e.g., two marshmallows) later. Delay of gratification is measured in terms of whether and how long the child can wait and not eat the available treat.

The adaptiveness of being able to delay gratification has been well demonstrated. Children who at 4 and 5 years of age had better delay of gratification skills had higher SAT scores, got along better with others, and dealt with life stressors more effectively. These data speak to the long-lasting benefits of early self-control and the importance of self-restraint skills.

## 4. DOMAINS OF SELF-CONTROL

### 4.1. Mental Control

It is no small feat to rid oneself of an unwanted thought, and research on thought suppression and mental control has illuminated the sheer difficulty of such intents. Psychologist Daniel Wegner built an important line of research around asking people to suppress thoughts of a white bear. People given such instructions find it incredibly difficult to accomplish this goal; moreover, when the suppression task ended, people experienced a surge in white bear thoughts. This effect has been replicated in studies on suppressing thoughts about sex, secret romantic relationships, and depressive thoughts.

### 4.2. Emotion Control

There are approximately six different categories of affect regulation: getting into, out of, or maintaining a positive or negative emotion. Most affect control efforts, however, are aimed at just one of these categories: ameliorating negative emotions. Researcher James Gross has conceptualized emotion regulation as occurring both pre- and post-emotional stimulation. Situation selection involves altering emotional state by choosing specific types of people, places, and objects to change one's mood. Situation modification strategies are efforts to alter a situation to change its emotional effects. Attentional deployment isolates aspects of the situation to manage emotional state. For instance, people who are scared of flying may choose to watch the in-flight movie. Cognitive change involves reconstruing the situation in alternative ways to lessen the impact of the emotional stimulus. Last, response modulation refers to the direct control of emotional responses after they are initiated.

## 5. BEHAVIORAL CHANGE

Recent models of behavioral self-control have highlighted the importance of distinguishing the start of behavioral change from the maintenance of behavioral change. Psychologist Alexander Rothman found that what people want to get out of a behavioral change program determines whether they make a behavioral change, but satisfaction with progress determines whether the change will be maintained. A person considering dieting will therefore consider how he or she will look in a new bathing suit when the diet is complete. If this expectancy is positive, the person is likely to start a diet. Once started, however, whether the person is satisfied with the perceived progress predicts continuation of the program. This analysis of behavioral change has been a powerful tool to researchers and clinicians for determining how to motivate people to begin self-change programs and to stay on track.

### 5.1. Adherence to Moral Rules

Adhering to cultural rules and living up to a society's values is an important life ability that also requires self-control. Research suggests that children who are toddler aged have little self-control, and that self-control begins to emerge at approximately 5–10 years of age in response to punishment fears. However, Grazyna Kochanska and colleagues have shown that in children of toddler and early school age, inhibitory control predicts children's conscience, including moral behavior, moral cognitions, and moral identity and self-concept. Judith Rich Harris suggested that by preadolescence and adolescence, children demonstrate self-control primarily in response to interpersonal and peer isolation cues. By the time people are in their early 20s, they are behaving in a self-controlled manner that is primarily in response to internalized cultural values.

## 6. INDIVIDUAL DIFFERENCES

This section reviews trait scales of self-control. They have been derived from different histories and backgrounds and thus have provided the field with a variety of options for measuring different types and manifestations of self-control.

The oldest of the three scales reviewed is the self-control (Sc) subscale from the California Psychological Inventory (CPI). The 38-item scale was initially created

to measure “the kind of joyful, ebullient abandonment of restraint that one sees at certain times such as attendance at a carnival.” Accordingly, Harrison Gough and colleagues set about to measure “impetuosity, high spirits, caprice, and a taste for devilry” and hence the Sc scale of the CPI was born. This scale gained followers in the 1960s and 1970s, but it did not enjoy a very popular run due to some psychometric problems. Many of the items on the CPI Sc scale are too varied (overly heterogeneous) and lack validity. Moreover, the scope of the scale is too narrow.

The socialization subscale of the CPI is also widely used in applied settings. It has 46 items and is best known as a scale to measure impulse restraints, revealing whether a person is reliable, thorough, industrious, well socialized, and rules and directions are followed without problem.

The Five Factor Personality Model, popularized by Paul Costa and Robert McCrae, has also been used to predict self-controlled behaviors. The personality dimension of conscientiousness best approximates a self-control scale, although extraversion has also shown some predictive abilities. Conscientious individuals are better able to make and carry out long-term plans, which means that they have better self-control, especially regarding the practical aspects of self-control (e.g., planning, being reliable, and keeping organized).

From the study of criminology, Harold Grasmick and colleagues set out to operationalize ideas put forth in Michael Gottfredson and Travis Hirschi’s book, *A General Theory of Crime*, which conceptualized aggression and criminality as poor self-control. The resulting Low Self-Control Scale has been used by criminologists to study crime, aggression, and delinquency, and it has proven its mettle by showing that people who score low in self-control are more likely to have problems abiding the law and following codes of conduct.

The Low Self-Control Scale contains 24 items, including statements such as

- I seem to have more energy and a greater need for physical activity than most other people my age.
- I will try to get things I want even when I know it’s causing problems for other people.
- When things get complicated, I tend to quit or withdraw.
- I like to get out and do things more than I like to read or contemplate ideas.
- I’m not very sympathetic to other people when they are having problems.

- I often act on the spur of the moment without stopping to think.

- I frequently try to avoid projects that I know will be difficult.

- I’m more concerned with what happens to me in the short run than in the long run.

Participants are asked to indicate their agreement with these statements as descriptions of “the type of person you are.” The scale is a unidimensional construct.

The most recently developed scale is the Self-Control Scale by June Tangney and Roy Baumeister. Their scale was created to measure varieties of self-control abilities and failures that have psychological and practical implications, such as breaking habits, resisting temptation, and maintaining self-discipline. This scale has a 36-item long form and a 10-item short form. The long form has five subscales: self-discipline, deliberate/nonimpulsive action, healthy habits, work ethic, and reliability.

The Self-Control Scale contains 36 items, to which participants are asked to assign a numeric value (where 1 = “not at all like me” and 5 = “very much like me”) that “best represents what you believe to be true about yourself for each question.” Representative items include

- I have a hard time breaking bad habits.
- I am lazy.
- I say inappropriate things.
- I never allow myself to lose control.
- I do certain things that are bad for me, if they are fun.
- People can count on me to keep on schedule.

People who score high on this scale get better grades in college, report less psychopathology, have higher self-esteem, are less likely to engage in binge eating or alcohol abuse, have a more secure attachment style, report better relationships with family and friends, and show fewer negative emotional reactions.

## 7. SUMMARY

This article discussed the core elements of self-control, including theoretical, empirical, and practical treatments of it. Self-control is one of the greatest human abilities because it gives us the flexibility to (psychologically) move beyond the immediate physical setting and orient our responses toward abstract, higher order goals. It is likely that self-control abilities are one of the key dimensions that allowed humans to advance evolutionarily, given its advantages for securing safety and increasing reproductive possibilities.

The psychological literature is clear on the importance of self-control for overcoming immediate urges, controlling thoughts, modifying emotional states, and altering behavior. Having good self-control, at a personality level, is related to better outcomes in the realms of physical health, mental well-being, interpersonal relationships, and cultural rule-abiding (i.e., criminality). In short, there are very few realms in which having good self-control is not a benefit to the self.

### See Also the Following Articles

Emotion

### Further Reading

- Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation*. San Diego: Academic Press.
- Baumeister, R. F., & Vohs, K. D. (In press). *Handbook of self-regulation research*. New York: Guilford.
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer-Verlag.
- Faber, R., & Vohs, K. D. (In press). To buy or not to buy? Self-control and self-regulatory failure in purchase behavior. In R. F. Baumeister, & K. D. Vohs (Eds.), *Handbook of self-regulation research*. New York: Guilford.
- Gross, J. J. (1999). Emotion and emotion regulation. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 525–552). New York: Guilford.
- John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 102–138). New York: Guilford.
- Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology*, *54*, 687–696.
- Sayette, M. A. (In press). Self-regulatory failure and addiction. In R. F. Baumeister, & K. D. Vohs (Eds.), *Handbook of self-regulation research*. New York: Guilford.
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (In press). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*.
- Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological Science*, *11*, 249–254.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34–52.



# Sexual Behavior

John D. Baldwin and Janice I. Baldwin

University of California, Santa Barbara, California, USA

1. Before Puberty
  2. Puberty and Later
  3. Early Sexual Problems
- Further Reading

## GLOSSARY

**double standards** The notion that society often treats men and women differently, expecting and accepting unequal behavior from the two genders.

**observational learning** The ability of people to learn about behavior by observing other individuals who perform the behavior; men tend to imitate men more than they do women, and women tend to imitate women more than they do men.

**operant learning** The ability of individuals to learn from the consequences of their own actions; behavior that is followed by rewards tends to become more frequent in the future, and behavior that is followed by punishment tends to become less frequent in the future.

**Pavlovian conditioning** The notion that stimuli that precede and predict the onset of a stimulus–response reflex become conditioned stimuli that have the capacity to elicit similar reflexive responses.

**punishers** Behavioral consequences that are “bad outcomes”; they suppress the behavior they follow, producing inhibitions about repeating the behavior.

**reinforcers** Behavioral consequences that are “good outcomes”; they strengthen (or reinforce) the behavior they follow.

**sexual turn-ons** Conditioned stimuli that have become associated with the sex reflex through Pavlovian conditioning; they are also called erotic stimuli and are capable of eliciting sexual arousal.

One of the best ways in which to understand human sexuality is to take a developmental approach, tracing sexual behavior from infancy onward. Babies are capable of basic sexual responses from the early days of life, and it is informative to study how more complex sexual behavior develops from the early responses of infants. This knowledge is of great practical value for all students of sexuality, including sex therapists and any individuals who want to improve their sexual experiences in life. The developmental approach allows one to begin with the most basic elements of sexual behavior and then add layers of extra information as one traces the emergence of sexual behavior from infancy to adulthood. Developmental theories are also well suited for showing how nature and nurture interact. The development of human sexuality involves complex mixes of biological and social inputs—both nature and nurture—mediated by multiple neural systems that operate at both conscious and unconscious levels, intertwining biologically based reflexes, learned responses, and various degrees of self-control. Naturally, it takes many years of sexual learning for a person to acquire all of the layers of complexity that are typical of adult sexuality. Not surprisingly, some people acquire far more layers of intricate sexual knowledge and behavior than do others.

## 1. BEFORE PUBERTY

Long before the onset of puberty, many infants and children have sexual experiences, including spontaneous

sexual arousal, erections, vaginal lubrication, self-stimulation, and sometimes sex play. Even though childhood sexual activities are less complex than the sexual conduct seen after the onset of puberty, they show that the sexual organs are well developed and operative early in life. People are not born as sexual “blank slates,” as some early scholars once assumed. Before birth, various physiological mechanisms have developed to produce the sexual responses seen in early life.

The first sexual responses of babies are purely reflexive rather than consciously planned or intended. There are reflex mechanisms located in the lower half of the spinal cord that function from birth to produce vaginal lubrication, penile erection, and other basic sexual responses. Soon after being born, many boys have penile erections, and some girls have vaginal lubrication and clitoral tumescence. Ultrasound imaging of pregnant women reveals that male fetuses sometimes have erections in utero, several months before birth. Given current technology, it is difficult to measure the sexual responses of female fetuses, but there are indications of genital swelling and vaginal lubrication.

An infant's sex reflexes can be elicited by mild stimulation of the genitals. Penile erection in baby boys and vaginal lubrication in baby girls may occur when bed sheets or clothing rub the infants' genitals, when babies explore touching their genitals with their hands, or when other people fondle the infants' genitals. Prolonged touching can lead to sexual excitement, and a small percentage of children learn to masturbate to orgasm during the first year of life.

In some cultures, adults gently touch their babies' genitals as a way in which to pacify the infants, indicating that genital stimulation is pleasant for infants. In the U.S. culture, adults who stimulate children's genitals might be charged with child sexual abuse. However, infants often learn to touch their own genitals, even if they have never been touched by other people.

### 1.1. Early Operant Learning

It is very natural for children to use their hands to explore nearly anything that is within their reach, including parts of their own bodies. In this process of exploration, many infants discover that touching their own genitals brings pleasure. Technically speaking, genital stimulation is a “primary reinforcer” because nerves from the genitals are biologically connected to the “pleasure centers” in the brain and genital stimulation produces the rewarding feelings associated with sex. When infants first touch their genitals, they are merely

exploring the world they can reach with their hands, with no intentions of discovering sexual pleasure. They do not innately know how to masturbate or produce sexual pleasure. However, their random explorations can produce positive reinforcers if they happen to touch their genitals, and these reinforcers increase the likelihood that, on later occasions, exploring infants will repeat those types of self-stimulation that were pleasurable in the past.

As children have repeated experiences with self-exploration and operant learning, some master increasingly effective ways of producing sexually pleasurable sensations. The types of genital self-touching that bring the most pleasure are reinforced and become stronger habits, whereas the types of self-stimulation that do not produce much pleasure become less frequent due to lack of positive reinforcement. As a result, some infants learn relatively effective forms of self-stimulation that adults may label with the term “masturbation” even though babies cannot verbalize or subjectively understand such words. Nevertheless, some infants learn how to stimulate themselves to orgasm before their first birthday.

Human sexual behavior is based on a combination of nature and nurture. Biological processes (i.e., nature) provide the infant with sex reflexes as well as the neural connections needed to experience sexual arousal, pleasure, and orgasm—and to learn from those experiences. Both personal experiences and social influences (i.e., nurture) are important as well. The power of nurture can be readily seen in the ways in which parents influence their children's operant learning by their responses to the children's early sexual exploration. Some parents recognize how natural it is for children to touch themselves, and they allow their children to explore many forms of self-touching. At the other end of the continuum are parents who punish their children for touching their genitals, perhaps by using a spanking, a firm grip on the arm, or a stern tone of voice. Parents who use strong punishment can suppress their children's masturbation and produce long-lasting inhibitions. Of course, many parents fall somewhere between the two ends of this continuum, allowing some pleasurable sexual experiences while giving various mixtures of punishment and inhibitions.

Parents who permit their children to explore and discover genital stimulation without punishment allow the children to learn masturbatory skills through the natural reinforcers of sexual pleasure, although they may also teach their children to restrict masturbatory activities to private locations such as bedrooms. On the other hand, many parents use considerable punishment

when they see signs of early sexuality; hence, many children stop masturbating during early childhood. Strong punishment may cause them to feel inhibited about masturbating, not only during childhood but also for many years later. The more punishment a child receives for masturbation, the stronger the inhibitions are likely to be, especially if the child is told that the masturbation is dirty, evil, or sinful.

Common sexual behaviors reported by mothers of 2- to 5-year-old children include masturbation, touching the mother's breasts (or other women's breasts), and trying to peer at adults or children when they are undressing or nude. As the children become older, their sexual behaviors gradually diminish. Mothers of 6- to 9-year-olds reported that their children primarily masturbated at home and attempted to look at people who are undressing or nude. By 10 to 12 years of age, the only common sexual behavior that mothers reported was an interest in the opposite sex.

### **1.2. Early Pavlovian Conditioning**

Children's emotional responses to all aspects of sexuality result from the pleasurable conditioning that occurs as a side effect of reinforcement and punishment. Children who are allowed to explore their sexuality and discover sexual pleasure without punishment experience positive Pavlovian conditioning that minimizes inhibitions about sexuality. If parents avoid associating guilt and punishment with sex and explain that sex is a natural healthy part of life (especially when done in privacy), children can grow up feeling good about their sexuality. In contrast, children who have been punished for engaging in sexual activities may learn to associate sex with fear and anxiety through Pavlovian conditioning. In the future, doing sexual activities—or even thinking about them—can elicit negative emotions and inhibitions.

Childhood masturbation and Pavlovian conditioning can lead to the early development of sexual fetishes, although this is not common. A classic example in the sex literature is that of a man who developed a sexual fetish for rubber that could be traced to early childhood, when his mother had him sleep on a rubber sheet. He discovered pleasurable experiences when he accidentally rubbed his penis against the rubber bed sheet. Because he repeated this activity on many subsequent occasions, he developed a strong and long-lasting attraction to rubber objects (due to Pavlovian conditioning), and he began to collect all sorts of rubber items. Even though his mother often took his rubber toys away from him, he would collect more, often using them during

masturbation. By adulthood, his positive conditioning for rubber had become strong enough to be called a "rubber fetish," and he enjoyed and defended this fetish as a special part of his sex life.

### **1.3. So What Is "Normal" or "Typical" During Childhood?**

In 2001, Kaeser and colleagues studied kindergarten, first-grade, and second-grade children to determine the common or "normal" sexual behaviors at that age. Most common were talking among peers and close-age siblings about genitals and reproduction, showing one's genitals (if the peer or close-age sibling would reciprocate), playing "doctor," masturbating, imitating seductive behaviors such as kissing and flirting, and occasionally using "dirty" words or telling dirty jokes. A variety of behaviors were judged as possibly problematic, including a preoccupation with either masturbation or sexually explicit conversations with peers. Activities that were judged to be definitely problematic included having sexually explicit conversations with people who were notably different in age and repeatedly touching the genitals of others. Behaviors judged to need therapeutic intervention were acts such as oral, vaginal, or anal penetration of dolls, animals, or children and simulated intercourse with peers after removing clothing.

### **1.4. Double Standards**

Many children are exposed to double standards about sexual activities. First, parents are twice as likely to punish their daughters for masturbation than to punish their sons, leading boys to believe that their parents are more accepting of sexual self-stimulation than girls believe their parents to be. Also, parents are three times more likely to tell their daughters that masturbation is sinful or morally wrong than to tell their sons. Thus, it is not surprising that girls and women often feel more uncomfortable than do boys and men about exploring sexual forms of self-pleasuring. This does not mean that women cannot overcome negative feelings about sexual self-stimulation, even those that were learned during early childhood. Sexual inhibitions are not "fixated" or "imprinted" during childhood experiences, as Freud and some early psychologists assumed. Sexual inhibitions can be reduced, through Pavlovian counterconditioning, by nearly any kind of pleasurable experiences that a person has associated with sex. A person with many sexual inhibitions can overcome

them, especially if he or she has a sexual partner who is very positive about sex as a natural and wonderful aspect of human experience and makes sure that both partners have very pleasurable experiences with sex.

Another double standard about childhood sexuality can be seen in the ways in which parents give names to their children's genitals. In our society, many parents teach their sons to name their genitals as a "penis" or some euphemism such as a "pee-pee." In contrast, parents often give their daughters less useful names for the vagina such as a "pocketbook that you'll open after you get married." Few girls are told about their clitoris and its name. Parents might fear that if they tell their daughters that they have a clitoris, the girls might ask, "So what's it for?" Many parents would be reluctant to answer that its only function is to produce sexual pleasure.

### 1.5. Early Observational Learning

During the early years of life, children often learn bits and pieces of information about sexuality by observing adults, television, movies, magazines, books, and early sex education programs. Children are often more attentive to sex information and learn more from observation than adults realize. The belief that observant children are not learning much about sex results from the fact that most children rarely talk about or act out the sexual behaviors they have observed. Why? Many children are punished if they act on their sexual observations, and this inhibits such behavior. In addition, children do not derive as much pleasure (or sexual reinforcement) as do teens and adults when they do act out sexual behaviors that they learned by observation because the pleasures of sexual stimulation are less intense before puberty than after puberty. This, in turn, helps to explain why prepubertal children cannot understand why teens and adults find sex so fascinating. Of course, some prepubertal children learn very little about sex because their parents prevent them from seeing sexual materials.

Parents are often quiet when their children are observing sexual information, as if they are hoping that their children are oblivious to the sexual implications of media images and innuendos. Nevertheless, there are so many modern television shows and Web sites that show blatantly sexual images that some modern parents are becoming aware of the importance of talking with their children about the information presented on television and through other sources. Concerned parents often explain how they perceive the media messages so that their children will learn how to evaluate media

images and dialogues. Such intervention helps children to learn how their parents process information from mass media while also passing on the parents' values to the children.

Unfortunately, some children receive no parental advice about the sexual information that is easy to observe in both the electronic and print media. Images and stories of rape and random violence can leave these children adrift with few tools for evaluating media messages about sex. These children are most likely to perceive sexual assault as natural and normal parts of social life, although not all of them will act on that perception.

Girls tend to receive considerably more guidance for evaluating media sex messages than do boys. Many parents and teachers know that naive girls can be seduced or forced into sexual activities that might lead to sexually transmitted infections (STIs), unintended pregnancies, and/or injuries. Thus, many parents, teachers, and media programs tend to target girls more than boys with messages about "responsible sex." Perhaps society should end this double standard and teach more boys about the merits of responsible sex. This can be done in a very positive way by telling boys that they also benefit from reducing the risks of STIs, unintended pregnancies, and violence. In addition, boys could come to realize, "She will love me more if I am responsible, concerned, and loving." Currently, adults may be doing boys a disservice by giving boys fewer guidelines than they give girls for considerate and mature approaches to sex.

## 2. PUBERTY AND LATER

The interactions of nature and nurture are especially conspicuous during puberty. Because of pubertal hormonal changes, both the sexual organs and secondary sex traits undergo the adolescent changes that make both boys and girls start to look more sexually mature, leading society to increasingly treat them like adults. The hormonal changes also cause the sex organs to produce increasingly pleasurable feelings when touched, speeding the pace of operant learning from sexual experiences.

### 2.1. Six Basic Sex Differences

During puberty, both boys and girls begin to have increasing numbers of experiences with spontaneous sexual responses and nocturnal orgasms (triggered by unconscious mechanisms deep in the brain), but there are several major sex differences in these hormonally

induced changes. First, pubertal boys begin to have spontaneous sexual arousal 2 to 3 years before girls do, giving boys a significant head start over girls in learning about sex. Second, pubertal sexual arousal is more noticeable and distracting for boys than for girls. A boy who has a spontaneous penile erection is far more likely to notice this conspicuous change than a girl is to notice when she has spontaneous clitoral enlargement and lubrication in the vagina. The girl's response is subtler and more difficult to notice, whereas the boy's response is bigger in that as his penis elongates from 3 or 4 inches (the average length before stimulation) to approximately 6 inches (the average length once erect). Because the boy's penis is well endowed with nerves that connect to the reward centers of his brain, the boy is likely to associate sexual arousal with pleasure. Even though the girl's genitals also connect to the pleasure centers in her brain, the subtlety of female sexual responses makes them less noticeable and less likely to create the powerful positive sexual conditioning that boys experience. Third, boys are significantly more likely to experience spontaneous nocturnal orgasms than are girls. Among young people who experience nocturnal orgasms, boys report them approximately four times more frequently than do girls. These orgasms are often associated with sexual dreams, and this strengthens the Pavlovian conditioning that links erotic images with sexual feelings, especially for boys. Fourth, boys experience nocturnal orgasms at a younger age than do girls, giving boys another major head start in sexual learning.

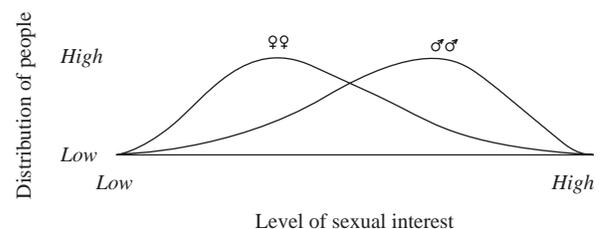
All four of these biological differences lead pubertal boys to become more curious about and interested in sex sooner than do same-age girls, although there is much variability among both boys and girls in the frequency and time of onset of all four of these experiences. Superimposed on these pubertal changes are all of the different levels of inhibitions or positive associations that each individual has learned about sex from his or her parents, family, and culture, with the double standard tending to create more sexual inhibitions for girls than for boys. This interaction of nature and nurture tends to produce the male–female differences in sexual interest shown by the bell-shaped curves illustrated in Fig. 1. The different shapes of the two skewed bell curves show that teen boys tend to be more interested in sex than are same-aged teen girls while also revealing considerable variability in the sexual interest of both boys and girls. For example, some girls fall near the right end of the continuum, indicating that they have more sexual interest than do all of the boys who are to their left on the continuum.

Adolescence brings about two additional concerns about girls' sexuality that can affect their sexual behavior. Sexually active girls tend to worry more about unintended pregnancies—and, in some cases, the issues related to abortion—than do most boys. Also, STIs can cause more serious damage to the reproductive organs of girls than to those of boys; hence, sexually active girls tend to worry more about the risk of contracting STIs than do sexually active boys.

## 2.2. Double Standards

Societal double standards about sexuality can augment the sex differences seen in Fig. 1. Adolescent girls and adult women are often given the pejorative labels of being “loose” or “sluts” if they have several sexual partners, and this can inhibit some from having too many partners, although there is little consensus on the definition of “too many.” In contrast, adolescent boys and adult men often do not feel bad if they have numerous sexual partners and might even be labeled as “studs.” Thus, these social labels tend to give girls more inhibitions than boys experience when thinking of having multiple partners. There is another double standard that relates to active and passive social roles. “Nice girls” are not supposed to take active or assertive roles in initiating sexual activities, and many girls feel uncomfortable if they do act too sexually knowledgeable or eager. In contrast, boys are often told that their role is to be active, gain sexual experience, and become sexual experts—and most boys feel happy when they succeed at these. Not surprisingly, in many societies, boys tend to think more about sex and initiate more sexual interactions than do girls.

Studies on the goals for the “first date” among high school and college students reveal that 19- and 20-year-old men typically wanted to touch their partners' genitals (under their clothing) on the first date, whereas same-age women wanted only hugging, physical



**FIGURE 1** Distribution of people along the continuum of sexual interest.

contact, and sensuous kissing. After these students started going out regularly but not exclusively, the men wanted oral sex, whereas the women were just beginning to want their genitals to be touched. When college freshmen were asked whether it was appropriate to have sex if two people had only recently met, 38% of the women and 66% of the men said “yes.”

After the “sexual revolution” of the 1960s, many women began to criticize the societal double standards that placed unnecessary inhibitions on female sexuality. As sex education became more widespread, women became increasingly empowered to want to know more about sexuality and have better sex lives. Today, there are increasing numbers of books and magazine articles geared to helping women explore their sexuality, and many more adolescent girls are curious about their sexuality than was true during the 1950s. The development of effective methods of contraception has also allowed both women and men to engage in sex without as much fear of having unintended pregnancies as would have been the case during the 1950s. These social developments have led to more rapid changes in sexual behavior than would be the case if sexual activities were determined by biology alone.

### **2.3. Recent Trends**

Between the onset of the sexual revolution and the early 1990s, there was an increase in adolescent sexual behavior, teen pregnancies, and teen abortions in the United States. However, adolescents have tended to become more sexually conservative since the early 1990s. There has been a decline in the number of teens who have engaged in sexual intercourse, had pregnancies, and obtained abortions. One reason for the reduction in risky teen sexual activities is the societal response to the fear of HIV/AIDS, including increased education about this disease, which has instilled a sense of caution in the age cohorts that reached puberty after 1990.

In the United States, many young adolescents experiment with masturbation, and by 16 years of age approximately 90% of boys and 60% of girls have masturbated. Boys typically masturbate approximately five times per week, and girls do so much less frequently. Typical adolescents begin engaging in sexual intercourse by 16 or 17 years of age, but approximately 20% of students have had sexual intercourse by 15 years of age. During the senior year of high school, 65% of students report having engaged in sexual intercourse. By the end of high school, approximately 20% of adolescents have had four or more sexual partners.

Not all adolescents are certain about their sexual orientation. In a study of junior and senior high school students, 10.7% reported that they were unsure of their sexual orientation and 1.1% reported a bisexual or mainly homosexual orientation. The remaining 88.2% described themselves as predominantly heterosexual. Some of the young people who are uncertain about their sexual orientation experiment with “both sides”; that is, they try sex with both boys and girls so as to understand what their sexual orientation is. Having social support from caring parents, teachers, friends, and counselors can help an adolescent to manage the complex psychological issues that can arise during these years of questioning one’s sexual orientation. Before 2003, homosexual behavior was illegal in some states, and this did not help adolescents and young adults to adjust easily to having gay, lesbian, and bisexual identities. In 2003, the U.S. Supreme Court struck down state laws against homosexual behavior, asserting that the constitutional rights to privacy protect people’s choices of intimate sexual activities.

Modern society has fewer taboos about sexual behavior than existed a century ago, and more young people are exploring varied sexual activities than in the past. The 100-year trend in sexual interactions shows an increase in petting, oral sex, penile–vaginal intercourse, and anal sex among teens, although since the 1990s young adults have become more cautious about activities they perceive might put them at risk for contracting HIV/AIDS.

### **3. EARLY SEXUAL PROBLEMS**

Some people have an ageist belief that only adults have sexual problems, whereas young people have naturally blissful sex. This is far from true. Although young people can be swept away by the passion and novelty of sex, they often learn about sexual activities through trial and error. Not unexpectedly, many experience numerous errors and frustrations if they have little sex education or prudent guidance in their sexual learning.

The sexual problems that are most common for young adults are understandable in light of the information presented in the prior two sections. Young men tend to become very excited by sexual activities and reach orgasm quickly, which can create the problem labeled as “premature ejaculation” or “early ejaculation.” In contrast, young women tend to have limited knowledge about the stimulation they need to reach orgasm, and they have more inhibitions about sexual exploration and playfulness than do young men. As a

result, many young women's biggest sexual problems center on orgasm. For example, they find it difficult to become sexually excited, stay excited, and reach orgasm. Men often reach orgasm too quickly, whereas women tend to have the opposite problem.

The following subsections describe some of the common problems that young people have with sex along with some possible solutions.

### 3.1. Common Male Problems

Because young men tend to have more experience with spontaneous sexual feelings and masturbation than do young women, young men often know what kinds of stimulation they like best. Many young men discover that using strong and fast sexual movements nearly always leads to orgasm. As a result of this operant learning, young men often develop the habit of masturbating quickly, and this can sometimes bring them to orgasm in a couple of minutes. In fact, rapid masturbation might seem to be a necessity if a young man fears that someone, such as a parent or roommate, may walk in and discover what he is doing if he engages in sexual self-stimulation for a prolonged period of time. Later, when having sex with a partner, the young man might feel the need for rapid sex if he has only the limited privacy offered by a family room or dormitory. All of these early sexual experiences can reinforce and strengthen a young man's habit of using strong and rapid movements to keep sexual activities brief and secret. The fact that his sexual activities at high muscle tension can lead to a very rapid sexual response might not be a problem during masturbation and in some relationships, especially if his partner also fears being caught in the act if the partners take too long.

Although rapid ejaculation might make sense to teens, it can create problems as people mature, have more access to privacy, and begin to explore forms of sexual activity that can extend and heighten the pleasures for both partners. At that time, a man's early habits of rapid movements and early ejaculation can become a liability. This is especially true if the man loses interest in sex after his orgasm but his partner would love to have more sexual interaction. When early ejaculation creates problems in a relationship, it can earn the label of "premature ejaculation." During the 1940s, it was not uncommon for sexual relations to last only 2 minutes before the man ejaculated and sex came to an end. But after the sexual revolution, people have come to realize that it can be pleasurable to extend sexual relations for

longer times, especially because some women need 5 or 10 minutes—or longer—to reach orgasm.

There are various methods that men can use to learn how to slow their sexual response and avoid ejaculating too early. Some men can learn the following technique by reading one or two paragraphs on it; however, the stronger a man's habits are for fast and intense self-stimulation, the longer it usually takes to learn to slow down and master new habits for lingering longer over the pleasures of sexual stimulation.

One of the best techniques a man can learn for avoiding early ejaculation is called the "stop-start method." Rather than starting sexual actions and continuing them all the way to orgasm, a man learns to alternate between starting and stopping his physical movements—and his sexual fantasies. Whenever he feels that his sexual activities are causing him to become too excited too quickly, he stops his sexual self-pleasuring and concentrates on bringing pleasure to his partner. Then he resumes cycling back and forth between starting and stopping, learning that the stop periods slow him down and help to ensure that his partner is also experiencing considerable pleasure. Most couples find that the stop-start method brings both of them more gratification than when the man merely starts his most exciting activities and progresses quickly to an early ejaculation.

Men who have learned powerful habits of rapid sex, due to years of positive reinforcement for coming fast, might need to practice the stop-start method systematically for months to unlearn their old habits of speed and discover the multiple rewards of the new habits of self-control. They might need to read Zilbergeld's 1992 book *The New Male Sexuality*, with longer and more detailed descriptions of the technique and several related subtle nuances, to master the new habits.

### 3.2. Common Female Problems

Women rarely complain about reaching orgasm too fast, although some do. It is much more common for women to report that it is hard for them to become sexually excited, stay excited, and reach orgasm. Among the causes for this are inhibitions about masturbating, talking openly about sex, and seeking sexual information, all of which often result from the double standards of male and female sexual socialization. As a result, many women have a limited understanding of their own bodies, which parts yield the most intense sexual feelings when touched, and how best to stimulate those parts. A lack of knowledge and sexual skill on the part of women—or their partners—can deprive them of the

best types of stimulation for becoming sexually aroused and reaching orgasm.

Most sex therapists and professional self-help books encourage women to experiment with several pleasurable sexual learning experiences. First, women should explore touching all parts of their bodies in the privacy of their bedrooms, perhaps after a soothing bath and with some favorite music in the background. The goal is to discover the parts of their bodies that feel good to touch, noting especially those areas on and around the clitoris and breasts that trigger erotic and sexual feelings. When stimulating those erotic areas, women also need to learn what types of touch feel best, for example, gentle circular motions or back-and-forth stimulation. Some women purchase vibrators to learn how the gentle mechanical oscillations can bring sexual pleasures. Most women learn that gentle stimulation on and around the glans and shaft of the clitoris produce the most sexual excitement, which usually mounts over time, eventually reaching orgasm. Some women discover the effectiveness of water stimulation, either by lying on their backs in a bathtub and allowing water from the spigot to flow over the clitoris or by using a handheld shower to spray water on their external genitals.

Once a woman discovers which parts of her body is erotically sensitive and what types of stimulation she likes, she can progress to either or both of the following techniques. The woman can experiment by herself with self-touching and/or vibrator stimulation to see whether she can reach higher levels of sexual arousal and eventually have orgasms. She can also ask her partner to join her in experimenting with various types of stimulation of her breasts and clitoral area. She and her partner can explore countless different ways of stimulating her clitoral glans and shaft such as gently rubbing up and down, back and forth, or in circular motions. Many women like oral stimulation of the clitoral area, and in fact many more women reach orgasm through oral stimulation than by penile–vaginal intercourse. Both women and men need to realize that stimulation of the inside of the vagina with a penis, dildo, or penis-shaped vibrator is far less likely to lead to female orgasm than is stimulation around the clitoral area. Because men usually think of the penis as their key sex tool and source of pleasure, they often orient their sexual activities to penile–vaginal intercourse, but this can easily deprive women of stimulation to the clitoral glans and shaft.

To create additional sources of erotic stimulation that can augment the physical arousal derived from stimulation of the breasts and clitoral area, sex

therapists often encourage women to develop sexual fantasies. For centuries, men have paired erotic images and fantasies with the sexual excitement of masturbation and coitus, thereby turning their fantasies into sexual turn-ons through Pavlovian conditioning. Increasing numbers of women are learning to eroticize mental images and fantasies by pairing them with the sexual excitement of masturbation and partner stimulation.

Another problem that females report more often than do males is painful intercourse, which is technically called “dyspareunia.” Anyone who experiences painful intercourse should see a physician for a personalized diagnosis and medical advice because there are many possible causes of sexual pains. In serious cases, it is prudent to avoid penile–vaginal sex until after consulting with a physician. People who cannot abstain from penile–vaginal intercourse might try using a water-based lubricant such as KY Jelly or Astroglide to reduce the friction of coitus. For women with dyspareunia, manual or oral stimulation to the clitoris and surrounding areas can lead to lubrication and orgasm before penile intromission—and even without it. In cases of female pain, another solution involves a much slower and more gradual introduction to erotic play and intromission than most males tend to do without a special request. If these practices are not helpful, a physician should be able to identify the exact cause of the pain and suggest at least one solution. For example, the woman may have a vaginal infection or a urinary tract infection. Either the woman or her partner may be allergic to the latex in condoms. A number of useful Web sites provide additional information about both male and female sexual problems.

During the past half-century or so, there have been so many advances in the science of sexuality that increasing numbers of people are enjoying the pleasures of sex without the problems of pain, unintended pregnancies, or STIs. Unfortunately, society still has a long way to go in reducing the persistent problems of sexual assault, rape, domestic violence, HIV/AIDS, and homophobia. The fact that sexuality can be discussed, written about, and legislated more freely today than was the case a mere 50 years ago bodes well for continued progress in resolving these remaining problems.

### See Also the Following Articles

Gender and Culture ■ Homophobia ■ Homosexuality ■ Risk-Taking and Sexual Behavior ■ Sexual Behavior and Culture

**Further Reading**

- Baldwin, J. D., & Baldwin, J. I. (1981). *Beyond sociobiology*. New York: Elsevier.
- Baldwin, J. D., & Baldwin, J. I. (1997). Gender differences in sexual interest. *Archives of Sexual Behavior*, 26, 181–210.
- Friedrich, W. N., Fisher, J., Broughton, D., Houston, M., & Shafran, C. R. (1998). Normative sexual behavior in children: A contemporary sample. *Pediatrics*, 101, E9.
- Kaesler, F., DiSalvo, C., & Moglia, R. (2001). Sexual behaviors of young children that occur in schools. *Journal of Sex Education and Therapy*, 25, 277–285.
- Pfaus, J. G., Kippin, T. E., & Centeno, S. (2001). Conditioning and sexual behavior: A review. *Hormones & Behavior*, 40, 291–321.
- Remafedi, G., Resnick, M., Blum, R., & Harris, L. (1992). Demography of sexual orientation in adolescents. *Pediatrics*, 89, 714–721.
- Sargent, T. O. (1988). Fetishism. In D. M. Dailey (Ed.), *The sexually unusual: Guide to understanding and helping* (pp. 27–42). New York: Harrington Park Press.
- Zilbergeld, B. (1992). *The new male sexuality*. New York: Bantam Books.



# Sexual Behavior and Culture

Lise D. Martel, Skyler Hawk, and Elaine Hatfield

University of Hawaii at Manoa, Honolulu, Hawaii, USA

1. Sexual Permissiveness
  2. Sexual Expression
  3. Contraception and Birth Control
  4. Sexuality and HIV/AIDS
  5. Conclusions
- Further Reading

## GLOSSARY

- agrarian** Living from the management or distribution of land.
- AIDS** Abbreviation for acquired immune deficiency syndrome.
- double standard** Arbitrary use of two different sets of rules for two distinct groups.
- globalization** The sharing and adoption of new techniques resulting from the increasing interconnectedness of the nations of the world.
- HIV** Abbreviation for human immunodeficiency virus, which is responsible for the development of AIDS.
- hunter-gatherers** Nomadic people who live by hunting wild game, fishing, and gathering nuts, fruits, berries, and other plant products growing in the wild.
- industrial society** Community where industry is the principal activity of the society.
- pandemic** Disease affecting a whole country or the whole world.
- sexual permissiveness** Toleration of a wide variety of sexual behaviors and attitudes.

Sexual behaviors and values are important components of all cultures. Societies may differ greatly in the meanings they attribute to sex, their sexual

permissiveness, the expressions of their sexual values, their views on contraception and pregnancy control, and the health issues they face. Yet despite their infinite variety and the individual differences within each culture, some broad generalities can be identified. Societies' commonalities and differences in behaviors and beliefs are determined by cultural, socioeconomic, and historical contexts, and during recent years they have been determined by globalization and technological advancement as well. By engaging in social comparison, people garner their understandings of naturalness and normality, thereby forming their attitudes toward sexual issues and determining what is sexual and sexually acceptable.

## 1. SEXUAL PERMISSIVENESS

A society's general level of sexual permissiveness most likely originates from complex interactions among several cultural factors. For example, a culture's technological sophistication, the influence of particular faiths, and the importance of economic exchange at marriage all may have impacts on the place of sex in a society. Acceptable forms of sexual expression are also determined by the importance that is placed on reproduction. For example, engaging in homosexual activity may be viewed more negatively in those cultures that hold procreation, but not pleasure, as an acceptable motivation for sex.

### 1.1. Permissive Societies

Many cultures see sexual expression as natural and beneficial. Cultures that can be labeled as sexually permissive hold generally positive attitudes toward sexual development, which is seen as essential to healthy social adjustment. There are usually fewer social restrictions placed on sexuality. Children in these societies are not sheltered from sexual knowledge. As with the Muria of central India, they might even be highly exposed to nudity and sexual activity through ritual and institutional practices. This is not seen as problematic because healthy sexuality is considered to be necessary for marital happiness and mature functioning. Both men and women in such cultures tend to equally value pleasure associated with sexual expression. Premarital sex is commonly accepted as a reality for both men and women.

### 1.2. Restrictive Societies

In contrast, sexually restrictive cultures attempt to limit both the occurrence of sexual expression and the ways in which it can be expressed. Sex itself is either seen as holding a unique and dangerous power or considered to be harmful because it is closely connected with negative outcomes such as early pregnancy and venereal disease transmission. Adults in restrictive societies generally try to limit children's understanding of sexuality for longer periods of time, but for different reasons. For example, in the United States, parents' long-term goals for children's development include later marriage and prolonged education. Because the possible negative consequences of sexual activity are seen as antithetical to these goals, abstinence may be emphasized and may become the cultural norm for moral and acceptable behavior. Thus, in restrictive cultures, virtue is often associated with premarital chastity. Sexual abstinence before marriage is stressed for one sex (usually women) or both sexes. In cases where women's chastity is emphasized, men are typically permitted at least some greater amount of sexual freedom.

### 1.3. Sexual Double Standard

When inequalities in sexual freedom are found, sexual activity and associated pleasure tend to be more valued by the gender whose behavior is less restricted. This describes what is referred to in Western literature as the sexual "double standard." In many cultures, men are often encouraged to seek out and acquire sexual

experiences with many partners, whereas women are expected to delay sexual activity until after marriage. This double standard can produce different sexual motivations for men and women. Men are more likely to endorse reasons of enjoyment and pleasure, whereas women are more likely to endorse reasons of love and emotional commitment. Although this double standard has eroded considerably in industrial countries, it can still be observed under some circumstances in that people are generally more tolerant of men who engage in casual sex or unusual sex practices than they are of women who do so. In most African and Asian nations, a double standard is still frequently reported.

## 2. SEXUAL EXPRESSION

Sexual expression varies widely from one culture to another. These differences include, but are not limited to, cultural views and behaviors related to premarital and extramarital sexuality, homosexuality, transgenderism, genital mutilation rituals, contraception and birth control, and issues related to the HIV/AIDS pandemic.

### 2.1. Premarital Sexuality

#### 2.1.1. Agrarian and Hunter–Gatherer Societies

In agrarian and hunter–gatherer societies, premarital intercourse is often perceived negatively. Although the absence of reliable birth control and the disastrous social consequences of unwanted pregnancies are good incentives for compliance, early age of marriage and late sexual maturation also make this social expectation bearable. In such cultures, a promise of marriage is often the only acceptable situation where premarital intercourse is considered acceptable. For some cultures, it is sexual maturity rather than marriage that counts. For the !Kung of southern Africa, married men may take prepubescent brides but must wait for their menarches to have sex with them.

#### 2.1.2. Industrial Societies

In many industrial societies, the increasingly permissive attitudes toward premarital sexual activity can be partly explained by women's empowerment through oral contraception and advances toward economic equality between men and women. Also, delaying marriage until a later age increases the gap between sexual maturity and marriage and makes premarital abstinence an unrealistic

goal for many. Yet despite the prevalence of more permissive attitudes, for most young people in industrial societies, sex is still expected to involve some level of affection and commitment between partners.

## 2.2. Extramarital Sexuality

### 2.2.1. Agrarian and Hunter–Gatherer Societies

In many agrarian and hunter–gatherer societies, extramarital relations are considered to be “appropriate” for men under certain socially established conditions. For example, in Thailand, young men are viewed as having insatiable sexual drives that they have a right to satisfy, even after marriage, through visits to sex workers. In some African societies, the importance of a large family to a man’s status gives him the right to copulate with another woman if his own wife is infertile.

### 2.2.2. Industrial Societies

Researchers agree that two changes are evident during recent years in industrial societies. First, the sexual double standard is eroding. Men and women are becoming similar in their willingness to experiment with extramarital sex. Second, men and women are experimenting with extramarital relations earlier, more often, and with more partners than was the case in the past.

## 2.3. Homosexual Orientation and Behaviors

Homosexuality generally refers to sexual attraction or behaviors toward people of one’s own biological sex. Differences in how “sex” is defined in a particular culture, and the meanings that are imparted on different sexual behaviors between people of the same sex, can often lead to contrasting ideas about what “homosexuality” actually entails and what behaviors are associated with it. In cultures where sex is defined in very narrow terms, such as consisting solely of penile–vaginal intercourse, a great deal of intimate behavior may be acceptable between people of the same sex. Sometimes, behaviors that would be considered highly sexual in one society are given a different nonsexual meaning in other cultures. For example, young adolescent boys in the Sambia culture of Papua, New Guinea, ritually fellate older men as a way of absorbing their strength and masculinity. However, they are later expected to continue on to exclusive heterosexual marriages and relationships, and adult same-sex relationships are not well

accepted. In Ancient Greece, important older men often took younger men as sexual partners concurrently with the older men’s heterosexual marriages.

Homosexuality is highly connected to perceptions of gender roles in a number of cultures. For many peoples in Latin, African, and Middle Eastern countries, as well as among many ethnic minorities in America and Europe, men may engage in same-sex behaviors yet still identify as “heterosexual.” They are able to engage in these behaviors, without incurring the stigma that accompanies a label of homosexuality, due to the roles that they play in these sexual activities. Men in these cultures who take the “inserting” role in same-sex encounters are still seen as masculine and heterosexual, presumably because they are adhering to sexual–cultural norms for their gender. On the other hand, men who take the “insertee” role, which is commonly viewed as the woman’s role in sex, are typically viewed as feminine and are labeled “homosexual.”

In contrast, in the United States and much of Western Europe, homosexuality is perceived to be an intrinsic aspect of the individual rather than being defined by gender roles. Therefore, men who identify as homosexual experience a much greater interchangeability in their sexual role taking. The Western conceptualization of homosexuality has been bolstered during recent years by a plethora of scientific information and political assertions that homosexual orientation has a genetic etiology.

As globalization continues to proliferate, many cultures are shifting their notions of homosexual behavior to be more consistent with Western (particularly American) ideologies. Still, cultural differences are numerous. In multicultural societies such as the United States, cultural differences in the conceptualization of homosexuality should be considered in social marketing campaigns, as well as in other interventions, that are geared toward homosexuals of specific minority groups.

## 2.4. Transgenderism

The term “transgender” applies to people who experience a deep lifelong preference for the identities or gender roles usually associated with being a member of the opposite sex. Some people who consider themselves to be transgendered choose to live out the conventional roles assigned to them by society. Many others take a variety of approaches to socially identify with what they feel is their true gender. Some choose to wear clothes and ornaments of the opposite sex; adopt the postures, gestures, behaviors, and roles of the

opposite sex; and/or undergo sex reassignment surgery. They might become transvestites or cross-dressers, drag queen or kings, or transsexuals. Yet this does not imply that becoming a cross-dresser or a drag queen or king is always motivated by transgenderism.

In most Western societies, the challenges that transgenderism poses to traditional Western gender roles still result in widespread stigmatization. In fact, in Western culture, transgenderism, transvestism, and transsexuality are often labeled as forms of a mental disorder called gender dysphoria. In contrast, in many non-Western societies, transgendered individuals, transvestites, and transsexuals are valued for their uniqueness. For example, historically in the Zuni culture of New Mexico as well as in many island cultures of the Pacific, male children who displayed feminine characteristics were considered to be endowed with great power. They dressed as a woman and performed functions of healer and arbiter. In today's Samoa, humorous entertainment provided by cross-dressers is used to diffuse violence in public gatherings. In India, the Hijras, a group of men who have had their genitalia removed and have sex only with other men, make a living by dancing, singing, and offering blessings at weddings. During recent years, many Hijras have been elected to public office due to a perception that they are less likely to be corrupted than are other public officials.

### **2.5. Genital Modification Rituals**

Male circumcision is probably the most common form of genital modification. From Ancient Egypt to modern-day Africa and America, circumcision has been practiced under all kinds of justifications. In America, circumcision was popularized into the 1900s as a cure for masturbation and as a way in which to ensure good hygiene and prevent diseases. Today, undergoing circumcision is still a common practice in many cultures, especially Jewish and Moslem cultures.

In many Sub-Saharan societies and Moslem Arab groups, women undergo genital alterations and mutilations. The modifications range from removing the clitoral prepuce (sunna circumcision) to removing the entire clitoris and most of the labia and stitching together the two sides of the vulva, allowing only a small opening for the passage of urine and menstrual blood (infibulation). Many of these gruesome procedures result in health problems or even death (in extreme cases). In the United States, they have been outlawed since 1996.

### **2.6. Future Directions**

The meanings that a culture gives to sex, the sexual expressions it recognizes, and its level of permissiveness are in constant flux. Globalization and rapid technological development are increasing the rate of such changes and continually force all cultures to reexamine their beliefs about issues related to sexual permissiveness and sexual expression. Many Western practices, such as prolonged adolescence and delayed marriage, serve to remove sexuality from family control and place it in the hands of the individual. Agrarian and hunter-gatherer societies increasingly are adopting the ways of the West, forcing some rapid changes on their cultures. Some changes, such as relaxing premarital sexual standards and promoting gender equality, permit more individual freedom. Other changes, such as laws regulating genital modification rituals, impose new restrictions on cultures.

## **3. CONTRACEPTION AND BIRTH CONTROL**

In both developing and industrial societies around the world, women are perceived to be responsible for preventing unwanted pregnancies. In many developing nations, women are under the economic control of men, have little access to modern contraceptives, and are expected to bear the responsibility of preventing unwanted pregnancies even if they have low control. Not surprisingly, most of the world's births occur in developing nations. In industrial societies, the situation of women is quite different. Success of pregnancy prevention is increased by gains in equality of rights, economic independence, and access to modern contraceptives that do not require the cooperation of their male partners.

### **3.1. Developing Societies**

In rural cultures of the developing world, children are economic assets. From an early age, they become active contributors to the survival of their families through a variety of labor tasks such as those related to food cultivation and sibling caretaking. In an effort to maximize efficiency of the family unit and reduce health risks, women in these societies generally try to maintain a birth spacing of 2 to 3 years. To control this spacing, they usually rely on abstinence and on the reduced fertility offered by lactation. In many African and Southeast Asian cultures, there is an expectation that

husbands will abstain from having sex with their wives for a period of time after their wives give birth, but husbands are not necessarily expected to cease sexual activity completely. Extramarital sex, visits to sex workers, and (to a lesser extent) polygamy are common culturally acceptable alternatives for men. Women are usually expected to remain abstinent postpartum.

Despite the fact that a large number of people in developing countries live in rural areas, as a result of globalization and industrialization, an increasingly larger number live in overcrowded cities. For these people, the economic assets provided by children are often replaced by the children becoming an additional strain on already compromised financial means.

### 3.2. Industrial Societies

In most industrial societies, where childhood is usually spent in school, children represent an economic cost rather than an asset. Pregnancy is often planned and heavily controlled by the individual through the use of contraceptives. In some countries, such as China, pregnancy is also restricted by the state in an effort to manage overpopulation.

In the United States, at the time of this writing, the current political administration promotes abstinence as the only acceptable alternative to prevent pregnancy and sexually transmitted diseases among teenagers. Yet many organizations have successfully promoted the use of various modern forms of contraception to prevent pregnancy in teenagers and adults alike. Recent trends show a decline in unwanted pregnancies among U.S. adolescents. Still, the United States maintains a disproportionate number of abortions in comparison with other Western industrial countries. In countries such as Holland, Sweden, France, and Great Britain, it is believed that the general endorsement by individuals, society, government officials, and the media of the message that protected sex (rather than abstinence) is expected from teenagers has facilitated the decline in abortions and the delay in the onset of sexual activity.

### 3.3. Future Directions

Under the influence of globalization, beliefs about individualism, human rights, and gender equality are having a growing impact on the sexual behaviors and attitudes of people from developing nations. Unfortunately, these nations usually lack the medical and financial resources to adopt modern contraceptive methods that have proven to be successful in the

wealthier industrialized nations. In the increasingly interconnected world, overpopulation in developing countries is a universal concern. All societies must collaborate to address the old and new challenges that arise from living in an interconnected world.

## 4. SEXUALITY AND HIV/AIDS

The awareness of the link between sexuality and health is not new. An estimated 333 million new cases of curable sexually transmitted diseases occur worldwide every year. Yet it was only after the first diagnosed case of AIDS that sexually transmitted diseases came to the forefront of global health concerns. Since the early 1980s, more than 42 million people worldwide have been infected with HIV, making HIV/AIDS the most devastating sexually transmitted disease in the world's history.

### 4.1. Condom Use

#### 4.1.1. Developing Societies

HIV/AIDS has been especially devastating in Sub-Saharan Africa, Southeast Asia, and Latin America. In an effort to control the spread of HIV/AIDS, the most common approach used worldwide has been to promote the use of male condoms and provide education on the virus and its transmission. This approach is especially adequate in developing countries where most of the transmission results from heterosexual intercourse. Yet the success of this intervention varies widely across cultures and even between genders. Many societies do not have the literacy, education, and access to condoms that are necessary for successful prevention of HIV/AIDS. The use of condoms is often heavily stigmatized. Beliefs that condoms are appropriate only with prostitutes and that only promiscuous women carry condoms are common. Health beliefs can also interfere with the use of condoms. For example, in Cambodia, condoms are believed to disrupt the balance of the body and burn the womb.

Nonetheless, condoms have had a tremendous impact on the AIDS epidemic in many developing countries. For example, in Southeast Asia, where the commercial sex industry had fueled the AIDS epidemic, several countries have succeeded in increasing condom use among sex workers and their clients. Thailand, with its 100% condom use policy in brothels, is a prime example of the successful implementation of a condom use program in a high-risk population.

#### **4.1.2. Industrial Societies**

In industrial countries, the male condom is still the only effective protector against HIV/AIDS. During recent years, many attempts have been made to develop competing alternatives. The male condom has been criticized for being a barrier to women's empowerment in protecting their health. It was hoped that the female condom, which consists of a malleable tube with a flexible ring at both ends (one that can be inserted into the vagina by the woman the other resting at the opening of the vagina), would give women the empowerment they had been seeking. Unfortunately, proper use of the female condom still requires active cooperation from their male partners during penile insertion. In addition, despite promising theoretical efficacy, its use efficacy proved to be much lower than that of the male condom.

### **4.2. Gender Differences**

#### **4.2.1. Developing Societies**

In developing societies, HIV/AIDS is equally prevalent among men and women. Yet women are at greater risk for acquiring HIV than are men because women usually are less educated and do not have the societal position or financial means to obtain condoms. Their disadvantage is further compounded by the fact that HIV is more easily transmitted from men to women. Furthermore, the only known effective barrier to HIV infection, condoms, requires male cooperation. For women, risk of infection is sometimes exacerbated by the practice of dangerous cultural habits. For example, in Sub-Saharan Africa, which has the highest prevalence of HIV/AIDS in the world, the custom of dry sex is common. In this practice, women prepare themselves to pleasure their husbands by drying their vaginas with a mixture of powdered stem and leaf and water, soil mixed with baboon urine, detergents, salt, cotton, or shredded newspaper. These increase dryness, friction, vaginal lacerations, and the suppression of the their vaginas' natural bacteria that, in turn, increase women's risks of acquiring sexually transmitted diseases, especially HIV/AIDS. The Muslim Hausa of Nigeria perform cuts on the walls on the vagina to cure women suffering from a variety of maladies. In addition to exposing the women to risk of hemorrhage and infection, the cuts provide the AIDS virus with direct access to the bloodstream during intercourse.

Targeting women with interventions in developing countries could have a large impact on the HIV/AIDS

pandemic because women can also transmit the HIV infection to their infants through delivery and breast-feeding. In developing countries, women rarely have access to highly active antiretroviral treatments to limit the transmission during delivery. Also, in countries with high prevalences of waterborne diseases, such as cholera, breast-feeding is often the only viable means of survival for infants. Therefore, prevention interventions with mothers would also have a significant impact on the health of their children.

#### **4.2.2. Industrial Societies**

In Western countries, since the 1980s, gay or bisexual men and intravenous drug users have been recognized as high-risk populations for HIV/AIDS. Still today, 82% of the estimated cumulative 877,275 cases of AIDS in the United States are among men. Of these, 55% are those who have sex with other men and 26% are injection drug users. Although women of childbearing age are the fastest growing population of AIDS cases in America, 70% of new HIV infections still occur among men. Yet because of prevention intervention efforts, the estimated number of new HIV infections in the United States has decreased from approximately 100,000 to 40,000 per year. Moreover, with the advances in HIV treatments, especially the advent of highly active antiretroviral treatments, great progress has been made in reducing the number of new AIDS cases as well.

### **4.3. Future Directions**

Efforts should continue to be made to improve women's control over their own protection against HIV/AIDS. More likely, the solution will come in the form of a microbicide if issues of genital irritation can be resolved.

In both developing and industrial countries, one group that has not been adequately targeted in HIV/AIDS intervention is adolescents. In addition to being behaviorally and physiologically more vulnerable to acquiring HIV/AIDS and other sexually transmitted diseases, adolescents are often socially disadvantaged. They lack independent incomes and often are not offered the privacy of information that is necessary for open communication with health care providers.

Currently in the United States, funding for prevention efforts is shifting from targeting noninfected individuals to targeting infected individuals. The progress made in preventing new cases of HIV/AIDS seems to justify this new approach, yet efforts should be

maintained to continue prevention efforts with high-risk populations and fast-growing groups.

## 5. CONCLUSIONS

Globalization has propelled the societies of the world in an interaction that transforms cultures and redefines the sexuality of every nation. It has both positive and negative effects. On the one hand, it encourages the sharing of perspectives and medical advancements that can help societies to adopt sexual norms and beliefs that are compatible with healthy lives. On the other hand, it facilitates the spread of disease and reduces cultural diversity in the human sexual experience. It is hoped that in the future, through collaboration between industrial and developing nations, both healthy sexuality and gender equality will become a part of every distinct society and discrimination will become a rarity in the lives of individuals who belong to stigmatized groups such as homosexuals, transgendered individuals, and those infected with HIV/AIDS.

### See Also the Following Articles

Gender and Culture ■ Homosexuality ■ Marital Therapy  
 ■ Risk-Taking and Sexual Behavior ■ Sexual Behavior

## Further Reading

- Ford, C. S., & Beach, F. A. (1951). *Patterns of sexual behavior*. New York: Harper & Row.
- Francoeur, R. T. (Ed.). (1997). *The international encyclopedia of sexuality* (Vols. 1–3). New York: Continuum.
- Francoeur, R. T., & Noonan, R. J. (Eds.). (2002). *The international encyclopedia of sexuality* (Vol. 4). New York: Continuum.
- Galliano, G. (2003). *Gender: Crossing boundaries*. Belmont, CA: Thomson Wadsworth.
- Gibney, L., DiClemente, R. J., & Vermund, S. H. (Eds.). (1999). *Preventing HIV in developing countries: Biomedical and behavioral approaches*. New York: Kluwer Academic/Plenum.
- Hatfield, E., & Rapson, R. L. (1993). *Love, sex, and intimacy: Their psychology, biology, and history*. New York: HarperCollins.
- Hatfield, E., & Rapson, R. L. (1996). *Love and sex: Cross-cultural perspectives*. Needham Heights, MA: Allyn & Bacon.
- Hotvedt, M. E. (1990). Emerging and submerging adolescent sexuality: Culture and sexual orientation. In J. Bancroft, & J. M. Reinisch (Eds.), *Adolescence and puberty: The Kinsey Institute series* (Vol. 3, pp. 157–172). New York: Oxford University Press.
- LeVine, R. A., Dixon, S., LeVine, S., Richman, A., Leiderman, P. H., Keefer, C. H., & Brazelton, T. B. (1998). *Child care and culture: Lessons from Africa*. Cambridge, UK: Cambridge University Press.
- Shostak, M. (2001). *Nisa: The life and words of a !Kung woman*. Cambridge, MA: Harvard University Press.



# Sexual Harassment

Janet Sigal

Fairleigh Dickinson University, Teaneck, New Jersey, USA

1. Prominent Cases and Laws Against Sexual Harassment
  2. Theories and Models of Sexual Harassment
  3. Measures of Sexual Harassment
  4. Incidence of Sexual Harassment
  5. Outcomes and Responses
  6. Prevention and Intervention
  7. Conclusion
- Further Reading

## GLOSSARY

**gender harassment** Taunts, gender-based insults and gestures, and/or the presence of pornographic materials in the workplace.

**hostile environment sexual harassment** When working conditions serve as a source of distress for the employee.

**quid pro quo sexual harassment** Requesting sexual favors in exchange for promotion or other employment-related decisions.

**sexual coercion** Job-related threats and/or bribes to victims for giving sexual favors.

**sexual harassment** Unwelcome sexual advances, requests for sexual favors, and other verbal or physical contact of a sexual nature when (a) submission to such conduct is made, either explicitly or implicitly, a term or condition of an individual's employment; (b) submission to or rejection of such conduct by an individual is used as the basis for employment decisions affecting the individual; or (c) such conduct has the purpose or effect of substantially interfering with a person's work performance or creating an intimidating, hostile, or offensive working environment.

**unwanted sexual attention** Persistent requests for dates (without any sanctions or job-related threats associated with refusals), unwanted touching, and/or harassing victims through phone calls or letters.

Sexual harassment is a pervasive and persistent problem in the United States as well as around the world. According to estimates, approximately half of the women in the United States will experience sexual harassment at some time, either in the workplace or in educational settings. Men also are victims of sexual harassment, although at a lower rate. Sexual harassment occurs internationally at a rate similar to, or possibly even higher than, that in the United States. Because this problem affects many people worldwide, it is an issue that must be explored and investigated. Before prevention programs can be developed, laws must be understood, the incidence of sexual harassment worldwide must be examined, the consequences for victims must be established, and models must be explored to determine why sexual harassment occurs.

## 1. PROMINENT CASES AND LAWS AGAINST SEXUAL HARASSMENT

### 1.1. Prominent Cases

Sexual harassment became the focus of intensive research during the 1990s, due partly to the existence of

high-profile cases in the United States. In 1991, during a televised U.S. Senate hearing in which Clarence Thomas was being considered for appointment to the U.S. Supreme Court, professor Anita Hill accused Thomas of having sexually harassed her when she had worked for him at the Equal Employment Opportunity Commission (EEOC). Although Thomas was confirmed as a Supreme Court justice, the widely viewed hearings brought the issue of sexual harassment to the attention of the American public. Another highly publicized 1991 event, in which several naval officers were accused of sexually harassing women at the Tailhook convention for naval aviators, also was responsible for focusing public attention on this issue. Whether sexual harassment occurs in the workplace (e.g., when an employer demands that an employee exchange sexual favors for a promotion) or in educational settings (e.g., when a professor demands sexual compliance for a high course grade), the behavior is illegal and poisons the climate of the workplace or educational institution.

### 1.2. U.S. Sexual Harassment Laws

Table I lists some cases and the precedents set by these court decisions. Sexual harassment laws related to the workplace stem from Title VII of the Civil Rights Act of 1964 (and the extension of this act in 1991) that prohibits any form of discrimination. Laws against sexual harassment have been developed within the framework of sex discrimination in that harassment targets employees due to their gender and prevents these individuals from obtaining equal opportunities in the workplace. Laws prohibiting academic sexual harassment are related to Title IX of the Education Amendments.

In 1987, a court supported the establishment of the reasonable woman standard concept. This concept developed as a consequence of studies demonstrating that women were more likely to perceive inappropriate sexually related behaviors as sexual harassment and to expect organizations to rectify the situation than were men. Because women tend to be at higher risk for sexual harassment than are men, women might feel more threatened and more sensitive to the presence of harassment behaviors constituting a hostile environment. Therefore, it was reasoned that if jurors were encouraged to determine whether a reasonable woman would be upset and made to feel uncomfortable by these behaviors, they might be less biased against a woman who alleges that she has been sexually harassed than if the jurors adopted a reasonable person standard. It has not yet been clearly demonstrated that the

**TABLE I**  
**Significant Sexual Harassment Court Decisions**  
**in the United States**

<i>Court case</i>	<i>Decision</i>
<i>Davis v. County Board of Education</i> (1999)	Schools will be held responsible for peer harassment if it is severe and the schools were aware of the harassment.
<i>Oncala v. Sundowner Offshore Services</i> (1998)	Same-sex harassment and harassment of men constitute actionable sexual harassment.
<i>Lois E. Jensen v. Eveleth Taconite Co.</i> (1997)	First class-action sexual harassment hostile environment lawsuit was filed.
<i>Harris v. Forklift Systems</i> (1993)	"Abusive and hostile workplace" was established as a precedent for lawsuits, whether or not the employee suffered "psychological damages."
<i>Yates v. Avco Corporation</i> (1987)	The "reasonable woman standard" was established.
<i>Meritor Savings Bank v. Vinson</i> (1986)	A clearer definition of "hostile environment" was provided, indicating that it involves "unwelcome behavior;" stating that the behavior had to be "persistent and severe" for it to be considered "actionable;" and limiting the employer's liability to when the incidents were "known or should have been known."

reasonable woman standard is superior to the reasonable person standard in court cases.

### 1.3. International Laws Against Sexual Harassment

Because the world is globally connected, it is important to examine international laws against sexual harassment. In 1992, a study indicated that only 9 of 23 countries examined had specific laws against sexual harassment. In 2000, 31 countries that had legal statutes against sexual harassment were identified, but it

was suggested that punishment and liability for the behavior vary from country to country due to different legal systems and conceptualizations of the issue.

Legislation in Canada and the European Union (EU) is very similar to that in the United States. An interesting comparison can be made between the United States and France, the model for the EU legislation, in terms of laws relating to sexual harassment (or *harcèlement sexuel*). Quid pro quo and hostile environment sexual harassment laws stemming from a sex discrimination framework have been the basis of lawsuits in the United States, and plaintiffs have received large monetary compensation in civil trials. For example, a \$7.1 million judgment was assessed against a major law firm and one of its former partners. French laws, however, focus entirely on quid pro quo sexual harassment emanating from an authority figure as the basis for allegations of sexual harassment that is considered to be a form of violence against women. Because sexual harassment is part of the French Penal Code, presumably a transgressor may be sent to prison for this offense. However, that consequence is unlikely to occur, and the maximum fine imposed (\$16,000) is far less than that imposed by U.S. juries in sexual harassment cases. French law also does not cover peer harassment. Legislators used anti-American sentiment to denounce initial attempts to develop French sexual harassment laws as “importing American excesses of litigiousness, Puritanism, and the battle of the sexes.”

In the Far East, legislation to prevent sexual harassment has been passed in Australia, Bangladesh, Japan, The Philippines, Sri Lanka, and Hong Kong. Court guidelines against sexual harassment exist in India as well. Although Mainland China has amendments, such as the “Law on Safeguarding the Rights of Women,” these guidelines have been criticized due to a lack of enforcement and penalties.

In Latin America, few countries have well-developed laws prohibiting sexual harassment. This is partly because harassing behaviors are not perceived as unwanted in some of these countries.

## 2. THEORIES AND MODELS OF SEXUAL HARASSMENT

There are three major categories of models: power models, biological and evolutionary models, and an integrative organizational model.

### 2.1. Power Models

Several approaches suggest that power is the dominant force behind sexually harassing behaviors. The organizational model states that because men are disproportionately represented in powerful positions of organizational authority, they use these positions to gain sexual satisfaction. The sociocultural model suggests that men have been socialized and rewarded for being aggressive and dominant and that women have been socialized for being submissive and nonassertive. These rewarded sex role behaviors then transfer into the workplace setting and serve to discourage women from competing in the “man’s” world of work. As an example, one male employee stated that it took 6 months before a woman coworker in his department “got the message that she wasn’t welcome and didn’t belong.” The sex role spillover model proposes that when women are in nontraditional workplaces, they are responded to in terms of gender identity instead of role identity; that is, they are treated differently from male coworkers and are treated as women as opposed to, for example, police officers or construction workers.

Although these power models have received some research support, they appear to be too simplistic to account for all sexual harassment, including peer harassment when there are no power differentials. In addition, these approaches fail to explain why some men exploit their positions and engage in sexual harassment and others do not.

### 2.2. Biological and Evolutionary Models

The biological and evolutionary models suggest that aggressive sexual behavior is natural and expected for men, both in the workplace and elsewhere. Research generally has not supported this approach.

### 2.3. Integrative Organizational Model

An integrative organizational model proposes two essential conditions that are favorable toward the expression of sexual harassment behaviors: climate and gender ratio in the workplace.

### 2.3.1. Climate

This factor refers to whether or not the organization has a climate that is supportive of sexually harassing behaviors. Although there may be an organizational policy against sexual harassment, if victims are afraid to come forward with complaints due to negative consequences, this climate would be classified as supportive of sexual harassment. For example, a female police officer who reported being sexually harassed was told by her supervisor that she could go forward with her complaint but that if she did, she could not expect her backup to appear when she called even if she was in a dangerous situation in the field and her life was on the line.

### 2.3.2. Gender Ratio in the Workplace

Sexual harassment is more likely to occur in a workplace where there are more men than women and where women are performing in nontraditional male-oriented jobs. Law enforcement and the military in the public sector, as well as private sector occupations such as construction work, might be related to this type of category.

The model suggests that sexual harassment has certain consequences for the victim, including lowered job satisfaction that can be associated with work withdrawal or work avoidance (e.g., frequently arriving late for work), health outcomes (e.g., headaches) that can be connected to job withdrawal (e.g., quitting the workplace), and psychological outcomes (e.g., anxiety, depression). The model proposes the following factors that help to determine the employee's reaction to sexually harassing experiences:

- Personal vulnerability (e.g., younger women are more susceptible to harassment)
- Response style (i.e., how the victim attempts to cope with the harassment experience)

Although the integrative model has been supported by U.S. and cross-cultural research, it also has been criticized for not indicating how sexual harassment occurs in the first place and why some men engage in sexual harassment and others do not. In addition, it does not specify the extent to which subjective appraisals, affected by the severity and frequency of the experiences, contribute to the effects of the harassment or the way in which additional job stresses (e.g., racism in the workplace) may interact with sexual harassment.

## 3. MEASURES OF SEXUAL HARASSMENT

A number of methodological problems characterized earlier incidence surveys that produced widely varying estimates of harassment behaviors. For example, two people could accurately state on a general measure that they had been sexually harassed, although the severity of the harassment incidents may have varied considerably. Another problem related to retrospective accounts (e.g., surveys that ask how often the respondent has been sexually harassed during the past 5 years) that are subject to memory issues as well as changes in social norms. Other issues involve the lack of psychometric data for the various scales and response biases in return rate that often limit the usefulness of these surveys.

Currently, the primary measure of sexual harassment in the United States, as well as in some other countries, is the Revised Sexual Experiences Questionnaire (SEQ-R), an 18-item self-report measure developed by Fitzgerald and colleagues in 1995. The SEQ-R is the most reliable and valid measure available and has been used in many settings, including workplaces, universities, and the military. It describes an extensive listing of behaviors that constitute gender harassment, unwanted sexual attention, and sexual coercion. Most important, it does not ask respondents whether they have been sexually harassed until the end of the questionnaire. Studies have shown that respondents often answer that they have not been sexually harassed, even if they have experienced several types of harassment behaviors, possibly because participants have different definitions of sexual harassment, are reluctant to label their experiences as sexual harassment, or perceive the harassment behaviors to be normal or expected.

## 4. INCIDENCE OF SEXUAL HARASSMENT

According to generally accepted statistics, approximately 50% of all women in the United States will experience sexual harassment at some time either in the workplace or in educational settings. Because initial American studies focused primarily on White women, those data are presented first, followed by a consideration of sexual harassment of men, minority women, and cross-cultural victims.

## 4.1. U.S. Incidence Data on Sexual Harassment of White Women

### 4.1.1. Workplace Incidence Data

Although certainly not all sexually harassed women have experienced the level of harassment reported by Michele Vinson, who was forced to have sexual relations with her boss 40 to 50 times during her employment at the Meritor Savings Bank (*Meritor v. Vinson* 1986 court case), the incidence of sexual harassment for American White women is high. Three often-cited landmark studies performed by the U.S. Merit Systems Protection Board (USMSPB) yielded similar incidence rates of sexual harassment of White women. A stratified random sample of 23,964 federal workers in the 1981 USMSPB survey revealed that approximately 42% of the female respondents had experienced sexually harassing behaviors over the previous 2 years. The most common form of harassment consisted of verbal comments or sexual remarks (33%), followed by physical touching (26%), and being pressured for dates (15%), with nearly 10% having been pressed for sexual cooperation and a small percentage having experienced sexual assault. Many women reported having been victimized repeatedly over a long period of time. Young and single women were the most likely targets of harassment behavior.

In another classic, frequently cited study, in 1985, Gutek found that 53% of the women in a community sample that was representative of the workforce had been sexually harassed at least once. Of these respondents, 19.8% had experienced insulting comments, 15.0% had experienced insulting looks and gestures, 24.2% had experienced sexual touching, and 7.6% had been expected to engage in sexual activity as part of their employment. Women in male-dominated occupations experienced sexual harassment at a higher rate than that of women in other occupations.

At least two major studies have been conducted by the U.S. military. In 1991, the U.S. Navy administered the Naval Equal Opportunity Sexual Harassment (NEOSH) survey, which revealed that 44% of enlisted women and 33% of women officers had been sexually harassed, figures comparable to those in the USMSPB studies. More recently, the Department of Defense surveyed 90,000 personnel in all branches of the armed forces. In this survey, 78% of the women had at least one sexual harassment experience. Women in the marine corps had the highest incidence rate in this study (87%), whereas those in the air force had the lowest incidence rate (74%).

### 4.1.2. Academic Sexual Harassment

It is difficult to determine the actual incidence rate of academic sexual harassment because the figures from various studies are quite discrepant, ranging from 13 to 70%. Explanations for these diverse statistics include differences in the type of survey (e.g., specific behavioral questions, a general question about whether or not the respondent has been sexually harassed), varying response rates, representativeness of the educational institution (e.g., most surveys are completed at 4-year colleges or universities and do not sample community colleges), and whether the focus is on faculty harassment or peer harassment. However, even conservative estimates suggest that from 30 to 50% of women who have completed these surveys have been harassed, a figure that corresponds to workplace statistics. Once again, the frequency of milder forms of sexual harassment is higher than the occurrence of sexual coercion. For example, in one study, approximately 30% of the women were subjected to gender harassment, 15% to unwanted sexual attention, and 9% to sexual coercion.

Evidence suggests that sexually harassing behavior may begin at the high school level. A classic study by the American Association of University Women (AAUW), questioning students from the 8th through 11th grades, reported that 85% of the girls (87% of the White girls) had experienced sexual harassment, with 80% of the harassment emanating from male peers and 20% from male adults associated with the school. This figure may be somewhat inflated because methodological issues have been raised about the procedure of the survey, but other research supports the contention that high school sexual harassment is a significant problem that is just beginning to be addressed in the literature. It is important to determine the incidence of peer harassment in this setting because patterns of sexually harassing behaviors that are developed and reinforced as a form of "male bonding" in the high school setting may continue at the university level and in the workplace.

## 4.2. Specific Victim Groups

### 4.2.1. Men as Victims of Sexual Harassment

In the USMSPB studies, from 14 to 19% of men reported being sexually harassed in the federal workplace. In the Department of Defense study mentioned earlier, 38% of men reported sexual harassment experiences, with 35% having experienced gender harassment and less than 1% having experienced attempted

rape. In addition, the AAUW high school study included the statistics that, overall, 76% of boys had been sexually harassed, with 56% having been the recipients of sexual comments and 42% having been touched in a sexual manner.

The impetus for the recent interest in men as sexual harassment victims stems from the 1998 Supreme Court decision in *Oncale v. Sundowner Offshore Services* that stated that same-sex harassment meets the requirement for sexual harassment cases in that the victim is singled out due to his gender. This behavior was classified as producing a hostile work environment regardless of whether the plaintiff in a civil suit was the target of sexual assault (as was the case in the Joseph Oncale lawsuit), was intimidated by male coworkers, or had his masculinity questioned in severe and persistent incidents. Related evidence in the USMSPB study indicated that approximately 22% of the male harassment victims identified in the survey had been sexually harassed by other men (37% in the 1988 study), thereby constituting 91% of all same-sex harassment victims. In the Department of Defense study, 51% of the male respondents reported same-sex harassment experiences as compared with 2% of the women, although overall, women had experienced more than twice as much harassment (76%) as had men (36%). Another study showed that 44% of the men were targets of same-sex harassment, with the majority of the harassment incidents involving "crude/offensive behaviors" such as jokes, gestures, and derogatory remarks about men. Recent studies have used the Sexual Harassment of Men (SHOM) scale, a revised version of the SEQ-R, to identify men who have been sexually harassed. Although male victims generally seem to have fewer negative reactions to their experiences, a sizable percentage of men have experienced hostile environment conditions in the workplace, with the primary form of same-sex harassment being related to attacks on the masculinity of the male targets. The actual incidence rate of male victims may be higher than has been indicated so far given that men might be reluctant to report being victimized due to fear of not being believed or possibly because complaining might go against stereotyped views of appropriate male behavior.

#### **4.2.2. Women of Color**

Few studies have been conducted to examine sexual harassment incidence figures of women of color as compared with those of White women, and the pattern of findings in these studies has been inconsistent. Some

studies concluded that women of color have greater experiences of sexual harassment than do White women, and others illustrated the opposite pattern. Earlier studies often did not provide incidence data broken down by race or ethnicity and often omitted economic status information. If a woman of any race has a low income and low status, these factors may increase her risk for becoming a sexual harassment victim. In addition, it is important to understand that for women of color, sexual harassment (as a form of sex discrimination) and race discrimination often are intertwined. The issue becomes even more complex when women of color are sexually harassed by men of color. For example, women of color had conflicting reactions to the testimony of Hill against Thomas in his Senate confirmation hearings. Although they might identify with Hill as a woman of color, it is possible that their racial identification was more important than gender identification in this situation, and they also may have been reluctant to see Thomas, a Black man, being prevented from succeeding due to Hill's testimony. Another important factor that must be considered is that the category of women of color is too diverse. For example, African American, Hispanic American, and Asian American women may experience differential rates of sexual harassment and respond in various ways to this harassment.

#### **4.2.3. Cross-Cultural Victims**

The methodological issues that exist for cross-cultural comparisons of sexual harassment incidence rates are similar to those described in the preceding section on women of color, but on a larger scale. Lack of information on incidence rates in the countries sampled, use of convenience samples that are not representative, lack of a definition or even the concept of sexual harassment in some countries, use of a variety of measures, and translation issues all are barriers to a definitive comparison of incidence rates cross-culturally. However, these comparisons are significant not only because harassment is a worldwide problem but also due to the presence of multinational companies in various countries. For example, a charge of sexual harassment was directed against a male Japanese plant manager at an American subsidiary of a Japanese company. The American woman employee, who was a supervisor, alleged that the manager had created a hostile work environment by inviting her to dinner several times, giving her gifts, massaging her knee after a ski injury, and trying to kiss her. The manager denied the allegations and explained

his behavior by stating that he always invited groups of workers to dinner, that giving gifts in appreciation was a Japanese custom, that he never tried to kiss her, and that he only wanted to stop her pain by massaging her knee. The company settled her claim, although the president suggested that the manager was just trying to follow orders to become friendly with American workers and that no sexual harassment was involved. This case illustrates the confusion that can arise in identifying sexual harassment in multicultural settings.

This subsection includes data from a sampling of countries. Canada generally has demonstrated incidence rates similar to those of the United States, as have countries in Northern Europe. An International Labor Office survey determined that 30% of Austrians sampled, 13 to 58% of Dutch respondents, and 21% of French citizens had experienced sexual harassment. Although figures vary from 30 to 80% in Germany, a conservative estimate would be that 59% of German women have experienced sexual harassment. In Belgium, 34% of women and 12% of men indicated that they had been sexually harassed, whereas the incidence rate was 48% for Italian respondents and 36% for Finnish respondents, with police respondents reporting the highest rate. Victimization in the United Kingdom ranged from 47 to 90%, with women in the police once again subjected to the greatest amount of harassment.

Data from Far Eastern countries exhibit somewhat higher patterns of harassment. A 2001 survey in Japan indicated that approximately 75% of all respondents had experienced sexual harassment at least once, and 70% of public officers in Korea reported having been sexual harassment victims. However, in 1995, a somewhat low rate of 25% of Hong Kong university students reported that they had been subjected to sexual harassment. In one of the few studies conducted in Pakistan, incidence rates revealed that 58% of nurses, 95% of agricultural workers, 91% of domestic workers, and 93% of women employees in private and public organizations had been victims of sexually harassing behaviors. In India, researchers suggested that harassment was widespread and was actually encouraged due to societal beliefs that men are naturally sexual "predators," women are sexual "prey," and so harassment was expected and considered acceptable.

Little data could be obtained from the Middle East (e.g., Israel was seen as having somewhat similar but possibly higher incidence rates compared with those in the United States, Turkey was described as a society in

which harassment might be extensive due to its patriarchal nature), Latin America (e.g., Brazil was portrayed as having similar rates of gender harassment and sexual coercion, but higher rates of unwanted sexual attention, compared with those in the United States), or Africa (e.g., there are no government statistics on incidence rates in South Africa).

### 4.3. Characteristics of Male Harassers

The major psychometric measure in this area is the Likelihood to Sexually Harass (LSH) scale, developed by Pryor in 1987, that describes 10 scenarios where there is a potential for engaging in sexual harassment without punishment. Men who score high on the LSH scale are portrayed as having traditional masculine sex roles. The classic type of harasser is the boss who engages in "sexual exploitation" or uses his power to gain sexual favors. Another type is a socially inept man who misinterprets women's friendly cues as indications of sexual interest. A third type that is more prevalent in male-dominated workplaces identifies men who are hostile toward women and who see the workplace as the in-group (us or men) against the out-group (them or women). Finally, men who are opposed to homosexuality (i.e., homophobes) may engage in harassment of women and men. Researchers have proposed that men who score high on the LSH scale are more likely to engage in harassment behaviors in workplace settings where harassment is accepted as normative behavior.

## 5. OUTCOMES AND RESPONSES

It is extremely important to examine the consequences of sexual harassment victimization as well as the coping responses that victims use to improve educational efforts in prevention programs, to inform expert testimony in sexual harassment cases, and to train therapists who treat victims of sexual harassment.

### 5.1. Consequences of Sexual Harassment

One analysis suggested that victims may experience the "sexual harassment trauma syndrome," which consists of the following components.

### **5.1.1. Emotional Reactions**

Victims often are angry, embarrassed, and fearful, and they experience learned helplessness or feelings of powerlessness.

### **5.1.2. Physical Outcomes**

These outcomes may include sleeping disruptions, headaches, phobias, and fatigue.

### **5.1.3. Alteration in Self-Perception**

Victims often suffer from lowered self-esteem and self-confidence. For example, employees may perceive themselves as being promoted due to their sexual attractiveness rather than their competence.

### **5.1.4. Behavior Change**

If victims feel responsible for the harassment, they may try to dress more conservatively and wear “less sexy” clothes. Their trust in men often is adversely affected, and they may avoid interpersonal relationships. As a result, friendship and alliances with coworkers might not develop, and victims often are excluded from mentoring relationships that can enhance their careers. In terms of academic sexual harassment, students might decide to change careers, or at least have their research experiences curtailed, as a result of their harassment experiences.

The integrative theory described earlier proposes that sexual harassment victims experience “work-related withdrawal” or attempts to remove themselves from the job environment. These effects can include lowered productivity, decreased concentration on work-related tasks, absenteeism, and being late for work. Another consequence may be “job-related withdrawal,” which refers to individuals’ plans to retire early or quit their jobs. In addition, the theory predicts that health-related effects will occur, similar to the physical outcomes described previously, as will damaging psychological effects from depression and anxiety. Although sexual coercion clearly will produce more extensive negative effects than will milder forms of harassment, there is evidence that the cumulative effects of ongoing persistent sexual harassment of any type will have a deleterious impact on the victim.

This “poisonous climate” of sexual harassment may affect the organization or educational institution by lowering organizational commitment and damaging

the reputation of the company or educational institution. The cost of sexual harassment to the organization or educational institution can be severe and includes turnover, increased medical costs, low morale, lowered productivity, and absenteeism. It generally is impossible to determine the actual human and monetary costs of this behavior, although the USMSPB’s 1988 study suggested that sexual harassment may have cost the U.S. government \$250 million over 2 years. Of course, if sexual harassment complaints reach a lawsuit level, the costs can be enormous.

## **5.2. Coping Responses**

One of the common reactions to the testimony of Hill, alleging that she had been sexually harassed by Thomas several years earlier, included “Why didn’t she just leave?” and “Why didn’t she tell him to stop the behavior?” These responses represent direct coping attempts to eliminate the harassment. If a victim does not behave according to these public assumptions, the situation might be judged as insufficiently offensive or might not be acknowledged as a case of sexual harassment. The frequency of use of different types of coping and the effectiveness of these various coping responses must be examined to assist harassment experts in advising victims. In addition, if researchers clearly indicate that most victims do not report harassment incidents and that victims often endure these incidents for long periods of time without any direct response (as was the case with Hill), organizations may be persuaded to adopt procedures that will discourage harassment and encourage victims to report any harassment to supervisory personnel.

### **5.2.1. Types of Coping Responses**

Although several category systems exist, the most useful approach divides the type of response into direct (active) versus indirect (passive) reactions to sexual harassment incidents. The most basic direct reaction to such an incident is to tell the harasser to stop the behavior in a face-to-face confrontation or in a letter. The next step is to report the behavior to a supervisor. The victim then may escalate the process to filing a formal complaint, although on the average only 1 to 7% of victims reach this stage. If these processes do not stop the harassing behavior, the final direct approach is to institute a lawsuit.

A number of indirect or passive responses to harassment have been noted. A victim may ignore the

behavior, trivialize the incident, redefine the behavior (e.g., "He's just fooling and doesn't mean it"), and/or attribute it to the harasser's unhappy marriage or the victim's provocative clothing. Alternative passive responses include avoiding the harasser, redirecting the conversation away from personal comments, and bringing a friend or colleague to meetings with the harasser. Avoiding the harasser is difficult if the perpetrator is a supervisor or professor. A victim may also seek social support by venting to friends and family. Finally, if the behavior continues, the individual may quit her job, transfer to another department, or drop the course (in the case of academic sexual harassment).

The most common response of victims is to ignore the behavior or attempt to avoid the harasser. However, harassers generally continue to behave inappropriately, and if no action is taken against these individuals, the harassment may escalate. Sexual harassment experts usually advise victims to respond directly to harassers, but there is a disconnect between this advice and the behavior of sexual harassment victims. The major reason why victims do not respond assertively is fear of retaliation or the perception that this is a high-risk coping response. Victims often are in low power positions and are afraid of losing their jobs if they complain about harassment, especially if the harassers are in supervisory roles. For example, a woman in a relatively high power position told her harasser to "cut it out," but when asked why she did not report him, she said that she wanted to hold onto her job. Studies indicate that using direct confronting types of coping styles often backfires for the employees in that one-fourth to two-fifths of individuals who complained were forced from their jobs. In one survey, 75% of the victims who directly confronted their harassers found that the behavior continued after these confrontations. In fact, some evidence suggests a paradox for victims. If they respond more assertively, there is a possibility that the behavior will stop and the harasser will be punished, but these victims also may experience more negative job-related outcomes (e.g., in terms of others' impressions of them and their commitment to the job) than will those who respond more passively.

A second related reason why assertive coping often does not occur is that complainants often are labeled as "troublemakers" or treated as "whistleblowers." For example, a female junior faculty member who reported harassment incidents by a male senior professor was told that he was a more valuable employee and that she could leave if she intended to persist in her complaint. In Japan, where "saving face" is a core value, victims

often are blamed for their harassment experiences and are seen as bringing shame to their companies' reputations and preventing the harassers from succeeding in their positions. When an employee goes forward with a complaint, she is suspended without pay while the accused aggressor remains in his job.

Another interpretation suggests that if the climate is supportive of aggressors, victims may avoid reporting this type of behavior because they believe that nothing will be done to stop the harassers. Finally, there are individual differences in the tendency to use direct coping styles, with some victims feeling too embarrassed or humiliated by their experiences to lodge complaints. Sexual harassment experts should reexamine their advice to victims considering the tendency of employees to adopt nonassertive responses to the offensive behavior.

## 6. PREVENTION AND INTERVENTION

As a result of increased public attention to the issue, most major companies have developed policies against sexual harassment and a number of consulting firms are specializing in training programs to prevent sexual harassment. Prevention is the preferred approach to eliminate harassment in the workplace and educational settings because once inappropriate behavior occurs, it results in a "lose-lose" situation for all concerned. No matter how the situation is resolved, the company will incur human and financial costs as well as potential damage to its reputation, the victim will experience negative outcomes, and the accused harasser, if found to be guilty of sexual harassment, will be subject to sanctions that may extend to termination of employment. In addition, the company could even be sued by the victim for not preventing the harassment and by the defendant for defamation of character or for terminating him without "due process."

If top management is only motivated to protect itself from lawsuits, a simple broad policy statement may be the extent of its prevention attempts. However, if top management truly wants to create a harassment-free culture or environment, the approach will be more extensive and will involve employees at every level. Even with the best of intentions, there are no sure-fire programs that can guarantee a positive climate in the workplace. In fact, research suggests that "one size does not fit all"; in other words, programs must be developed

within the framework of specific companies. Another crucial element is evaluation of the effectiveness of the prevention efforts. To date, critics have complained that most training programs have not been established as valid in terms of prevention. One of the difficulties is to determine the outcome measure. If the program is effective, will the number of complaints increase given that victims will feel more comfortable in filing complaints, or will the number decrease because sexual harassment behaviors have been prevented?

One approach suggests that there are three important stages of prevention and intervention in sexual harassment situations. The first relates to policies explicitly stating that sexual harassment will not be tolerated in the organization. The policy statement also should include specific examples of inappropriate behavior with associated penalties, should provide contact sources for victims, and should perform an educational function. Most important, this policy should be disseminated to all employees (e.g., through handouts to new employees when hired as well as through systematic distribution of the guidelines to other employees) and supported from top management down throughout the hierarchy. Unfortunately, policy statements alone are not effective, particularly if the climate or culture supportive of sexual harassment is not altered.

The second component is procedures that have been developed as part of the intervention process. Employees must be informed that if a formal complaint of sexual harassment is brought against an individual, both the rights of the accused and the accuser will be protected. At this point, a "neutral" investigator (an in-house employee or an outside consultant) should be in place to determine the validity of the charges.

As discussed earlier, however, the formal complaint route probably will not be chosen by many victims due to fear of retaliation or fear of being labeled as a "troublemaker." In addition, victims might not want their harassers to be sanctioned or punished. Therefore, the informal complaint approach might be more effective in some settings. A neutral negotiator is identified for victims, and face-to-face meetings may be arranged between victims and perpetrators. In many cases, all the victims want is for the harassment to stop. Possible outcomes of this informal process include stopping of the behavior with or without an apology, transfer of one or both of the individuals, and (in extreme cases) resignation of one or both of the individuals. Informal procedures have been criticized for lack of standardized objective procedures that may lead to ineffective resolution of the problem situation.

The third step in the intervention process is training given that policies and procedures alone often have been shown to be ineffective in preventing harassment. Training should be used to sensitize employees to types of sexual harassment and effects on victims, should include sanctions for harassment behaviors, and should occur over extended periods of time. These programs should include not only an informational component but also interactive exercises such as role-playing and watching videos with actors, followed by intensive discussions. Training programs have been shown to increase understanding of the behaviors constituting harassment but might not prevent these behaviors. In addition, if programs are voluntary, the people attending might not be the potential harassers, and if programs are mandatory, resentment and resistance might occur. In one study, for example, males scoring high on the LSH scale showed a small boomerang effect after watching a training video.

Creative and idiosyncratic approaches may effectively supplement the three basic stages described previously. At one university, in addition to training films, lectures, and a "sexual harassment awareness week," core faculty and staff volunteered to be resource individuals for students who had experienced sexual harassment. This group was trained not only in referral sources and policies and procedures but also in communication skills so that the faculty could counsel students and be supportive during these stressful situations. Once again, it is important for organizations to determine appropriate ways in which to implement these three steps and to develop effective means of preventing harassment.

## 7. CONCLUSION

Sexual harassment is a worldwide problem that causes distress to its many victims and damages the various organizational climates. More attention should be directed at developing adequate prevention and intervention approaches that contain an evaluation component to establish the effectiveness of these programs.

### See Also the Following Articles

Coping ■ Feminist Psychology ■ Gender and Culture ■ Gender Role Development ■ Homophobia ■ Homosexuality ■ Occupation and Gender ■ Prejudice and Discrimination ■ Risk-Taking and Sexual Behavior ■ Sexual Behavior ■ Sexual Behavior and Culture

### Further Reading

- Blumenthal, J. A. (1998). The reasonable woman standard: A meta-analytic review of gender differences in perceptions of sexual harassment. *Law and Human Behavior, 22*(1), 33–57.
- Equal Employment Opportunity Commission. (1980, April). *Guidelines on discrimination because of sex under Title VII of the Civil Rights Act as amended: Adoption of interim interpretive guideline*. Washington, DC: Government Printing Office.
- Fitzgerald, L. F. (1993). Sexual harassment: Violence against women in the workplace. *American Psychologist, 48*, 1070–1076.
- Fitzgerald, L. F., Drasgow, F., Hulin, C. L., Gelfand, M. J., & Magley, V. J. (1997). Antecedents and consequences of sexual harassment in organizations: A test of an integrated model. *Journal of Applied Psychology, 82*, 578–589.
- Fitzgerald, L. F., Gelfand, M., & Drasgow, F. (1995). Measuring sexual harassment: Theoretical and psychometric advances. *Basic and Applied Social Psychology, 17*, 425–445.
- Gutek, B. A., & Koss, M. P. (1993). Changed women and changed organizations: Consequences of and coping with sexual harassment. *Journal of Vocational Behavior, 42*, 28–48.
- O'Donohue, W. (Ed.). (1997). *Sexual harassment: Theory, research, and treatment*. Needham Heights, MA: Allyn & Bacon.
- U.S. Merit Systems Protection Board. (1995). *Sexual harassment in the federal workplace: Trends, progress, and continuing challenges*. Washington, DC: Government Printing Office.



# Social Comparison and Subjective Well-Being

Thomas Ashby Wills and Don Mendoza

*Albert Einstein College of Medicine, Bronx, New York, USA*

1. Is Social Loafing a Universal Effect with Disastrous Consequences for Group Performance?
2. A Brief Historical Overview
3. Theoretical Explanations
4. Implications of Social Loafing
5. Prevention and Control of Social Loafing

## GLOSSARY

**downward comparison** Comparing one's own attributes with those of another person who is worse off on the attributes.

**self-enhancement** Using social information to feel better about one's own situation.

**self-evaluation** Using social information to obtain an accurate assessment of one's own standing.

**self-improvement** Using social information to improve one's own performance.

**threat-derogation effect** Making more negative ratings of other people's attributes when one is threatened, in order to increase perceived social distance between the self and the other.

**upward comparison** Comparing one's own attributes with those of another person who is better off on the attributes.

Social comparison research considers how people's perceptions of their own situation are influenced through comparison with the attributes and situations of

other people. Research has identified processes of upward comparison and downward comparison, and studies have linked these to subjective well-being. Downward comparisons may serve to increase subjective well-being, particularly in the short term after people have experienced setbacks or failures. Upward comparison can be used for self-improvement, but evidence from field studies suggests that a constant background of unfavorable comparisons is negatively related to psychological and physical outcomes. Costs and benefits of comparison processes, and their implications for intergroup relations, are discussed in this article.

## 1. SOCIAL COMPARISON THEORY

Subjective well-being (i.e., feelings of positive mood and life satisfaction) is an important variable for applied psychology, both for its theoretical interest and for its relations with health and productivity. Social comparison theory is relevant for the study of subjective well-being because comparisons with other people can influence how people feel about themselves and their own situation. Demographic characteristics such as gender and ethnicity are not strongly correlated with well-being indices, so research has shifted toward considering basic psychological processes that

contribute to lower vs higher levels of subjective well-being. Social comparison is one of these processes.

Social comparison theory was developed to address the question of how people assess their own attributes and abilities when objective measures are not available. For some characteristics (e.g., weight) objective measures are readily available, but this is not so much the case for important personal attributes (e.g., attractiveness). For such attributes, when people ask themselves "Where do I stand on attribute X?" an answer must come from comparing the self with others. The form of the answer, then, is "Comparing myself with the attribute for persons A, B, C, and D, my standing is most like that of B." The outcomes of these comparisons can have important implications for whether people feel happy or unhappy about themselves and their own situation.

### **1.1. Upward Comparison**

The original formulation of social comparison theory by Leon Festinger was based on the assumption that people want an objective evaluation of their own attributes (e.g., academic competence). This led to the concept that people would tend to compare themselves with others who were better off than the self on the attribute in question, termed upward comparison. The predicted consequence of such comparisons was that uncertainty about one's own standing would be reduced and therefore one would feel more certain about performance (or self-improvement) in this domain. Support for these predictions was obtained by studies that showed a preference for comparison with slightly better-off others, mostly from contexts in which the subjects were already doing pretty well on the attribute in question and felt confident about attaining the status of the better-off comparison others. Further developments by Goethals and Darley led to the recognition that in the comparison process, people would also consider additional attributes that are believed to be related to the attribute in question (e.g., intelligence and time spent studying).

### **1.2. Downward Comparison**

The issue for this model of human functioning was what would happen when people believed that they were not doing so well on the attribute in question, a context that could be true for part of the population. This led to the formulation by Wills of downward comparison theory, which considers the types of

comparisons that occur in situations in which people are not so well off to begin with. For this context, the prediction was that people would tend to compare with others who are worse off than the self, termed downward comparison. The predicted consequence of this type of comparison was that people would feel better about themselves and their own situation. Support for this theory was obtained in situations in which people felt that their own status was poor or threatened. For example, downward comparisons were found to be common among people with physical illness, who reported ways in which their own situation was relatively favorable compared to others, and media studies showed that people felt better about their own city when the local newspaper portrayed a high level of crime in other cities. Further development by social psychologists such as Taylor and Lobel led to recognition that some downward comparisons may be with cognitively constructed images of typical others, and cognitive comparisons may have different consequences than forced contact with other people having problems.

## **2. RESEARCH ON PRINCIPLES OF SOCIAL COMPARISON**

A number of laboratory studies have tested propositions from social comparison theory. Because social comparison is a cognitive process and hence not directly observable by a third party, findings from laboratory research have been helpful for understanding comparison processes because known variables can be experimentally manipulated and conclusions about causality can be drawn from the observed effects of these variables.

Results from laboratory studies typically show a downward shift in comparison preferences when distress, threat, or failure is present. When people feel they are doing well, they tend to prefer information about others who are better off than the self. However, when negative mood is experimentally induced or when subjects are identified who have experienced a recent setback (e.g., in academic performance or social relationships), then there is a preference for information about others who are having similar problems or worse problems. This is consistent with the basic principles of downward comparison theory. Another aspect of social comparison is that threatened people may work to improve their own self-image through

derogating the characteristics of other people or groups, a process that represents the potential dark side of social comparison and has been termed the threat-derogation effect. This process has been demonstrated consistently among threatened people in laboratory settings, and research on intergroup perceptions by Brewer and others has established this effect in downward comparisons with outgroups.

The predicted consequences of social comparison have also been observed in the laboratory. When distressed people get access to information about worse-off others, they typically show an improvement in subjective well-being—that is, reduced negative affect or increased positive affect. Effects of upward comparison tend to depend on the situation, but several studies have shown that when subjects are exposed to information about others who are better off (more attractive, more intelligent, better adjusted, etc.), the subjects show increases in negative affect. The primary moderator of upward comparison effects is the controllability of the problem. When people feel that a difficulty is easily solvable and they can engage in self-improvement, they tend to react more favorably to upward comparisons; however, when they perceive their problems are not controllable, they react more negatively.

Laboratory research has also clarified that there are several possible motives for social comparison. Most common are self-enhancement, using comparisons to obtain a favorable picture of the self (e.g., I compare “to feel good about my own situation”); self-assessment, using social information to obtain a strictly accurate evaluation of one’s relative standing in a social group (e.g., “to see how I’m doing”); and self-improvement (e.g., “so I can get better”). These motives are all present in the population to varying degrees and may be activated under different conditions. This research has been extended in applied settings (e.g., with business entrepreneurs) to consider how social comparison motives for making money are related to subjective well-being.

There is also evidence of consistent individual differences in social comparison. For example, some studies indicate that women engage more often in all types of social comparison. Other research has indicated dispositional aspects of social comparison. A dimension called comparison interest indicates that some people are more interested, in general, in pursuing social comparisons. Also, a dimension termed dispositional envy has been identified, in which some people are particularly disposed to feel envious toward better-off others, and hence unhappy, because of comparison processes.

### **3. RESEARCH ON SELECTED AND FORCED COMPARISONS IN FIELD SETTINGS**

#### **3.1. Selected Comparisons**

Research in field settings has examined several questions about how social comparisons are related to subjective well-being. A number of studies have been conducted on people subjected to distress from physical illness, such as heart disease or cancer. Downward comparisons are frequent in these populations, with people reporting that they can think of others who are worse off in some respect. When patients are studied in clinic settings, in which a range of comparison others are available, a paradoxical aspect is also noted to social comparison in these populations. Helgeson and others have shown that patients show a preference for affiliating with other patients who are recovering from the disease, apparently because this makes them more optimistic about their own recovery. At the same time, however, they tend to compare cognitively by thinking about others who are worse off, and this is related to greater subjective well-being as indicated by feelings of current mood and perceived coping ability.

Studies of individuals with mental illness are fewer in number and still paint a somewhat confusing picture. People with clinical depression, whose basic issue is that they feel very negative about the self, seem less inclined to engage in active, self-aggrandizing comparisons, possibly because of a breakdown in self-enhancement processes. However, there is evidence that even moderately unhappy people find ways to engage in indirect forms of self-enhancement, and can be creative about identifying unique aspects of the self on which they perceive themselves as advantaged relative to others, as a way of maintaining some level of self-regard. It has been suggested that depression may be brought on by excessive reliance on social information and an inability to discount (or ignore) unfavorable comparisons, but there has been no longitudinal research in this area and the answer to the question remains unresolved.

#### **3.2. Forced Comparisons**

The other question about social comparison in real-world settings concerns what happens when people are repeatedly exposed to comparison information that they cannot avoid. One example of forced comparison is school settings, in which students receive a

considerable amount of information about others' performance, either through public postings and awards or through more informal communication. Findings from studies by Marsh and others have been consistent in showing that students feel worse about themselves, in terms of academic self-concept, when they are in schools in which many students are performing better than they are performing. Conversely, people feel better about themselves when they are in school environments in which most students have only average performance. The suggestion has been that constant exposure to unfavorable comparisons makes people feel less efficacious and less inclined to exert effort for self-improvement. These findings have been summarized as indicating the "big fish, little pond" effect—the concept that in terms of subjective well-being it can be better to be a relatively big fish in a small pond, even if one does not learn to swim quite as well.

Another noteworthy question is how people respond to comparisons of income, an attribute that may not be publicly posted but is still clearly evident in daily life. Studies using evidence from aggregate statistics and public health data have been consistent in showing income distributions in the population related to individuals' reports of subjective well-being. For example, reports of happiness and life satisfaction are decreased when there is greater inequality in distributions of income (i.e., when the average individual sees more people with higher income than him or her). This has been demonstrated both within the United States and across different countries. Also, evidence from studies using several different indices of inequality has shown that mortality rates are higher in census units where there is greater inequality of income. Thus, forced comparisons have definite consequences for subjective well-being.

#### **4. CONCLUSIONS AND IMPLICATIONS**

Social comparison is an important contributor to subjective well-being. Laboratory studies have established basic social comparison principles in controlled settings, and field studies have tested how these processes operate in the natural environment. This research has shown that when people perceive they are doing well they tend to engage in upward comparison, but when people are stressed or threatened there is a downward

shift in comparison processes. Downward comparison has been supported as a selective process of using comparison information so as to enhance subjective well-being. It is probably a mixed blessing as a coping strategy, if exclusive reliance on downward targets serves in the long term to deter self-improvement. The available theory and evidence suggest that downward comparison is best as a coping strategy when used judiciously in times of distress so as to maintain self-esteem and provide a bridge to sustained, active coping efforts.

Comparison with better-off others has been supported as a method to self-improvement, but aspects of the research suggest that upward comparison can be a double-edged sword. Findings from field studies have shown forced upward comparisons to be related to lower levels of psychological and physical well-being. Laboratory research suggests that the comparison process may become so automatized that people may not always be aware of the impact of social comparisons, particularly when they are cognitively busy dealing with the demands of daily life. This work suggests that people can sometimes "undo" a social comparison by discounting the relevance of the comparison dimension, but a constant background of unfavorable comparisons could be difficult to undo and could work to decrease subjective well-being. Linkages of forced comparisons to physical health have been identified, but the processes underlying these effects are not well understood and implications of this evidence for applied psychology need to be considered.

Social comparison research also has applied implications because of linkages to intergroup relations. The concern is that the threat-derogation effect may be operative among people whose subjective well-being is threatened because they believe that they are not doing well, their problems are uncontrollable, and their situation may not improve. This concern is bolstered by archival studies that show stereotyping and authoritarian attitudes in the U.S. population to be increased in times of threat (e.g., from economic recession). Whether this effect is related to derogatory attitudes toward endogenous targets (e.g., ethnic minorities), or other groups such as immigrants, has not been extensively studied in applied research in field settings and remains a topic of importance for investigation.

#### **See Also the Following Articles**

Well-Being

### Further Reading

- Banaji, M. R., & Prentice, D. A. (1994). The self in social contexts. *Annual Review of Psychology*, 45, 297–332.
- Bogart, L. M., & Helgeson, V. S. (2000). Social comparisons among women with breast cancer. *Journal of Applied Social Psychology*, 30, 547–575.
- Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology*, 66, 268–275.
- Buunk, B., & Gibbons, F. X. (Eds.). (1997). *Health, coping, and social comparison*. Mahwah, NJ: Erlbaum.
- Hagerty, M. R. (2000). Social comparisons of income in one's community: Evidence from national surveys of income and happiness. *Journal of Personality and Social Psychology*, 78, 764–771.
- Kawachi, I., Kennedy, B. P., Lochner, K., & Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American Journal of Public Health*, 87, 1491–1498.
- Lyubomirsky, S., & Ross, L. (1997). Hedonic consequences of social comparison: A contrast of happy and unhappy people. *Journal of Personality and Social Psychology*, 73, 1141–1157.
- Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Models of the big fish–little pond effect on academic self-concept. *Journal of Personality and Social Psychology*, 78, 337–349.
- Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their implications. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York: Russell Sage Foundation.
- Srivastava, A., Locke, E. A., & Bartol, K. M. (2001). Money and subjective well-being: It's not the money, it's the comparison motives. *Journal of Personality and Social Psychology*, 80, 959–971.
- Suls, J., & Wheeler, L. (Eds.). (2000). *Handbook of social comparison*. New York: Kluwer Academic/Plenum.
- Taylor, S., & Lobel, M. (1989). Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review*, 96, 569–575.
- Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, 90, 245–271.
- Wills, T. A., & Sandy, J. M. (2001). Comparing favorably: A cognitive approach to coping through comparison with other persons. In C. R. Snyder (Ed.), *Coping with stress: Effective people and processes* (pp. 154–177). New York: Oxford University Press.
- Wood, J. V., & Lockwood, P. (1997). Social comparisons in dysphoria and low self-esteem. In R. Kowalski & M. Leary (Eds.), *The social psychology of emotional and behavioral problems* (pp. 97–135). Washington, DC: American Psychological Association.



# Social Loafing

Francisco Gil

Complutense University, Madrid, Spain

1. Is Social Loafing a Universal Effect with Disastrous Consequences for Group Performance?
2. A Brief Historical Overview
3. Theoretical Explanations
4. Implications of Social Loafing
5. Prevention and Control of Social Loafing  
Further Reading

highly familiar or overlearned tasks are performed and the correct responses are dominant.

**social loafing** A reduction of individual effort when individuals perform collective tasks, as compared with working individually or alongside coactors.

**sucker effect** When individuals decide to reduce their own effort and input to group performance (and hence diminish the reinforcements they could receive) instead of being “suckers” and letting themselves be exploited by free riders.

## GLOSSARY

**additive tasks** Tasks in which members' inputs are added to produce a group result (e.g., generating the highest number of ideas in a brainstorming session); group productivity depends on all members' contributions.

**coactive tasks** Tasks in which people work individually in the presence of others who are performing the same task (e.g., doing administrative work in a collective office).

**conjunctive tasks** Tasks that must be performed by all group members (e.g., team cycling competitions); group productivity depends on the worst group member (e.g., the slowest cyclist).

**disjunctive tasks** Tasks that may be performed by one member of the group (e.g., solving a problem); group productivity depends on the best (in this case the most expert) group member.

**free riding** When individuals who have not made appropriate contributions to the task profit from the contributions of other group members.

**social facilitation** An increase of individual effort when individuals perform tasks in the presence of other people or alongside coactors; this effect tends to occur when

The social loafing effect is related to the loss of motivation that occurs in a group context, causing people to reduce their effort or to produce less when they carry out collective tasks than when they work alone. This effect is the opposite of the social facilitation effect, which is related to the increased achievement that occurs when performing individual tasks in the presence of others. Social loafing is related to two other effects that occur in certain enhancing conditions: the free rider effect, which occurs when individuals who have not made an appropriate contribution to the task profit from the inputs of other group members, and the sucker effect, which occurs when individuals decide to reduce their own effort to group performance rather than let themselves be exploited by free riders. Different theories (e.g., social impact, arousal reduction, identification and evaluation potential, dispensability of effort, self-efficacy, collective effort) have questioned the original supposition that social loafing is a universal process inherent to work in groups and have identified

different conditions in which it is more likely. Various factors, mainly those related to tasks, groups, and organizations, have also been found to prevent and control the social loafing effect.

### **1. IS SOCIAL LOAFING A UNIVERSAL EFFECT WITH DISASTROUS CONSEQUENCES FOR GROUP PERFORMANCE?**

Collective work and cooperation among individuals are fundamental to the establishment of interpersonal relations and bonds, the creation and functioning of organizations, and the development of society. Throughout history, the great achievements of humankind have been the result of this kind of cooperative action. Intuitively, it is assumed that working with others in a group has an energizing effect, causing people to work hard. In addition, the beneficial effects of groups are generally acknowledged (e.g., “unity is strength,” “many hands make light the work”), and collective efforts are rewarded. The performance of people working together in teams can be very high if the people are ready to cooperate with each other.

Nevertheless, people often witness the contrary, as was revealed by the early research of social psychology on group performance. People carrying out collective work are not always willing either to cooperate or to work hard. Under such conditions, they try to go unnoticed, hiding in the group or crowd and avoiding contributing fully to the collective product. At the same time, they are suspicious of and distrust their fellow group members and worry that others might not make their respective contributions.

One important research tradition confirms that people tend to make less effort when their contributions are combined with those of other people, in contrast to the effort and interest they show when performing individually. In such circumstances, it may be observed that “the whole is less than the sum of its parts.” This effect, known as social loafing, is related to two other effects, free riding and the sucker effect, and has been found to be sufficiently robust and generalized among the different types of tasks and populations to be considered universal, that is, an inherent feature of group performance. If this is the case, concern is warranted because groups are an ever-present feature of organizations and society. Most work is carried out in groups, and a large part of this work involves performing tasks in which the

respective contributions of all members are combined. Is social loafing a universal effect, and does it have disastrous implications for group performance?

To address these issues, verify the extent of social loafing, identify how it works and what consequences it has, and find out how to prevent it or counteract its effects, this article analyzes this phenomenon throughout the historical evolution of the research carried out and the main theoretical models used to explain it. Finally, it presents practical suggestions for preventing and controlling social loafing.

### **2. A BRIEF HISTORICAL OVERVIEW**

Scientific interest in social loafing is relatively recent. In 1913, Ringelmann performed an experiment that consisted of asking volunteers to pull on a rope as hard as possible, both individually and in groups of various sizes. He discovered a nearly linear reduction in the amount of strength exerted in groups compared with expectations based on individual performances. Although average group productivity increased, the result fell short of the maximum predicted as new members were added to the group. These results drew attention to the possibility that the performance of collective tasks could have a demotivating effect on group members. Initially, Ringelmann surmised that this could be due to coordination problems. Thus, the individuals might be having difficulty in combining their efforts in a way that was fully effective, and these difficulties would be more likely to emerge in large groups. A well-known “law” established by Steiner in 1972 states that a group’s real productivity is equal to its potential productivity minus losses due to defective processes such as coordination.

Ringelmann’s experiment was not replicated until 1974, when a number of inquiries into the causes of social loafing were undertaken. Some ingenious experimental strategies were designed in an attempt to reject the original explanation of coordination-related losses. In the replication of Ringelmann’s experiment carried out by Ingham and colleagues in 1974, the participants were blindfolded to prevent distraction and placed at the head of a group with instructions to pull on a rope. However, the rest of the group members were actually the experimenter’s confederates, and they did not really pull on the rope. In fact, the “group” was fictitious and had only one member who believed that he or she was part of a team. The outcome observed by Ringelmann was also obtained in this situation. That is, as the

apparent size of the group increased, performance dropped, confirming that the effect was not due to coordination difficulties; rather, it was due to motivation loss related to the size of the group.

Latané and colleagues obtained similar results in 1979 when they replicated the experiment with a different task. Blindfolded participants who also wore earplugs were asked to cheer or clap as loud as they could. Again, motivation decreased as the apparent size of the group increased. Participants seemed less willing to make an effort if they felt that others might not be doing so. The authors called this reduction in the effort of members when working in a group social loafing, and they claimed that the phenomenon was widespread—a kind of “social disease” with negative consequences for individuals, institutions, and society. By way of example, they mentioned the former Soviet Union’s collective farming systems.

Subsequent research confirmed the consistency and general prevalence not only of social loafing but also of two other closely related effects: free riding and the sucker effect. Thus, it was shown that social loafing was not confined to physical tasks (e.g., rope pulling, cheering) but also occurred in other cognitive, creative, evaluative, and work-related tasks. Moreover, it affects both men and women, as well as people of all ages, and was to be found in both Western and Eastern cultures. (The experiments were replicated in countries such as India, Japan, and Taiwan.)

Contrary to the original supposition, however, research has revealed that there is nothing inherently demotivating in the group context. Thus, social loafing cannot be considered a feature of group performance versus individual performance because it arises only under certain specific conditions, whereas in many other situations working in groups actually increases motivation.

### 3. THEORETICAL EXPLANATIONS

A variety of theories have been proposed in an attempt to explain the social loafing effect. Although all of these have their limitations and are open to criticism, they have allowed researchers to progressively expand their knowledge of social loafing. A summary of the main theories follows.

#### 3.1. Social Impact Theory

The social impact theory treats social influence as a function of the intensity, immediacy, and number of

sources and targets present. When a source (a person) insists on hard work during the performance of a collective task, demands from the source are diluted among the various targets (all group members) and, therefore, have little impact. Effort drops. People feel less responsible for task performance as the size of the group increases.

#### 3.2. Arousal Reduction

Another explanation, which builds on the social impact concept, is based on arousal reduction theory. Research on social facilitation indicates that arousal increases when a person performs a task in the presence of others (i.e., a coactive task). This causes people to improve their performance in simple or familiar tasks (dominant responses) but has the opposite effect when they are asked to carry out complex or unusual tasks. However, the situation is different with social loafing. People are performing collective tasks rather than coactive tasks, and in this case the participation of others reduces arousal because the group’s members as a whole become the targets of external social influence. Thus, different results are obtained depending on the circumstances. In complex tasks, individuals might do better working collectively than coactively because they do not feel as though they are the exclusive targets of attention.

#### 3.3. Identification and Evaluation Potential

Many interpretations of social loafing refer to the absence of identification and evaluation of individual effort in the performance of collective tasks. Because the input of each individual is combined with the inputs of others to produce a single group result, the contribution made by each individual becomes indistinguishable from the contributions made by other members of the group. As the size of the group increases, so does anonymity. This makes it easier for individuals to shirk—to “get lost in the crowd”—while nobody can blame them or make them responsible for underperformance. If, on the other hand, each person’s input is identifiable (e.g., if a microphone is placed in front of each person who is asked to clap or cheer in the experiment described earlier), social loafing drops. However, it is not sufficient for the input simply to be identifiable. It must also be evaluated, not only or necessarily by an external observer or the group but also by the individuals themselves.

From this point of view, social facilitation and social loafing can be considered two opposite phenomena explained by the same process, that is, the concern with evaluation. This contributes to heightened performance not only when simple tasks are performed individually alongside coactors but also when complex tasks are addressed collectively.

### 3.4. Dispensability of Effort

An important exception occurs when individuals perform disjunctive tasks, requiring input from only one member of the group. Under these conditions, individuals who believe that their efforts do not affect final group performance, although the input may be perfectly identifiable and capable of evaluation, seem unwilling to make the effort required to perform the task because they believe that other members will achieve the same result with less effort. If the most capable or skilled individual shares this perception, he or she will probably be ready to compensate the others. Otherwise, the free riding effect (i.e., other members not doing their share) and the sucker effect (i.e., unwillingness to let other group members exploit oneself) are likely to emerge. However, if the task is conjunctive (i.e., all group members are needed to perform the task), the effort put in by the most skilled individual is likely to fall.

### 3.5. Self-Efficacy

The two opposite phenomena of social loafing and social facilitation can be explained through the self-efficacy theory. Thus, the effort that individuals are willing to make will depend on the level of self-efficacy for the task. If they perceive themselves as having high self-efficacy for the task, they will probably be willing to work hard when they perform alone or alongside coactors (social facilitation). Individuals in this situation feel responsible and self-confident about their performance. In contrast, individuals who believe that they have low self-efficacy for the task are less likely to work hard, especially when their performance is being evaluated (social loafing).

### 3.6. The Collective Effort Model

The collective effort model (CEM) is an integrative model that incorporates elements from a variety of theories, including the expectation–value models of effort, the social identity theory, and self-evaluation

processes. The expectation–value theory states that individuals make more effort under certain circumstances such as when they believe that (a) they will achieve more if they work hard (expectation), (b) good performance will be rewarded (instrumentality), and (c) the rewards are valuable and desirable (value or importance).

CEM extends this theory to account for individual motivation in group settings. In collective tasks, effort depends on the perceived relationship among (a) individual and group performance, (b) group performance and results, and (c) group and individual results. According to the theory, individuals seek to maximize the expected utility of their actions. Therefore, they will make an effort in performing a collective task if they consider their efforts to be instrumental in obtaining results that they value personally. The model assumes that the links among the different variables are weaker when performing collective tasks, so that individuals' perceptions of the instrumentality of their efforts to achieve valued goals are more uncertain when working in groups.

The determining factor in the evaluation process is not so much the result itself as an individual's own evaluation based on certain objective and subjective criteria. Among the latter, one important criterion is the group to which an individual belongs and identifies. As social identity theory explains, to the extent that the individual identifies with the positive attributes and results of the group, he or she gets a positive self-evaluation that is incorporated into his or her personal identity.

A meta-analysis of 78 studies of social loafing conducted by Karau and Williams in 1993 confirmed that social loafing has a clear and widespread effect (it is present in all of the studies considered even though the studies are quite different in themselves) and that it increases with the group size. The results also found patterns that support the rationale and implications of CEM. On this basis, social loafing is less probable when individuals work in small groups rather than large groups, when they perform interesting tasks rather than irrelevant tasks, when they work with people they respect (as friends) rather than with strangers, when they perceive their contributions as essential to the final product rather than as redundant, and when they expect their companions to perform poorly rather than skillfully.

### 3.7. Other Integrative Models: Instrumentality, Value, and Equity

In addition to CEM, other comprehensive theoretical models that guide current research have been proposed.

For example, in 2003, Rutte included in his model the concept of equity as a mediating mechanism in addition to instrumentality and the value of the result. This corrected and refined some of the assumptions made in CEM. When individuals perform collective tasks, they judge their own contributions to the collective performance on the basis of the inputs made by other group members and they make an effort to bring about a fair exchange (i.e., balanced input–output relationship). The risk of free riding is always present in groups, although people dislike the thought because such behavior is unjustifiable. Consequently, they make less effort to avoid being suckered, and this can trigger a contagious spiral of demotivation among all members of the group. This integrative explanation combines two approaches derived from previous theories: the rational approach, from expectation–value theory, and the moral approach, from equity theory.

#### 4. IMPLICATIONS OF SOCIAL LOAFING

Karau and Williams's meta-analysis revealed the generalization of social loafing that makes it an issue of great concern because many of the conditions in the studies reviewed are present in most settings in which people work collectively, both in business and in public institutions. In different surveys carried out in these settings, workers themselves identified social loafing as one of the main problems of teamwork. However, although everybody acknowledged the phenomenon, nobody personally admitted to loafing.

The social loafing effect is especially problematic in educational settings due to the special emphasis placed on training activities based on collaborative groups. Current studies of the phenomenon pay special attention to new ways of group working such as virtual work teams and groups mediated by information technologies.

There are a number of observations that might be made about the general effect identified in the meta-analysis. With regard to the studies analyzed, nearly all were laboratory experiments carried out with ad hoc groups created out of context and having neither a past nor a future. When working with more realistic groups, different results are obtained. For example, in 1991, Erez and Somech analyzed groups of people who had known each other for at least the previous 6 months working under differing conditions of

realism (with or without specific goals, communication, reinforcement, and collective values). In this case, social loafing was the exception rather than the rule, occurring in only 1 of the 16 experimental situations.

The results of the meta-analysis contradicted the literature of organizational psychology in support of work teams. Currently, for example, major changes are being made to organizational structure to replace the organization of work around individual jobs with specifically designed team-based organizations, as described by Mohrman and colleagues in 1995. Only teams that integrate a variety of skills and bring together sufficient knowledge, expertise, and experience are capable of providing agile and flexible responses to contemporary organizations' demand for competitiveness and innovation. Groups are essential if organizations are to be made more competitive (e.g., by applying knowledge management strategies) because they are much more efficient learning systems (i.e., they learn more and better) than are isolated individuals.

Compared to individual work and competitive relationships, cooperation substantially improves productivity and strengthens interpersonal relationships. Cooperation involves understanding that individual success and group success are intertwined and that assistance, support, and integration actions will contribute to the team's success. Cooperation is the essence of work teams and provides the basic justification for the creation of links between the individual and the organization and between the organization and its setting. In 2003, these concepts were discussed in a handbook edited by West and colleagues.

Finally, there is a strong tradition of research in relation to group motivational gains, as described by Stroebe and colleagues in 1996. These gains arise for a number of reasons and are generated by the distribution of the relative strengths of the group that provide incentives that are not present in individual situations. These incentives are related to social compensation (increased motivation in the most skilled or strongest group member, who is aware that some less capable or weaker members will not be able to contribute to the group and is ready to make up for their weaknesses), interpersonal competition (based on the motive to be better than others, and should be evoked when it is possible to invert the existing rank order), and the Köhler effect (increased motivation in the weakest group members, who try to match their performance to that of a stronger companion).

Therefore, the relationship between individual productivity and group productivity presents evidence of both losses and gains. The losses are perhaps better known and, in any case, are more marked. However, their effect is limited to certain situations.

## 5. PREVENTION AND CONTROL OF SOCIAL LOAFING

The social loafing effect is not inevitable. In fact, there are a number of well-known factors that can prevent and control social loafing and can also improve motivation in groups. In general, the integrative models and research results show that it is necessary to promote situations that develop the expectation that effort is necessary to obtain results and increase their value and that such endeavors are guided by equitable criteria. Some of these factors are more directly related to the tasks that individuals must perform, others are related to group functioning, and still others depend on organizations. Nevertheless, it is hard to separate these factors in practice.

### 5.1. Task-Related Factors

Some procedures call for adequate design or reappraisal of the tasks assigned to the group members. Thus, boosting instrumentality and reinforcement requires designing tasks on the basis of the factors shown in [Table I](#) (e.g., meaningful tasks, intrinsically interesting or attractive tasks). Some of these strategies are in fact applied by the managers of certain kinds of teams such as elite sports teams, whose players are individually filmed and evaluated to identify their personal performance.

### 5.2. Group-Related Factors

Other factors are related to the composition and processes of the group as well as to its relationships with other groups. [Table I](#) reflects a number of significant group characteristics that help to prevent social loafing.

Certain competition- or high goal achievement-oriented groups (e.g., sports teams, elite combat units) promote specific strategies to achieve additional motivation, for example, establishing rules oriented toward maximum effort, encouraging strong feelings of group loyalty, and increasing arousal.

Team training and development programs are fundamental to promoting the majority of these factors. For example, training in processes and group skills,

whether behavioral or cognitive (e.g., communication, coordination of effort, shared mental models, transactive memory systems) helps to develop the perception that individual efforts are inextricably linked to group performance, to raise awareness of the characteristics of group members, and to increase convergence within the group with regard to both tasks and goals.

The strategies usually employed in team development are likewise essential to the prevention of social loafing inasmuch as they allow for improving individuals' skills in four main areas that were broadly discussed by Salas and colleagues in 1999: (a) setting and achieving goals; (b) developing communication, support, and trust; (c) identifying problems and generating, applying, and evaluating solutions; and (d) improving comprehension of the responsibilities and obligations of other members' roles. Therefore, the strategies used in team development are very useful for improving members' motivation to the extent that personal responsibility is enhanced, task relevance is recognized, commitment among members and to the group's shared goals is promoted, and the group's social identity is developed.

The role played by the leaders or managers is crucial to the extent that they are capable of forming the group, designing the physical and organizational setting in which it operates, and influencing members' perceptions, expectations, and behaviors. For example, the leader can make the group see that members' skills are necessarily mixed and can develop the expectation that every individual's input is essential. In addition to directly influencing members' behavior through supervision, direction, and control, the leader can appeal to their values and ideals to achieve organizational goals and promote change and innovation (transformational leadership).

### 5.3. Factors Related to the Organization

Other factors are related to the organizational setting in which groups and individuals are embedded. The design of the task, the organizational structure (communication, power, and decision structure), human resources policies and practices, and the organizational culture may create specific patterns that directly prevent social loafing and lend coherence and support to the strategies referred to earlier ([Table I](#)).

With regard to incentives, organizations can establish a system of individual and collective external reinforcement and can contribute to the development of internal reinforcement (collective or individual).

**TABLE I**  
**Factors Reducing or Eliminating the Social Loafing Effect**

<i>Level of analysis</i>	<i>Factors and strategies</i>
Tasks	<ul style="list-style-type: none"> <li>• Task evaluation (by supervisors, colleagues or the individuals themselves) allows the reinforcement of actual behaviors.</li> <li>• Different or unique tasks ensure that group members perceive that their inputs to the collective product are not redundant but rather necessary and complementary.</li> <li>• Tasks involve incentives to contribute.</li> <li>• Tasks reduce the costs of contributing.</li> <li>• Tasks are intrinsically interesting or attractive.</li> <li>• Tasks produce personal involvement.</li> <li>• Tasks are meaningful.</li> <li>• The collective results achieved should be considered valid and important for the individuals and their reference groups.</li> </ul>
Group	<ul style="list-style-type: none"> <li>• The group should be small.</li> <li>• The group should be made up of individuals who share a relationship based on respect and trust.</li> <li>• Group members should be aware of each other's skills as well as of each other's difficulties with some tasks or the inability to carry them out.</li> <li>• The group should have high cohesion.</li> <li>• The group should clearly set performance goals (in addition to or instead of individual goals), and these goals should be highly valued.</li> <li>• The group should receive feedback and reinforcement (in addition to or instead of individual feedback).</li> <li>• Group members should consider group results to be valid and important.</li> <li>• There should be group norms to regulate the behavior patterns required of members to achieve goals.</li> <li>• Group members should share the same view of the situation and mental models.</li> <li>• Group members should share a perception of high collective efficacy (i.e., team potency).</li> <li>• Group members should value the group and identify with it, establishing a positive identity within the group and with regard to other collectives.</li> </ul>
Organization	<ul style="list-style-type: none"> <li>• Team training and team development programs should be applied.</li> <li>• Select group members in accordance with their skills and the tasks required.</li> <li>• Set goals that are both challenging and realistic.</li> <li>• Define standards or comparative criteria.</li> <li>• Provide nonevaluative conditions to test and use skills.</li> <li>• Establish conditions to use skills and perform tasks.</li> <li>• Provide regular feedback on task performance.</li> <li>• Provide the necessary resources and support actions.</li> <li>• Establish an incentive system.</li> <li>• Create a setting that leads to the development and strengthening of self-efficacy and self-esteem.</li> </ul>

Shepperd reviewed these systems in a 1993 article. The most commonly applied method is to establish external reinforcements that are both collective (material or social) and individual. The organization can foster the development of individual and collective internal reinforcement. To enhance internal individual reinforcement, the organization can not only design interesting and meaningful tasks but also set specific performance goals to allow individuals to evaluate and compare their behavior on the basis of objective

criteria. By creating teams, acknowledging their autonomy, empowering their members, and using cultural artifacts (e.g., symbols, myths), the organization can also contribute to the development of internal collective reinforcements so that group members identify with the teams and the organization, feel proud to belong to them and feel responsible to contribute to them, and feel capable of coping with shared goals and of accepting collective successes as personal triumphs.

Although the use of collective reinforcements is very important, the conditions under which they are effective are unclear, according to a 1998 review by DeMatteo and colleagues. This is because they depend both on the characteristics of the reinforcement itself (size, frequency, and allocation procedures) and on the characteristics of the individuals, groups, and organizations concerned.

According to experts in the field, both the reinforcement system and the mechanisms used to set goals and provide feedback require the organization to guarantee a work setting that ensures a good match between team members' and teams' motivation and the organization's own goals. Similarly, the organization should avoid any dissociation or conflict between motivational factors.

To conclude this review of the factors that prevent social loafing, it is necessary to at least mention individual variables related to personality traits that can give rise to major differences among the members of a team. Among these variables, the need to belong and feel affiliated, a collectivist orientation, and the acceptance of Protestant ethics have been identified as key to the prevention of social loafing.

The prevention of social loafing and the enhancement of team members' motivation are complex topics that can hardly be explored fully in this article. However, it can be noted that they depend on a combination of variables that, in turn, present their own peculiarities, including the following:

- The tasks should not be confined to simple additive operations; rather, they should demand the synchronized combination of inputs from a variety of specialists.
- The teams should not be fixed, durable, and real; rather, they should be temporary, requiring members' partial and multiple affiliations, and also should also present some degree of virtuality.
- Goals and tasks should not be set in isolation from one another; rather, they should present varying degrees of interdependence.
- Decisions should be made not only on the basis of rational criteria but also taking into account moral and emotional factors.
- The feeling of belonging to a group or an organization and involvement are not always based on economic transactions or regulated by formal contracts; rather, they can also depend on ideals and values as well as on some kind of commitment.
- Work relationships are not confined to bonds between the individual and other group members but also include associations with other groups and organizations.

These variables represent the main topics of current research into work teams in the organizational setting. Results from this research will provide a better understanding of the keys to the motivation losses and gains that arise when people work in teams. This could allow for the achievement of maximum performance situations, where the individual's inputs are recognized and combined efficiently with the contributions made by other members of the team and organization while avoiding the temptation to overvalue the individual's contributions but without annihilating them in favor of a tyranny of collective values in this age of "groupism."

### See Also the Following Articles

Groups, Productivity within ■ Groupthink ■ Intergroup Relations and Culture

### Further Reading

- Baron, R. S., Kerr, N., & Miller, N. (1992). *Group process, group decision, group action*. Buckingham, UK: Open University Press.
- DeMatteo, J. S., Eby, L. T., & Sundstrom, E. (1998). Team-based rewards: Current empirical evidence and directions for future research. In B. M. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 20, pp. 141–183). Greenwich, CT: JAI.
- Erez, M., & Somech, A. (1996). Is group productivity loss the rule or the exception? Effects of culture and group-based motivation. *Academy of Management Journal*, 39, 1513–1537.
- Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology*, 10, 371–384.
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65, 681–706.
- Karau, S. J., & Williams, K. D. (2001). Understanding individual motivation in groups: The collective effort model. In M. E. Turner (Ed.), *Groups at work: Theory and research* (pp. 113–141). Mahwah, NJ: Lawrence Erlbaum.
- Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Industrial and organizational psychology* (Vol. 12, pp. 333–375). New York: John Wiley.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37, 822–832.
- Locke, E. A., Tirnauer, D., Roberson, Q., Goldman, B., Latham, M. E., & Weldon, E. (2001). The importance of

- the individual in an age of groupism. In M. E. Turner (Ed.), *Groups at work: Theory and research* (pp. 501–528). Mahwah, NJ: Lawrence Erlbaum.
- Mohrman, S. A., Cohen, S. G., & Mohrman, A. M. (1995). *Designing team-based organizations: New forms for knowledge work*. San Francisco: Jossey-Bass.
- Rutte, C. G. (2003). Social loafing in teams. In M. A. West, D. Tjosvold, & K. G. Smith (Eds.), *International handbook of organizational teamwork and cooperative working* (pp. 361–378). Chichester, UK: Wiley.
- Salas, E., Rozell, D., Driskell, J. D., & Mullen, B. (1999). The effects of team building on performance: An integration. *Small Group Research*, 30, 309–329.
- Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin*, 113, 67–81.
- Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.
- Stroebe, W., Diehl, M., & Abakoumkin, G. (1996). Social compensation and the Köhler effect: Toward a theoretical explanation of motivation gains in group productivity. In E. Witte, & J. H. Davis (Eds.), *Understanding group behavior: Small group processes and interpersonal relations* (Vol. 2, pp. 37–65). Mahwah, NJ: Lawrence Erlbaum.
- Weaver, J. L., Bowers, C. A., Salas, E., & Cannon-Bowers, J. A. (1997). Motivation in groups. In M. M. Beyerlein, & D. A. Johnson (Eds.), *Advances in interdisciplinary studies of work teams* (Vol. 4, pp. 167–191). Greenwich, CT: JAI.
- West, M. A., Tjosvold, D., & Smith, K. G. (Eds.). (2003). *International handbook of organizational teamwork and cooperative working*. Chichester, UK: Wiley.



# Social Networks

Mieneke W. H. Weenig

Leiden University, Leiden, The Netherlands

1. Introduction
  2. Social Networks
  3. Major Research Findings
  4. Conclusions
- Further Reading

## GLOSSARY

**bridge** A person who links two or more cliques in the network from his or her position as a member of one of the cliques.

**centralization** The extent to which ties in a network are directed to a central individual in the network.

**cliques** Subsystems in the larger network whose elements interact with each other relatively more frequently than with other members of the network.

**connectedness** The extent to which individuals in a network are linked by direct or indirect ties.

**density** Ratio of the number of existing ties to the number of possible ties.

**homophily** The degree to which two interacting individuals are similar in certain attributes such as beliefs, norms and values, and socioeconomic status.

**interlocking network** A network in which all members communicate with each other.

**knots (or stars)** Individuals who have relatively many links with other members of the network or clique; they are central persons in the network like spiders in a web.

**liaison** A person who links two or more cliques in the network but who is not a member of any clique.

**multiplexity** The degree to which a network link is used for various topics.

**network structure** The pattern of communication links in a system.

**sociometry** A means of obtaining quantitative data about who communicates with whom in a system.

**structural equivalence** The degree to which two individuals are linked to the same set of other persons but not necessarily to each other.

**tie (or link)** A communication relationship between two units (usually individuals) in a system.

**tie strength** A combination of the amount of time invested, the emotional intensity and intimacy, and the reciprocal services that characterize a tie.

A social (or communication) network is an aggregation of individuals who are tied to each other by patterned flows of information. Ties can vary in strength, depending on the frequency of use and on the degree to which the ties are used for multiple message contents. Weak ties are especially important for the diffusion of information, whereas strong ties are more important with respect to the flow of influence.

## 1. INTRODUCTION

Humans are social beings, and the social context plays an important role in what they do (behavior) and what they think (attitudes and beliefs). For instance, when an individual wants to buy a new car, he or she would probably base the ultimate decision on a mixture of three kinds of information: (a) the individual's own

experiences with cars; (b) information obtained from brochures, television, and magazines; and (c) information obtained from other persons such as experiences or advice from personal friends and acquaintances and advice from experts (e.g., the car salesperson). In addition, the individual would observe what cars his or her friends and neighbors have bought themselves and probably would not buy a car that he or she expect the social environment to reject.

In other words, individuals base their purchase decisions at least partly on information, expertise, and expectations from others that is communicated mainly through interpersonal networks. Thus, it is very important to know something about the social context of a target group and how it affects people to develop and implement more effective promotion (or “persuasion”) campaigns.

## 2. SOCIAL NETWORKS

The social context of a person, group, or community can be conceived as an aggregation of individuals. Some of them communicate with each other, whereas others do not. The pattern of all such communication ties is called a communication network or social network. Basically, such a network indicates who communicates with whom. Together, these “communication links” constitute a network of direct and indirect ties between people. The communication link or tie is the basic unit of each communication network. It is the dyadic link of interpersonal communication between two individuals.

Ties may vary in strength, depending on the frequency of use and on the degree to which a tie is used for multiple message contents (also called “multiplexity”). In 1973, Granovetter defined the strength of ties as “a (probably linear) combination of the amount of time invested, the emotional intensity and intimacy (mutual confiding), and the reciprocal services which characterize the tie.” Thus, close relatives and friendship relations can be regarded as strong ties.

The formation of a strong tie between two network members is more likely in case of homogeneity with respect to norms, values, important beliefs, socioeconomic status, and demographic characteristics. This is referred to as the “homophily principle.” In general, the more homogenous two individuals are, the higher the probability of a strong tie between them. However, in the immediate geographic environment, spatial distance appears to be a decisive factor. Therefore, in neighborhoods, the probability of a rather strong (at

least in terms of high frequency) tie between next-door neighbors is relatively high and less dependent on similarity with respect to social and demographic characteristics.

Apart from the strength of ties, the most frequently used characteristics of ties in social network research are the stability (duration over time) and direction of ties (unidirectional vs bidirectional). In addition, several structural characteristics of social networks have been identified, including the existence of cliques, knots, bridges, and liaisons. Finally, various network indexes have been developed for network analyses, including the degree of density, connectedness, and centralization.

## 3. MAJOR RESEARCH FINDINGS

### 3.1. Efficiency of Information Diffusion in Various Types of Networks

Some of the first studies on communication networks were performed during the 1950s by Massachusetts Institute of Technology researchers who were mainly interested in the relationship between network structure and efficiency of information diffusion. Basically, they used an experimental design in which participants could communicate with one another through various kinds of network structures, including a wheel, a chain, a Y, a circle, and a “comcon” (completely connected). Of these network forms, the wheel (where all information goes to one central person) was the most centralized network form, whereas the comcon (where all individuals can communicate with one another) was the least centralized network form. Results suggested that a centralized network is more efficient than a decentralized network, but later studies qualified this rule to apply mostly to relatively simple information and routine tasks.

The information diffusion process is also affected by the strength of ties. In his seminal article, Granovetter argued that weak ties, although used less frequently than strong ties, may nevertheless be more valuable for the diffusion of information than are strong ties because weak ties are more likely than strong ties to serve as bridges between subgroups (or “cliques”) in the larger network. Thus, bridges to other cliques prevent information from remaining inside clique boundaries and bring new information to a clique, a notion that has received support from several investigations. However, within a clique, strong ties seem to be more important for information diffusion.

### 3.2. Network Influences on the Diffusion of Innovations

Social network influences on people's attitudes and behavior have been investigated most extensively in the areas of the diffusion of innovations, marketing, and organizational communication. Research on the diffusion of innovations has addressed the question of which factors determine people's decisions to adopt innovations and how these adoption decisions spread in the social system. Rogers defined an innovation as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption." Innovation diffusion is the process by which the adoption of an innovation spreads over time in a social system through communication channels among its members. The adoption rate is the relative speed with which an innovation is adopted by all members of a social system.

Rogers identified "opinion leaders" as crucial members of a social system. An opinion leader is defined as an individual who "is able to influence other individuals' attitudes or overt behavior informally in a desired way with relative frequency." According to Rogers, major characteristics that distinguish opinion leaders from their followers include greater exposure to mass media and a more "cosmopolitan" orientation, more personal ties within the community as well as with people beyond the group's boundaries (including change agents), and a higher social status in comparison with their followers. Computer simulations have shown that diffusion occurs up to three times faster when it is initiated by opinion leaders than when it is initiated by random network members.

One of the most classic innovation diffusion studies is Coleman and colleagues' 1957 study on the diffusion of a new medical drug among a group of doctors. In this study, socially integrated doctors, described as doctors who received three or more sociometric choices as "friends" by their colleagues and thus were considered to be "opinion leaders" of the network, were observed to have adopted a new drug much earlier than did less integrated doctors (who were mentioned as friends less frequently). The moment that nearly all of the opinion leaders had adopted the new drug, the diffusion rate of the opinion leaders' followers accelerated rapidly. This was interpreted as the result of interpersonal communication about the new drug. However, more recently, Burt reanalyzed Coleman and colleagues' data and demonstrated that the results did not reflect interpersonal conversations among colleagues (so-called "contagion by cohesion").

Instead, Burt found that the data reflected a (nonverbal) process of social comparison and competition among doctors who occupied structural equivalent positions in the communication network of the doctors (i.e., had similar patterns of ties within the network), irrespective of whether they had direct ties with each other (so-called "contagion by structural equivalence"). Based on these data and on data from business and politics, Burt showed that the strength of relationships (i.e., cohesion) increases the adoption rate only in the case of weakly equivalent individuals. According to Burt, contagion occurs regardless of cohesion for strongly equivalent people, whereas contagion does not take place at all for nonequivalent people.

Although contagion might not necessarily require the existence of direct ties but rather might require an equivalent position in the network, some studies indicate that contagion does take place at least partly through the influence of direct ties. An example is a 1981 study by Rogers and Kincaid on the diffusion of family planning measures in Korea. It showed that individuals were more likely to adopt the innovation if a larger proportion of the persons with whom they most often talked about family planning consisted of individuals who had already adopted the innovation.

### 3.3. The Impact of Word-of-Mouth Communication

Other indications of the influence of direct communication links on adoption decisions come from consumer and marketing studies on the occurrence and impact of "word-of-mouth" (WOM) recommendations. A large number of studies on the antecedents of purchase decisions have shown that substantive percentages (sometimes even large majorities) of buyers mention WOM recommendations as important or major sources of information or influence concerning the purchase of a specific product or service. In fact, the frequency of occurrence and the apparent influence of WOM communication have even encouraged several manufacturers to implement promotion strategies of their products that are based nearly exclusively on stimulating WOM recommendations within their target groups by means of strategically providing free samples of their products (e.g., Häagen-Dazs ice cream). However, in spite of the acknowledged importance of WOM recommendations for purchase decisions by manufacturers and scientists alike, surprisingly little research has been performed on the

dynamics and antecedents of WOM recommendations. One of the most prominent exceptions is Brown and Reingen's 1987 research. These authors found that weak ties served as information bridges through which WOM information traveled from one subgroup in the network to another, whereas strong and homophilous ties were more activated for the flow of referral information and also were perceived as being more influential than weak ties.

Research on negative WOM communication is even scarcer, but findings suggest that, in general, negative WOM communication has a stronger influence on consumers' evaluations of brands and products than does positive WOM communication. This is consistent with the so-called "negativity effect" that is often found in the impression formation and decision-making literature. It means that negative information is given greater weight in judgments than is positive information of equal extremity. The most accepted explanation for the negativity effect is that negative information is generally perceived as more diagnostic than positive information for categorizing target objects into evaluative categories.

In 2002, Ahluwalia suggested that the influence of negative WOM communication on product evaluations depends on the causal attributions that recipients make after hearing negative WOM messages. For instance, if a recipient infers from a message that only one person has had a negative experience with a particular brand or product, whereas most people have had positive experiences with that product (low consensus), or when a recipient perceives the communicator to be a person who has negative opinions about many products or brands (low distinctiveness), the negative message is attributed to (i.e., is perceived as being caused by) the communicator rather than to the product to which the negative WOM message refers. In that case, product evaluations might even increase due to negative WOM communication.

Other recent research performed by Lacznik and colleagues suggested that the negativity effect applies mostly to consumers who are unfamiliar with a specific product or brand but who are motivated to evaluate that product as accurately as possible with an "open mind" to all kinds of information. On the other hand, if consumers are committed to a particular product or brand, they usually want to defend their choice. In that case, a positivity effect is more likely than a negativity effect, meaning that more weight is given to positive information that is consistent with consumers' existing (positive) opinion concerning the product than to negative (i.e., inconsistent) information.

In sum, the stronger influence of negative WOM communication on product evaluations seems to be restricted to situations where recipients are relatively unfamiliar with the product to which the negative message refers, attribute the negative information to the product and not to the communicator, and are motivated to make an accurate evaluation of the product.

WOM communication generally seems to occur more frequently among strong ties than among weak ties and with more impact on attitudes and behavior. However, whenever WOM communication occurs among weak ties, it seems to be evaluated more often as new and surprising information.

Apart from direct communication links, indirect communication links also may play an important role in the diffusion process. Several rumor studies have shown that secondhand information (i.e., information transmitted by indirect ties) is often more polarized (positively or negatively) or exaggerated than is firsthand information. Thus, secondhand information about an individual's positive or negative experiences with an innovation may be more "extreme" than the original information.

### **3.4. Network Influences on the Diffusion of Information Versus Adoption Decisions**

Before an individual can make a decision on whether to buy a product or adopt an innovation, he or she should first be informed or made aware of its existence and also learn something about its features or characteristics. However, in many studies, no distinction is made between the process of information diffusion and social influence on adoption decisions. Such a distinction was made in a study by Weenig and Midden in 1991. They investigated the influence of a target group's existing social network on the information diffusion process and effects (i.e., positive adoption decisions) of several small-scale communication programs that aimed to promote the adoption of energy conservation measures. In each case, the program was developed and implemented by a project group that consisted of one or two professional change agents and several volunteers from the target group. In all cases, program awareness was positively related to the number of weak ties that residents had within the neighborhood and was unrelated to the number of strong ties that they had within the neighborhood, thereby supporting Granovetter's hypothesis on the strength of weak ties. Program awareness and attention to program activities

were also higher among residents who were personally acquainted with at least one of the volunteers from the project group than among residents who had no direct ties to the project group. The presence of direct ties to the project group did not affect program effects (i.e., adoption decisions). Instead, adoption decisions were positively related to advice received from strong ties (i.e., friends or family members) and were unrelated to advice received from weak ties. This means that the strength of ties is important with respect to influence on adoption decisions. Brown and Reingen found similar results concerning WOM referral behavior (described earlier). Thus, weak ties are important for the diffusion of information, whereas strong ties are more important with respect to the flow of influence.

In one of the target groups, however, the adoption of cavity wall insulation appeared to be negatively related to the number of weak ties in the neighborhood. It appeared that this negative relationship reflected the awareness of a negative rumor—contending that cavity wall insulation would lead to moldy walls—that had been disseminated among residents. The negative relation between adoption of cavity wall insulation and number of weak ties means that awareness of this negative rumor occurred mostly through weak ties. This awareness resulted (both directly and indirectly) in fewer adoptions of cavity wall insulation—directly if the rumor was heard from more than one weak tie (consensus information) and indirectly if the rumor was heard from only one weak tie (when it triggered asking additional advice from strong ties).

#### 4. CONCLUSIONS

From the research described in the previous paragraph, three important conclusions can be drawn. First, the quantity of existing communication ties in the social network of a target group affects the amount and speed of diffusion of information. That is, the larger the degree of connectedness among target group members, the more alternative routes will be available for the diffusion of information. The flow of information will usually be fastest among strong ties because they are (by definition) used more often than are weak ties. However, because strong ties tend to cluster into interlocking subgroups (cliques) in the larger network, new information from outside a clique is likely to enter a subgroup through a bridge or liaison, and such ties tend to be weak.

Second, the strength (or quality) of ties determines the potential of social influence that may occur within

the target group. That is, people appear to be influenced most by their friends and close relatives or by other “near peers.” The potential of social influence from friends and close relatives has been stressed and demonstrated in several social psychological research traditions, including Bandura’s social learning theory (i.e., attractive persons are the most influential models), persuasion theory (i.e., friends are usually more persuasive than strangers), Festinger’s social comparison theory (i.e., similar persons serve as reference persons more often than do dissimilar persons), and conformity theory (i.e., more compliance to group norms occurs in cohesive or tightly knit groups). In fact, it might well be that the influence that strong ties exert is a mixture of all these kinds of social influence.

Third, opinion leaders, so far as they can be identified in a network or subgroup, may accelerate the speed of information diffusion and adoption decisions to a considerable extent, mainly as a result of their relative centrality in the network in combination with their ties with people outside the (sub)group’s network. Unfortunately, such opinion leaders do not exist in all communities. However, several studies have shown that opinion leaders can be “created,” to some extent, by means of training volunteers from the target group to become experts on the relevant topic. Such volunteers are able to reach and stimulate a large proportion of the target group provided they have many personal ties with target group members and have at least some interest in the topic. This means that, apart from at least some interest in the relevant topic, the number of ties within the target group should be an important selection criterion with respect to the recruitment of volunteers. These volunteers must subsequently be trained to become paraprofessional change agents. The duration and intensity of this training depend on the difficulty and complexity of the innovation and on the educational background of the volunteers, among other factors. For simple products, it might be enough to train the volunteers in how to provide free samples of the product to strategic persons (e.g., knots, bridges) in the network, whereas for more complex innovations, elaborate and lengthy training on various aspects (e.g., knowledge about the innovations, group dynamics, and communication processes) might be necessary.

This implies that it is crucial to have at least some knowledge of the existing social network of the target group. No doubt, the most comprehensive way in which to obtain this is through gathering sociometric data on the existence of ties, their strength, and their structure. However, even in their most simple form, such analyses

are complex, time-consuming, and costly. Moreover, for practitioners, they often result in unnecessarily detailed data, for in many cases change agents will not be interested in detailed descriptions of who communicates how frequently with whom. Instead, change agents may be more interested in the quantity and quality of existing ties at an aggregated level and in whether opinion leaders and bridges or liaisons can be identified. Local authorities, such as schoolteachers, church leaders, police officers, and community workers, can often provide such information. In that manner, essential information about network characteristics can be obtained quickly and easily and can be used subsequently to implement more efficient promotion strategies.

### See Also the Following Articles

Groups, Productivity within ■ Groupthink ■ Intergroup Relations and Culture

### Further Reading

- Ahluwalia, R. (2002). How prevalent is the negativity effect in consumer environments? *Journal of Consumer Research*, 29, 270–279.
- Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14, 350–362.
- Burt, R. S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, 92, 1287–1335.
- Burt, R. S. (1999). The social capital of opinion leaders. *Annals of the American Academy of Political and Social Science*, 566, 37–54.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Laczniak, R. N., DeCarlo, T. E., & Ramaswami, S. N. (2001). Consumers' responses to negative word-of-mouth communication: An attribution theory perspective. *Journal of Consumer Psychology*, 11, 57–73.
- Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.
- Rogers, E. M., & Kincaid, D. L. (1981). *Communication networks: Toward a new paradigm for research*. New York: Free Press.
- Shaw, M. E. (1964). Communication networks. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1, pp. 111–147). New York: Academic Press.
- Weenig, W. H., & Midden, C. J. H. (1991). Communication network influences on information diffusion and persuasion. *Journal of Personality and Social Psychology*, 61, 734–742.



# Social Skills Training

Stephen N. Elliott and Sylvia C. Lang

University of Wisconsin, Madison, Wisconsin, USA

1. Introduction
  2. Identification of Children in Need of Social Skills Training
  3. Social Skills Training Procedures
  4. Effectiveness of Social Skills Interventions
  5. Selection of Interventions for Students with Different Problems
- Further Reading

## GLOSSARY

**behavioral rehearsal** A process in which the intervention target rehearses a specific social behavior and receives informational feedback on the rehearsal.

**coaching** A direct verbal instructional technique in which a “coach” (intervention agent) provides the intervention target with a standard for a specific behavior, the opportunity to rehearse that behavior, and informational feedback on the rehearsal.

**discrimination training** The process of training an individual to differentiate between situations and/or settings in which specific behaviors or actions would or would not be appropriate.

**modeling** A process in which the intervention target observes the social behavior(s) of a model for purposes of the intervention target acquiring new, socially appropriate behaviors.

**social skills acquisition problem** Characterizes children who have not acquired the necessary social skills to interact appropriately with others, or children who have not acquired a critical step in the performance of a given skill.

**social skills performance problem** Characterizes children who have appropriate social skills in their behavior repertoires but do not perform the behavior at acceptable levels and/or at appropriate times or settings.

Social skills interventions emphasize the acquisition and performance of prosocial behaviors and typically utilize nonaversive methods to teach these prosocial behaviors. Some social skills interventions, however, must concurrently focus on the reduction of interfering problem behaviors. Thus, social skills training has four primary objectives: promoting social skills acquisition, enhancing social skills performance, reducing or removing interfering problem behaviors, and facilitating the generalization and maintenance of social skills. Common social skills training tactics can be characterized as therapist directed, therapist and peer directed, or peer directed. These procedures can be further categorized into three theoretical approaches: operant, social learning, and cognitive-behavioral. Research indicates that social skills training overall is a moderately effective procedure for improving the social functioning of children age 3 to adults age 21.

## 1. INTRODUCTION

Behaviors such as sharing, helping, initiating relationships, requesting help from others, and ignoring teasing or aggressive reactions from others are examples of socially appropriate behaviors that enable children young and old to successfully function in social settings. Developing these types of social skills is one of the most important accomplishments of childhood and can lead to successful social relationships and academic functioning.

Social skills may be defined as socially acceptable learned behaviors that enable a person to interact with others in ways that elicit positive responses and to avoid negative responses. When an individual enacts social skills successfully, the results generally are positive perceptions of that individual's social competence. The acronym of CARES was developed by Gresham and Elliott in 1990 to characterize major clusters of social skills frequently targeted for intervention. These skill clusters are cooperation, assertion, responsibility, empathy, and self-control. Guevremont developed an alternative, but equally appealing, situational approach to characterizing clusters of social skills. Specifically, his skill clusters are social entry skills, conversational skills, conflict-resolution skills, problem-solving skills, and anger-control skills. Regardless of the terminology or approach used to characterize social skills, the behaviors of interest to most individuals conducting social skills training are those observable nonverbal and verbal interpersonal skills that maximize social engagement and social reinforcement.

The following are key assumptions to the assessment and treatment of social skills: (i) Social skills are primarily acquired through learning that involves observation, modeling, rehearsal, and feedback; (ii) they include specific, discrete verbal and nonverbal behaviors; (iii) they require specific behaviors that are effective in initiating, maintaining, and terminating interpersonal interactions and relationships; (iv) they are interactive by nature and entail effective and appropriate behavioral performances; and (v) they are situationally specific behaviors and are influenced by the characteristics, demands, and expectations operating in specific environments. Collectively, these five assumptions stress the multidimensional (verbal–nonverbal and initiating–responding) interactive, situation-specific nature of social skills. As such, effective interventions need to address (i) target behaviors involving both verbal and nonverbal communications used to initiate or respond to others and (ii) the ability to discriminate when to exhibit certain social skills given changing social situations.

## **2. IDENTIFICATION OF CHILDREN IN NEED OF SOCIAL SKILLS TRAINING**

A number of methods, including rating scales, checklists, and sociometric nomination techniques, have been designed to identify children at risk for behavior

problems. In general, social skills assessments have one of two purposes: identification/classification or intervention/program planning. A standard battery of tests or methods for assessing social skills does not exist. To increase the likelihood of accurate identification/classification decisions, it is recommended that one use direct observations of the target child and nontarget peers in multiple settings; behavioral interviews with the referral source and possibly the target child; rating scale data, preferably norm-referenced, from both a social skills scale and a problem behavior scale completed by more than one source; and sociometric data from the target child's classmates. Regarding intervention decisions, data contributing to a functional analysis of important social behaviors are imperative. These types of data usually result from multiple direct observations across settings, behavioral role-plays with the target child, and teacher and parent/guardian ratings of socially valid molecular behaviors.

Children identified for social skills intervention historically have been either individuals who have been rejected by their peers because they behave aggressively toward others or students with significant cognitive disabilities. During the past two decades, a large number of students with learning disabilities have also been targeted for social skills interventions. A number of researchers have documented that mildly disabled students who are classified as behavior disordered, learned disabled, or cognitively impaired frequently exhibit a significant number of social skills deficits and could benefit from individual or group social skills interventions. Social skills have also been documented to be an important part of academic success for nondisabled students, and thus the number of classwide school skills programs has also increased substantially during the past decade. The intervention strategies used for individual students are directly applicable, with only minor changes, to small groups of students, a classroom, or an entire school.

## **3. SOCIAL SKILLS TRAINING PROCEDURES**

Social skills training procedures focus on positive behaviors and often use nonaversive methods (e.g., modeling, coaching, and reinforcement) to improve children's behavior. Therefore, use of these methods may enhance treatment acceptability and integrity and they can easily be built into the existing structure of a classroom or home environment. Social skills

interventions can be used with individuals or groups of students, and because they primarily concern increasing prosocial behaviors, all students can participate and potentially benefit from the interventions.

A large number of intervention procedures have been identified as potentially effective for social skills training with children age 3 to adults age 21. These procedures can be classified into three approaches that highlight common treatment features and assumptions about how social behavior is learned: operant, social learning, and cognitive-behavioral. In practice, many researchers and practitioners have used procedures that represent combinations of these basic approaches. However, these three approaches to intervention are used in this article to describe the basic procedures and to organize a review of their effectiveness.

### **3.1. Operant Intervention Procedures**

Operant intervention procedures focus on discrete, observable behavior and the antecedent and consequent events that maintain that behavior. Behavioral control is achieved most often through the use of reinforcement and/or punishment paradigms that are contingent upon a specified behavior.

Children often fail to interact successfully because of a nonresponsive or inhospitable social environment. Here, the use of antecedent control can serve to facilitate a positive environment and elicit positive social interactions. With these procedures, a teacher or other intervention agent makes wide use of cueing and prompting. The key assumption with antecedent control procedures is that a child possesses the necessary social skills but he or she is not performing them at an acceptable rate. Two of the most frequently used antecedent strategies are peer social initiation and cooperative learning. Peer social initiation entails training similar-age peer confederates to initiate and maintain social interactions with a withdrawn or isolated child. Overall, this procedure has been effective in increasing positive social behavior in withdrawn children. It should be noted, however, that when a child's interaction rate approaches zero, peer initiations are likely to be less effective. Cooperative learning, on the other hand, provides a prosocial environment in which children work together to complete specified activities. This procedure requires children to cooperate, share, and assist each other to complete a task and, as such, represents an effective method for increasing the likelihood of positive social behaviors.

Many operant learning procedures, which involve the manipulation of antecedents and consequents, have been used to decrease interfering problem behaviors while simultaneously increasing positive social behaviors. These procedures are based on the premise that behaviors are maintained by functional relationships. In other words, reinforcement contingencies (positive or negative) perpetuate low rates of positive social interactions and high rates of negative social interactions. For this reason, it is assumed that a child can perform positive social interaction skills but does not demonstrate the skills due to lack of reinforcement. Two of the most frequently used methods to decrease inappropriate social behaviors are contingent social reinforcement and differential reinforcement. Contingent social reinforcement involves publicly reinforcing a child for socially appropriate behaviors. Although this procedure has been found to be effective, it can require an extraordinary amount of time to ensure that the reinforcement is delivered on a consistent basis. Unless reinforcement is given in a consistent manner, it is doubtful the intervention will be as effective. This type of contingent social reinforcement perhaps is best used to maintain social interaction rates established through the use of other social skills interventions. Other operant learning procedures used to modify social skills are differential reinforcement of other behavior (DRO), differential reinforcement of alternative behaviors (DRA), and differential reinforcement of low rates of responding (DRL). DRO and DRA are based on the omission of a specific behavior rather than the commission. Therefore, after a certain amount of time, reinforcement is given after any behavior except the target behavior.

DRL involves reinforcement for reduction in the frequency of a target behavior. Reinforcers may be delivered either for reduction in overall frequency of a response within a particular time period or for increased elapsed time between responses (interresponse time). DRO and DRL procedures are probably most effective, however, when used in conjunction with other social skills interventions, whereby one decreases socially inappropriate behavior while concurrently teaching or reinforcing positive social behaviors.

### **3.2. Social Learning Intervention Procedures**

Social learning procedures can be traced back to the social learning theory of Bandura and Walters

(published in 1963) and Bandura (published in 1977). From this perspective, social behavior is the result of two types of learning: observational learning and reinforced learning. Social learning theorists differentiate between the learning and the performance of a response. This distinction enables social learning theorists to advocate the process of modeling as a means to acquire new, socially appropriate behaviors. Modeling also affects previously learned responses through its disinhibitory and/or cueing effects. Therefore, children are vicariously reinforced (positively or negatively) by observing a model receiving reinforcement for a behavioral performance. In essence, observers tend to inhibit responses that they see punished in others, whereas they are likely to perform modeled behaviors that elicit desired reinforcers. Modeling has substantial empirical support for the promotion of social skills with children and youth. A powerful development in the realm of social learning has been peer-mediated interventions. These interventions are based on the premise that peers can be effective change agents for other children with social skills performance deficits. Empirically, it has been shown that peers can serve to differentially reinforce appropriate social interactions and influence the occurrence of positive social behaviors. Peer-mediated interventions may be more successful than teacher-mediated interventions because peer confederates may be better able to consistently monitor and differentially reinforce their peers. Additionally, peer mediation can be cost-effective because it minimizes teacher involvement.

### **3.3. Cognitive–Behavioral Intervention Procedures**

Cognitive–behavioral intervention procedures are a loosely bound group of procedures that focus on a child's internal regulation of his or her behavior. In particular, cognitive–behavioral approaches to social skills training emphasize a child's ability to problem solve and to self-regulate behavior. Two of the most frequently used cognitive–behavioral social skills procedures are coaching and problem solving. Coaching, unlike modeling, is a direct verbal instruction technique that involves a "coach" (usually a teacher or a psychologist, and occasionally a peer) who has knowledge about how to enact a desired behavior. Most coaching procedures involve three basic steps: (i) A student is presented with rules for, or the standards of, a specific behavior; (ii) a selected social skill(s) is

rehearsed with the coach; and (iii) the coach provides specific informational feedback during behavioral rehearsal and gives suggestions for future performances. In some instances, modeling may be included in the coaching procedure to give the child a better understanding of the topography of the social behavior and, if praise is given, reinforcement may occur. Therefore, coaching, although conceptualized as a procedure that requires the child's cognitive skills to translate instruction into desired behaviors, can be enhanced with behavioral and/or social learning procedures. Coaching has received considerable empirical support as a social skills training procedure.

Several interventions have been developed that stress teaching children the process of solving social or interpersonal problems as a way to facilitate socially appropriate behaviors. As Weissburg noted in 1985, some of these intervention programs, which are largely classroom based, have been called social problem solving (SPS) programs, whereas others have been called interpersonal cognitive problem solving (ICPS) programs. ICPS programs generally place greater emphasis on cognitions that parallel social problem situations and employ narrower training procedures than SPS approaches; however, both use a similar training sequence to help students identify and cope with social problems. Briefly, the steps can be described as follows: (i) Identify and define the problem, (ii) determine alternate ways of reacting to the problem, (iii) predict consequences for each alternative reaction, and (iv) select the "best" or most adaptive alternative. Social problem-solving methods can be used with individual children or with entire classrooms.

## **4. EFFECTIVENESS OF SOCIAL SKILLS INTERVENTIONS**

The popularity and widespread use of social skills training procedures have resulted in several traditional and meta-analytic reviews of the effectiveness of these procedures with children. Schneider and Byrne, who conducted the first major meta-analysis of social skills training studies in 1985, reported social skills interventions were more effective for preschoolers and adolescents than elementary children. No gender differences in effect sizes were noted, although few studies have treated gender as an independent variable. In addition, social skills training was found to be more effective for students

exhibiting social withdrawal or learning disabilities than for aggressive students. In a 2001 meta-analysis of social skills training with antisocial and aggressive youths, Ang and Hughes found a moderate effect size for the skills training intervention. Interestingly, group composition appeared to mediate the effectiveness of the intervention. Intervention groups that were composed of deviant youth and their prosocial peers had higher effectiveness than did intervention groups composed only of deviant youth. This meta-analysis provides further evidence for the positive effect that prosocial peer models can have on social skills training interventions. Forness and Kavale's 1996 meta-analysis of 53 social skills intervention studies of students with learning disabilities found a similar positive, but less powerful, effect for social skills training. Specifically, they reported a mean effect size of only 0.21. This translates to a modest 8% percentile rank increase in performance when compared to untreated control students. Many of the studies reported by Forness and Kavale were brief interventions—less than 3 hours per week over a 10-week period.

Based on these reviews of research, there appears to be good support for the moderate effectiveness of social skills training procedures in general, and in particular for operant and modeling procedures. However, treatment effects often are short-lived and do not generalize to other settings or behaviors unless specific intervention steps are taken. For treatments delivered in social skills training to be considered truly effective and valid, levels of change must evidence treatment generality. Treatment generality is composed of maintenance (behavior change that persists over time) and generalization (behavior changes that occur under nontraining conditions). Although treatment generality is often evaluated as a “hope-for-the-best” side effect of social skills training, maintenance and generalization should be systematic facets of any intervention. Several authors have discussed procedures to enhance treatment generality. These generality facilitators include (i) targeting behaviors that will be maintained by natural reinforcement contingencies; (ii) training across different behaviors, settings, and persons; (iii) training loosely (i.e., varying stimuli such as reinforcers and tone of voice); (iv) systematically withdrawing or fading intervention procedures to approximate the natural environment; (v) reinforcing behavior change in novel and appropriate settings; (vi) including peers in training; and (vii) providing booster sessions.

## **5. SELECTION OF INTERVENTIONS FOR STUDENTS WITH DIFFERENT PROBLEMS**

The selection of social skills interventions rests heavily on the classification of social skills difficulties as resulting from deficits in either response acquisition or response performance. Gresham and Elliott's Social Skills Rating System extended this two-way classification scheme to include areas of social skills problems, social skills strengths, and potential concurrent interfering problem behaviors. This scheme distinguishes whether a child possesses the ability to perform a target skill and whether interfering behaviors (e.g., aggression and anxiety) are present.

### **5.1. Social Skills Acquisition Deficits**

This social skill problem characterizes children who have not acquired the necessary social skills to interact appropriately with others, or children who have not acquired a critical step in the performance of a given skill. Training in skill acquisition frequently uses direct treatment approaches, such as direct instruction, modeling, and coaching.

### **5.2. Social Skills Acquisition Deficits with Interfering Problem Behaviors**

This social skills problem describes children with emotional (e.g., anxiety, sadness, and impulsivity) and/or overt behavioral (e.g., verbal or physical aggression, tantrums, and excessive movement) responses that interfere with skill acquisition. Thus, with social skills acquisition deficits accompanied by significant interfering behaviors, the intervention objectives are to teach and increase the frequency of prosocial behaviors while concurrently decreasing or eliminating interfering problem behaviors. Interventions designed to remediate emotional responses typically involve emotional-arousal reduction techniques, such as desensitization or flooding, paired with self-control strategies, such as self-talk, self-monitoring, and self-reinforcement. Interventions that can help reduce overt behaviors are often referred to as reductive procedures. These procedures include the use of differential reinforcement, group contingencies, and mild aversive techniques (e.g., reprimands, time out from positive reinforcement, response cost, and overcorrection). If aversive techniques are used, it is a best practice to couple aversives with socially positive procedures.

### 5.3. Social Skills Performance Deficits

Children with social skills performance deficits have appropriate social skills in their behavior repertoires but do not perform the behavior at acceptable levels and/or at appropriate times or in appropriate settings. Interventions for social skills performance deficits typically use manipulations of antecedent and consequent contingencies such as peer initiations, social reinforcement, and group contingencies.

### 5.4. Social Skills Performance Deficits with Interfering Problem Behaviors

Children with social skills performance deficits accompanied by interfering problem behaviors have acquired given social skills, but performance of the skills is hindered by emotional or overt behavior responses and by problems in antecedent or consequent control. Self-control strategies to teach inhibition of inappropriate behavior, stimulus control training to teach discrimination skills, and contingent reinforcement to increase display of appropriate social behavior are often used to ameliorate this social skills problem. Occasionally, when the interfering behaviors persist, it may be necessary to use reductive methods in addition to positive techniques.

### 5.5. Social Skills Strengths

Many children exhibit social skills strengths that often may be overlooked and may be undermined through lack of attention. Therefore, it is a best practice to take a proactive approach to maintaining those strengths through reinforcement of desired social behaviors. Children with strong social skills can also become part of a treatment strategy for children with social skills deficits by serving as models or as participants in peer pairing and peer initiation strategies.

### See Also the Following Articles

Attention Deficit Disorders: School-Based Interventions

- Attention Deficit/Hyperactivity Disorders (ADHD)
- Emotional and Behavioral Problems, Students with
- Study Skills Instruction

### Further Reading

- Ang, R. P., & Hughes, J. N. (2001). Differential benefits of skills training with antisocial youth based on group composition: A meta-analytic investigation. *School Psychology Review, 31*(2), 164–185.
- Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*(4), 302–306.
- Cartledge, G., & Milburn, J. F. (1995). *Teaching social skills to children: Innovative approaches* (2nd ed.). New York: Pergamon.
- Demaray, M. K., Ruffalo, S. L., Carlson, J., Olson, A. E., McManus, S., Leventhal, A., & Busse, R. T. (1995). Social skills assessment: A comparative evaluation of published rating scales. *School Psychology Review, 24*, 648–671.
- DuPaul, G. J., & Eckert, T. L. (1994). The effects of social skills curricula: Now you see them, now you don't. *School Psychology Quarterly, 9*, 113–132.
- Elliott, S. N., & Gresham, F. M. (1991). *Social skills intervention guide: Practical strategies for social skills training*. Circle Pines, MN: American Guidance Service.
- Elliott, S. N., & Gresham, F. M. (1993). Social skills interventions for children. *Behavior Modification, 17*, 287–313.
- Forness, S. R., & Kavale, K. A. (1996). Treating social skills deficits in children with learning disabilities: A meta-analysis of the research. *Learning Disability Quarterly, 19*(1), 2–13.
- Gresham, F. M. (1994). Generalization of social skills: Risks of choosing form over function. *School Psychology Quarterly, 9*, 142–145.
- Gresham, F. M., & Elliott, S. N. (1989). Social skills deficits as a primary learning disability? *Journal of Learning Disabilities, 22*, 120–124.
- Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Gresham, F. M., & Reschly, D. J. (1988). Issues in the conceptualization, classification, and assessment of social skills in the mildly handicapped. In T. R. Kratochwill (Ed.), *Advances in school psychology* (pp. 203–247). Hillsdale, NJ: Erlbaum.
- Guevremont, D. (1990). Social skills and peer relationship training. In R. A. Barkley (Ed.), *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment* (pp. 540–572). New York: Guilford.
- Lentz, F. (1988). Reductive techniques. In J. C. Witt, S. N. Elliott, & F. M. Gresham (Eds.), *The handbook of behavior therapy in education* (pp. 439–468). New York: Plenum.
- Mathur, S. R., & Rutherford, R. B. (1991). Peer-mediated interventions promoting social skills of children and youth with behavioral disorders. *Education and Treatment of Children, 14*, 227–242.

- Merrell, K. W. (2003). *Behavioral, social, and emotional assessment of children and adolescents*. Mahwah, NJ: Erlbaum.
- Mize, J. (1995). Coaching preschool children in social skills: A cognitive-social learning curriculum. In G. Cartledge, & J. F. Milburn (Eds.), *Teaching social skills to children and youth* (3rd ed., pp. 237-261). Boston: Allyn & Bacon.
- Sheridan, S. M., & Walker, D. (1999). Social skills in context: Considerations for assessment, intervention, and generalization. In C. R. Reynolds, & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 686-708). New York: Wiley.
- Stokes, T. F., & Osnes, P. (1989). An operant pursuit of generalization. *Behavior Therapy*, 20, 337-355.



# Social Support

Christine Schwarzer and Petra Buchwald

Heinrich Heine University, Düsseldorf, Germany

1. Introduction
  2. Different Meanings of Social Support
  3. Social Support and Health
  4. Support Groups and Self-Help Groups
  5. Social Support in the Elderly
  6. Social Support and Gender
  7. Social Support and Social Skills
- Further Reading

## GLOSSARY

- emotional support* Provision of liking, reassurance, and respect.
- instrumental support* Aid with services and tasks of daily living.
- perceived support* Support perceived to be available if the need for it arises in the future.
- received support* Support actually received if needed in the past.
- social network* A group of individuals with whom an individual may exchange support.

Social support is information that leads individuals to believe that they are valued, respected, and loved and that helps them to cope with major life stressors and the challenges of everyday life. Social support is a special kind of social interaction that can appear in different ways and that can be both a psychological and a tangible resource provided by a social network,

for example, by friends, family members, and/or colleagues.

## 1. INTRODUCTION

Social support refers to the fundamental experience that people need to stay healthy both physically and mentally, to get proper aid and comfort, and to integrate morally and behaviorally into the group or society at hand. The positive effects of social contacts seem to have been well known since the beginning of humankind. Although the positive effects of social support are an integral part of a healthy life, there is much more to learn about the theoretical underpinnings, conceptualizations, and operationalizations of this construct. During the past 30 years or so of research, social support has become one of the central variables in many fields of psychology, epidemiology, public health, and education. It seems to be a potent variable for helping to explain the beneficial or detrimental effects of other people being around an individual who is coping with the challenges of daily life and/or major life stressors. But some crucial questions still remain within the community of researchers. What is the real nature of social support? When does social support occur? Are there gender differences with respect to giving or receiving social support? How can we measure social support? Is social support always helpful? What are the consequences of research outcomes in this field for applied psychologists? This article focuses on questions such as

these and provides an outlook for future research and practical applications.

## 2. DIFFERENT MEANINGS OF SOCIAL SUPPORT

Because social support occurs within the interaction of people, it is central for applied psychologists to find out what constitutes its effects. Is it just the fact of being together with others? Is it the certitude that one is loved and honored by others who console? Or is it just the helpful hands that others lend by giving financial or other tangible aid? The different viewpoints that can be found in the research literature stem from various theoretical perspectives. More than a century ago, Durkheim was one of the first to observe a higher risk of committing suicide among socially isolated individuals than among those with regular social bonds. In line with this, there exists a research tradition of looking for the objective quantitative roots of social support. This structural view is interested in the size, structure (e.g., number of males and/or females), and density of social networks and in the amount of embeddedness they give to people. However, the number of network members or the number of contacts among them is not necessarily related to psychological well-being or the subjective perceptions of being helped. Because relationships may be unhelpful as well as supportive, the focus of research has shifted from network size to an understanding of individual perception of social support as well as the context in which it is given. In contrast to the structural perspective, the more functional view is interested in relational content, which can be subdivided into perceived availability of support and actually received support. The functional view relies on self-report data of the individual either on how many helping acts (and from whom) he or she can remember for a given time in the past (received support) or on who will help in a stressful situation in the future (perceived support). Following the latter point of view, researchers are not as interested in the structure as in the content of support and focus on what people really do to each other in a network, for example, providing emotional support, informational support, instrumental support, or appraisal support. Emotional support involves the message of being loved, cared for, esteemed, and connected to other people in a network of mutual obligation. Informational support can be guidance with problem solving and advice. Instrumental support includes tangible elements

such as housing, transportation, and money. Appraisal support offers personal feedback concerning personal attitudes and behavior. These various research approaches make the construct a multifaceted one that can hardly be defined in a way that incorporates all possible perspectives. But in a global sense, social support can be defined as information that people receive within social interactions, including actual assistance and feelings of embeddedness in a social network that is believed to provide love and care.

Social support can be provided by formal or informal sources. An informal and very important support source is the family (e.g., parents, spouses, other relatives). Other informal sources of support might include friends, partners, coworkers, peer groups, and neighbors. Formal social support may come from services such as self-help groups, drop-in centers for information, and formal community services. For example, schools, which often seem to be a source of stress for pupils, can also be a source of social support for children who experience problems with their parents. Hence, these formal support sources can be focused on specific populations of individuals who may be at risk, such as children and older people living in isolation, or they may be open to all people in general. Empirical studies indicate that both formal and informal support sources are positively related to health and well-being, although people seem to rely more on informal support sources for emotional help and seem to prefer formal sources for receiving instrumental support. In addition to human help, household pets can provide support. Analyses of pet support measures show that pets fulfill many of the same support functions as do humans (e.g., emotional support, social integration, closeness) and that pet support, like human support, is associated with less stress and better adjustment. However, studies that have analyzed sources of support are limited. Thus, further research should collect information about sources of social support in special groups differentiating special kinds of support.

Sometimes, social support not only is appreciated but also can be seen as a source of stress. The mere offering of support can be very stressful for a recipient if he or she does not want help from others in that particular situation. Rook was one of the first researchers to make a significant contribution to investigating the negative aspects of social support. There is some empirical evidence for the negative effects of received assumptive help, that is, help that is neither asked for nor needed. This implies a certain dilemma: If help might be appraised as threat or control, it might be better not to

offer any help. On the other hand, not offering assistance might hurt a potential recipient who really needs it. In both cases, the potential helper has reasons to worry that his or her motives will be questioned. Furthermore, if the offer of help is rejected, the helper's self-esteem may be threatened. Therefore, support providers who acquire social skills, such as knowing appropriate behaviors for certain situations, produce less negative supportive actions than do those who do not. Social skills not only improve the interactions with others but also boost the understanding of others' needs and desires. Therefore, socially skilled helpers provide support when it is needed and decrease support in situations where others prefer to help themselves.

According to the threat to self-esteem model of Nadler and Fisher, receiving help is experienced as self-supportive when recipients feel appreciated and cared for but is experienced as self-threatening when recipients feel inferior and overly dependent. If recipients feel supported, they react with positive feelings, acceptance, and gratefulness to the helpers. But if recipients feel threatened by the help they receive, this is accompanied by negative emotions and evaluations of both helpers and recipients. There are three conditions under which receiving help could be perceived as threatening. First, individuals with high self-esteem regard themselves as highly competent and are especially sensitive to the implication that they are unable to take care of themselves. As a result, they tend to react more negatively to receiving help than do those with low self-esteem. Second, being helped by other competent people highlights the contrast of being in need of assistance between helpers and receivers. Third, receiving help from significant others on important ego-relevant tasks can be threatening to recipients' self-esteem.

However, high self-esteem does not seem to prompt negative reactions to social support from those who are close to recipients. In close relationships, those negative effects of similarity do not occur because similarity is expected and desired. Feelings of inferiority are less likely to arise because mutuality in such close relationships makes receiving help less threatening. In addition, age might be an important moderator of experiencing negative effects of support in close relationships. Dependency is acceptable for children, and they react negatively to the help of their parents less often. On the other hand, older parents will not as easily accept help from their adult children.

On the basis of the previously described threat to self-esteem model, many empirical studies have been launched. Table 1 summarizes some outcome variables

TABLE 1  
Positive and Negative Effects of Social Support

<i>Positive effects</i>	<i>Negative effects</i>
<ul style="list-style-type: none"> <li>• High self esteem</li> <li>• Self-worth</li> <li>• Self-efficacy</li> <li>• Optimism</li> <li>• Joy</li> <li>• Health, life satisfaction</li> <li>• Sense of belonging</li> <li>• Embeddedness</li> </ul>	<ul style="list-style-type: none"> <li>• Low self-esteem</li> <li>• Inferiority</li> <li>• Helplessness</li> <li>• Hopelessness, depression</li> <li>• Anger</li> <li>• Illness, psychological impairment</li> <li>• Overprotection</li> <li>• Restriction of freedom, demands of conformity, social control, social conflicts</li> </ul>
<ul style="list-style-type: none"> <li>• Gratefulness</li> <li>• Adaptive coping</li> </ul>	<ul style="list-style-type: none"> <li>• Indebtedness to the donor</li> <li>• Dysfunctional coping</li> </ul>

that were part of empirical studies analyzing the positive and negative impacts of social support. These studies also showed that the association between unsupportive interactions and outcome variables did not result from a confounding with the severity of an illness or the level of psychological impairment.

### 3. SOCIAL SUPPORT AND HEALTH

Since Cassel hypothesized in 1974 that, under stressful conditions, the lack of social support will lead to a higher risk of becoming ill, research has shown that supportive interactions among people are protective against the health consequences of life stressors. It appears that social support can protect people in crises from a wide variety of pathological states, from low birthweight to death. However, recent research reviews of the empirical evidence of social relationship factors being protective against life stressors have yielded rather inconsistent results. One of the main reasons for such inconsistent findings might be the rich diversity of definitions and conceptualizations of social support and the complexity of the support process. A person's ability to cope effectively depends on the stressor itself, the type of preferred coping resource, and the resources available to the individual. There are several major types of coping resources, and social support is only one of them.

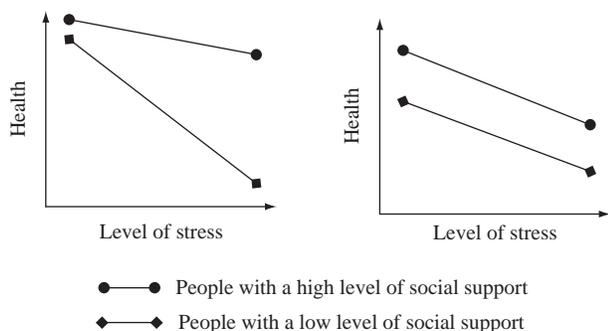


FIGURE 1 Stress buffering effect (left) and direct effect (right) of social support.

Investigating the question of stress buffering versus the direct effect provided a better understanding of the effects of social support. According to the stress buffering hypothesis, social support is needed only when people are under high stress. In that case, social support can buffer the negative stressful impact of critical life events such as divorce, loss of a loved one, chronic illness, pregnancy, job loss, and work overload on health. This means that the health and mental health benefits of social support are evident mainly during periods of high stress, whereas social support may have few physical or psychological benefits during periods of low stress. The stress buffering effect is depicted on the left-hand side of Fig. 1. Another model that explains how social support affects people suggests a direct effect and is illustrated on the right-hand side of Fig. 1. Here, social support has a direct positive effect on physical or mental health independent of stress levels. In other words, regardless of whether people are experiencing stressful situations, social support is generally beneficial.

Empirical evidence for both direct and buffering effects of social support has emerged. In 1986, Hobboll reported that the direct and stress buffering effects occur simultaneously and that one type of effect is more prominent than the other in a given instance.

An overwhelming amount of evidence shows that social support has effects on both psychological and physical health. One remarkable result for physical health is that social support reduces the risk of dying from disease, suicide, or accidents by about the same percentage as does smoking. Moreover, increased social support from family, friends, and/or colleagues is related to decreased psychological distress. With help and support, people are psychologically better able to manage critical life events, such as unemployment and divorce, as well as everyday problems and hassles.

In Fig. 2, a theoretical framework for social support as a coping resource helps to illuminate the effects of social support on health, referring to stress theoretical underpinnings.

#### 4. SUPPORT GROUPS AND SELF-HELP GROUPS

The rapid increase in the number of support groups and self-help groups all over the world over the past 15 years or so is likely due to rising health care costs, an increase in alternative health care, and an emphasis on prevention. Self-help groups can be seen as an attempt by people in stressful situations to enhance the amount of mutual help and cooperation they receive. They are groups of people who come together for support because they share common problems or experiences and

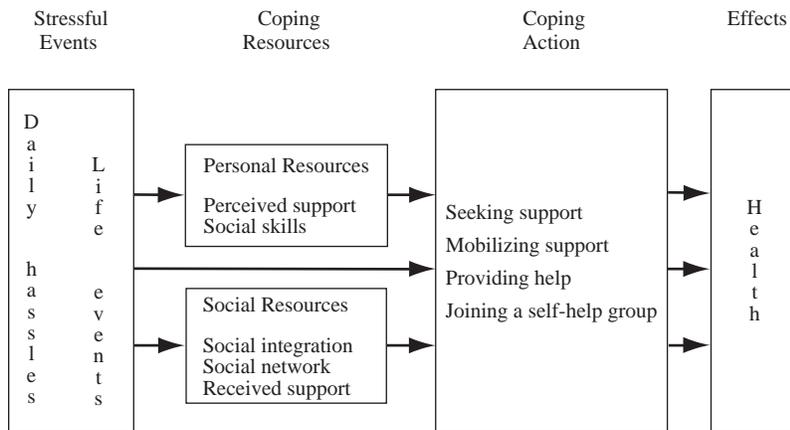


FIGURE 2 Theoretical framework for social support as a coping resource.

meet to give and receive support. Making the distinction between social support groups and self-help groups is important for practitioners in deciding which groups will best serve their clients' needs. Support groups are member centered, although they are guided by trained leaders who provide expertise in group counseling but might not share the concerns that bind the groups together. In contrast, self-help groups are mostly leaderless or guided by nonprofessionals. For example, people who are learning to deal with specific problems find support in self-help groups such as Alcoholics Anonymous. There are groups not only for alcoholics but also for former drug addicts, divorced people, single parents, cancer patients, and so on. Support groups and self-help groups both help people to cope not only because they provide assistance but also because people can learn coping techniques from others with similar problems. Recent research reveals that young unmarried White women with less social support in other areas of their lives are more likely to join self-help groups than are other individuals. Spiegel showed, for example, that women with advanced breast cancer who joined a support group not only benefited emotionally from this experience but also lived longer than did similar women who did not attend a support group.

Self-help groups can be seen as a special kind of communal coping, in contrast to individual coping efforts. Self-help groups provide advantages that are not found in individual ways of coping. During times of stress and emotional trouble, knowing that others have similar problems can be very reassuring. In addition, seeing others' problems improve can be a source of hope and motivation. Because self-help group members typically have comparable problems, they can learn from each other's mistakes and share insights. Furthermore, when a group member receives similar comments about his or her behavior from several group members, the message may be more convincing than when it comes from a single friend or therapist. Therapists often direct their patients to self-help groups to supplement their individual therapy. For example, a person with an alcohol problem can find comfort and help in a group with others who have the same problem. They exchange useful information, share their coping strategies, and gain hope by seeing others overcome or successfully manage their shared problems. Research on self-help groups for alcoholism and obesity strongly support the hypothesis of the effectiveness of self-help groups either alone or in combination with individual counseling.

## 5. SOCIAL SUPPORT IN THE ELDERLY

The elderly must face a life situation that is often characterized by stressful events such as the loss of loved ones, progressive health impairment, and disability that make them vulnerable to emotional disturbances. The elderly can be regarded as a risk population that is confronted with dependency, social isolation, illness, and the threat of death. Especially for people age 65 years or over who have experienced both physical and psychological limitations and impairments as side effects of aging, the beneficial effects of social support have been increasingly reported. Social support can be seen as one social key resource for the elderly, and many studies consistently show that perceived support exerts the strongest effects on health and well-being in late life. The general consensus is that those elderly persons who perceive a high amount of social support are better able to withstand the negative impact of stress than are their counterparts who receive less support. However, as the literature has shown, this process is by no means simple and social support functions in complex ways. For example, support indicators, such as the number of regular contacts and the amount of actually received social support, also play an important role in physical and emotional well-being in old age.

To enhance the understanding of how social support operates, it seems necessary to link social resources with personal resources within a resource-oriented life span model of successful aging because social resources, such as assistance, guidance, and presence provided by others, may be filtered through the existing characteristics of a person. This idea is generally confirmed by research on social support and control beliefs highlighting the social foundation of the control processes. For example, one could show that feelings of control beliefs and self-efficacy are enhanced by received social support. Perceived social support, in contrast, has weak relationships with perceived control but closer relationships with indicators of subjective health. However, received social support might be interpreted as being more central to initiating and maintaining behaviors that enable the elderly to perceive a sufficient feeling of control. Especially in old age, people's opportunities to accomplish their own performances are limited; they are handicapped, and specific abilities are reduced. This means that older people often cannot experience control in special domains, but they do have vicarious experiences, verbal persuasion, and other social influences that convey control information. By providing continued

support and enhancing control beliefs, social network members may help older people to avoid a downward spiral of resources that ultimately leads to a relapse of health. Following Baltes's theory of optimization with compensation, one can argue that older people cope with such an ongoing spiral of losses by giving up control in some life domains so as to maintain it in others. Or, they optimize control by compensation of resource loss with received support from others.

In addition, received social support is an important resource not only with regard to control beliefs but also with regard to emotions. Within one sample of senior citizens, it was observed that older people were more anxious than younger people. Low social integration and low social support were associated with higher levels of anxiety. Furthermore, it turned out that received social support, as measured by the number of visits, served as a moderator for the relationship between loss and anxiety. Those who recently experienced the loss of friends or family members and who were not regularly visited by others were the most anxious, whereas being contacted regularly served as a buffer against anxiety.

## 6. SOCIAL SUPPORT AND GENDER

Gender seems to make a difference as well, especially with respect to various types of support. Women are more likely than men to develop intimate friendships, to cooperate with others, and to provide support for their colleagues, friends, and loved ones. There is already broad empirical evidence for gender differences in the participation in social networks. Male participation is seemingly more extensive but less intensive. Men prefer large numbers of acquaintances, whereas their female counterparts prefer smaller numbers of intimate friendships. Differences in intensity and quantity of participation are resumed to be attributable to the activity-focused type of relationships preferred by men (e.g., sports, politics), in contrast to the emotion-focused type of relationships favored by women (e.g., chatting groups). However, such different preferences should be viewed more as general patterns than as rigid prototypes.

Furthermore, the quality and quantity of the participation in social networks seem to be a function of the life span. For example, whereas girls are principally help seekers during childhood, they develop into help providers during adolescence and adulthood.

Gender is related not only to differences in social network participation but also to differences in the

willingness to seek and mobilize social support. Whereas women tend to draw on extended network support such as friends, family members, and partners, men tend to rely solely on their partners. Men's reluctance to seek social support beyond their partners might be due to the attributed femininity of emotional intimacy. Also, for relatively minor problems, men seek help less frequently than do women because mobilizing social support is more threatening to their self-esteem. Mobilization of support is also subject to life cycle and emancipation effects. Although emotional intimacy and affirmation from spouses are crucial to the emotional well-being of married men and unemployed wives, this priority is altered by female participation in the labor market. Employed women appreciate equity in the marriage in conjunction with affirmation from their spouses. However, social support entails not only the receipt of support but also the provision of support. Men and women show similar gender-specific preferences for support in that both choose more female support sources. This trend holds true not only in the private sphere but also in the public sphere, for example, in care services where women are overrepresented.

In general, gender-specific behavior concerning social support might be due to different norms for males and females in Western cultures. Part of the male sex role is to be courageous and adventurous, whereas part of the female role is to be nurturant and caring, valuing close long-term relationships. As a result, men are more likely to help in situations that demand heroic acts, and women are more likely to help in long-term relationships that involve less danger. The effects of these norms are striking. Meta-analyses have shown that men with a low level of social support are two to three times more likely to die than are men with a high level of social support. The results for women are similar but not as striking. Women with a low level of social support are one and a half to two times more likely to die than are women with a high level of social support.

## 7. SOCIAL SUPPORT AND SOCIAL SKILLS

People under stress have good reason to seek other people's help. The resources that the social network offers may reduce the impact of stressful events. Thus, social support is a major way in which to get resources that are not directly available for individuals. Therefore, an important field of intervention is enabling people to

receive meaningful help. What people need are social skills to communicate needs and desires and to mobilize social support when it is needed. Social skills training is widely used to train lonely people to make friends. People who have deficient social skills may learn how to act in a variety of social situations, for example, by observing others with appropriate social skills. They also might practice their new skills by role-playing with video-supported microteaching or modeling from films before applying the new behavior in real life. The results on such training have been very successful, although the main problem is to find the best skill for a special person in a particular situation with a special problem. In addition, the appropriateness will vary with gender, culture, social class, and context.

### See Also the Following Articles

Elder Caregiving ■ Gender and Culture ■ Social Networks  
 ■ Social Skills Training ■ Stress

### Further Reading

- Bowlby, J. (1969). *Attachment and loss, Vol. 1: Attachment*. London: Hogarth.
- Buchwald, P., & Schwarzer, C. (2002). Analyzing the mutual influence of social support and control beliefs in the elderly. In C. Schwarzer, & M. Zeidner (Eds.), *Developmental issues in stress and coping* (pp. 141–163). Aachen, Germany: Shaker Verlag.
- Cohen, S., Gottlieb, B., & Underwood, L. (Eds.). (2000). *Social support measurement and intervention: A guide for health and social scientists*. Oxford, UK: Oxford University Press.
- Cohen, S., & Syme, S. L. (Eds.). (1985). *Social support and health*. New York: Academic Press.
- Duck, S. (Ed.). (1990). *Personal relationships and social support*. London: Sage.
- Gottlieb, B. H. (Ed.). (1981). *Social networks and social support*. Beverly Hills, CA: Sage.
- Hobfoll, S. E. (1986). *Stress, social support, and women*. Washington, DC: Hemisphere.
- Sarason, B. R., Sarason, I. G., & Pierce, G. R. (Eds.). (1990). *Social support: An international view*. New York: John Wiley.
- Schwarzer, C. (1992). Anxiety, stress, and received support in the elderly: A longitudinal analysis. *Anxiety Research*, 4, 287–298.
- Vanzetti, N. A., & Duck, S. (Eds.). (1995). *A lifetime of relationships*. Pacific Grove, CA: Brooks/Cole.
- Vaux, A. (1988). *Social support: Theory, research, and intervention*. New York: Praeger.



# Spatial Cognition

Reginald G. Golledge

University of California, Santa Barbara, California, USA

1. Introduction
  2. Development of Spatial Cognition: A Multidisciplinary Effort
  3. Spatial Cognition and Spatial Representation: Two Sides of the Same Coin
  4. Human Spatial Cognition
  5. Spatial Cognition in Animals
  6. The Neurobiology of Spatial Cognition
  7. Age-Related Differences in Spatial Cognition
  8. Sex-Related Differences in Spatial Cognition
  9. Spatial Cognition without the Benefit of Sight
  10. New Frontiers in Spatial Cognition Research
- Further Reading

## GLOSSARY

**cognition** Deals with knowledge, including its acquisition, storage, retrieval, internal manipulation, and use. Stored and manipulated knowledge helps organisms achieve behavioral goals. Thus, it includes processes such as perception, learning, thinking and reasoning, memory, recall, problem solving, and language.

**cognitive map** A hypothetical construct implying the existence of an internal representation of an environment and the spatial relations among the environmental features.

**cognitive mapping** The dynamic process that involves acquiring, storing, internally manipulating, and using information about the external environment.

**path integration** A process by which a traveling organism self-monitors rates of travel and distance traveled as movement away from a home base takes place.

**place cells** The fundamental element of the hippocampal spatial mapping system; the predominant cell type in the hippocampus, comprising perhaps 90% of the cells. Place cells in the hippocampus are conceived as small, localized fields, which allow them to play a role in many spatial mappings. Place cells are part of a memory system showing location-appropriate information.

**spatial cognition** The internalized reflection and reconstruction of space in thought.

**spatial representation** An internal representation of an environment upon which an active information-seeking structure can be imposed. This internal representation of the environment enables an organism to change the location or orientation of its entire body (as opposed to perceptual, postural, and manipulatory space). Spatial representation is often used in preference to the term spatial cognition when dealing with animals other than humans.

In the multidisciplinary area of spatial cognition research, underlying theory has mostly come from cognitive and development psychology and neuropsychology. The contribution of other disciplines, such as geography, planning, architecture and design, anthropology, computer science, artificial intelligence, and information science, has often been focused on the applied facets of spatial cognition. Applications include specifying the contribution of spatial cognition to navigation and way-finding, understanding environmental settings, building environmental schemata, learning in both familiar and unfamiliar environments, spatial knowledge acquisition across the ages, the development and measurement of

spatial skills and abilities, sex-based differences in spatial cognition, and cognitive maps as cultural universals. This breadth of interest enables examination of the complementary concepts of spatial cognition and spatial representations in both the human domain and the domain of other animals (particularly mammals). Questions relating to cognitive mapping and the neurobiology of spatial information processing lead to examination of age- and sex-based factors in spatial cognition and to the study of spatial cognition in disabled people. Spatial cognition is considered a necessary basis in many research contexts, and this article concludes with speculations on promising areas for research.

## 1. INTRODUCTION

Philosophical debates about space have continued for millennia. They can be summarized into two major themes: (i) whether spatial knowledge results from experience (the empiricist position) or whether it is constructed by the mind (the rationalist or nativist position) and (ii) whether psychological spatial concepts reflect physical realities of external spaces or whether there may be striking differences between psychological and physical space.

Spatial cognition gained visibility in part because more than 100 years of experimental work in perception provided little insight about how we knew where in the world (or in the macrospace surrounding us) we were located and how we knew to select destinations and move toward them. Thus, spatial orientation and spatial representation, when added to spatial perception and spatial information processing, provided the missing links to relate human activities and knowledge to real-world objective spaces. Once established, similar processes and reasoning could be transferred to understanding the informational content of subjective (psychological) spaces. The particular contribution of different settings to environmental knowledge was acknowledged in ecological and environmental psychology as well as in urban planning, environmental design, and behavioral geography. Thus, tasks such as finding one's way around an environment, searching for objects, taking shortcuts, and recovering positional knowledge after becoming lost or disoriented all began receiving attention under this umbrella of spatial cognition.

Geographic metaphors have played a significant role in the external representation of cognized spatial information, with particular emphasis on maps, sketches, and diagrams as illustrative modes. In the early years of

interest in spatial cognition, there was confusion regarding the term cognitive map. Some regarded it as a map-like way in which spatial information was stored in the brain; some used the term to describe spatial products or externalizations of stored spatial information.

In the human domain, the cognitive map system allows people to locate themselves at specific places within a frame of reference, remember what is located where, and determine how to get from one place to another in response to needs and motivations. The cognitive map also facilitates learning about environments and the spatial relations among environmental features.

Reasons for assuming that spatial cognition is a cultural universal include

- Organizational similarities of the nervous system
- Common body structures and processes, including sensory and motor activities
  - Learning and socialization similarities
  - A similar need to cope with complex physical environments
    - Use of cognitive maps that engender spatial relations and spatial inferences
      - The influence of gravity and the consequent perception of a three-dimensional world with a horizon boundary
        - The influence of scale in the processing of spatial information in memory, reasoning, or language
        - Classification of complex environments into categorical and hierarchical regions
          - The importance of labeled features in memory and their use in problem solving
            - Use of multiple frames of reference, including egocentric (body-related), exocentric (object-to-object), and environment-related
              - The common understanding of angularity
              - The tendency to apprehend three-dimensional spatial relations of environments by reducing them to two-dimensional representations
                - The use of spatialization to produce visualizations of nonspatial information

The past two decades in particular have seen considerable research emphasis on understanding spatial cognition in humans, animals, and technical systems (computers and robots). The concept is common in many disciplines, and contributions toward understanding have come from the areas of psychology, neuroscience, philosophy, linguistics, geography, anthropology, architecture, computer science, and artificial intelligence. The result has been a rapidly growing

literature focused on activities such as wayfinding and navigation, route learning, acquisition of spatial knowledge in familiar and unfamiliar environments, strategies for regaining orientation after becoming lost, sex-based differences in spatial cognition, individual differences in spatial abilities, the development of spatial knowledge across the life span, and the building of artificial intelligences in robots that adopt strategies and perform spatial tasks either in mathematically optimal ways or in ways that closely mimic human thought and actions.

A goal common to all these areas is to understand the internal spatial representations that humans and other animals build from the experienced or imagined world. Spatial cognition goes beyond the analysis of sensory input via vision (the focus of image processing and the psychology of perception) because it includes recognition and categorization of features and objects using all the senses. Spatial concepts include, as a minimal (primitive) set, place name or identity, location, magnitude, and time. Higher order (more complex) concepts can be derived from these; the latter include spatial distribution, spatial arrangement, pattern, shape, size, texture, distance, direction, angle, reference frame, scale, link, network, region, and many others that make up a spatial concept ontology. However, spatial cognition also goes beyond object recognition to include how spatial or environmental information is encoded and how people, animals, and machines either think about space or construct representations that can be used in various reasoning processes.

Spatial cognition is revealed by performing tasks (e.g., navigation and wayfinding), by constructing external representations of stored information (e.g., spatial products such as sketch maps), by the effective use of spatial language to communicate (e.g., spoken or written spatial language), and by the ability to spatialize nonspatial information (e.g., by representing it in the form of graphs, diagrams, sketches, maps, images, art, sculpture, dance, gestures, or other modes of expression). Since it is generally accepted that vision is the spatial sense par excellence, many representations are achieved via visualizations (e.g., by diagramming). Spatial information can also be represented and used effectively in haptic and auditory domains, and a considerable portion of the natural languages used by different cultures for interpersonal communication consists of spatial concepts. These range from place names to fuzzy prepositions and prepositional phrases, such as “above,” “below,” “close to,” “far from,” “along,” and “across.”

Spatial information is well represented diagrammatically. Two traditional forms of representation are

iconic and relational. In the iconic diagram, concrete ties among objects are established by scaled spatial relations, which are isomorphic to the originally observed relations (e.g., drawings of parts of the human anatomy). Relational diagrams are used to show spatial arrangements between entities represented by abstract, nonvisual relations, such as may be found on flowcharts, data graphs, and different topological diagrams.

## 2. DEVELOPMENT OF SPATIAL COGNITION: A MULTIDISCIPLINARY EFFORT

Spatial cognition is generally accepted to have emanated from Tolman's place learning theory, published in 1948. Researchers in psychology argued about differences between place learning and response learning. Other disciplines (particularly planning, design, architecture, and human geography) embraced the fundamental idea that people learned environments by encoding, storing, recalling, and using place-based information, and they took the lead in exploring ways to externally represent a person's knowledge about variously scaled environmental settings as a way to recover, represent, and use stored spatial knowledge for purposes of planning and policy making. Methods such as sketch mapping, verbal descriptions of places, and the investigation of latent spatial structures stored in memory were examined. The latter were made evident using nonmetric multidimensional scaling procedures based on paired comparison proximity or similarity data.

Spatial cognition thus developed as an interest in the way humans and other animals think about space and how they encode, store, and internally manipulate data to produce spatially relevant information that can be used in solving problems of everyday life. These problems require an understanding of spatial concepts such as location, identity, magnitude, space-time, distance, direction, orientation, frame of reference, connection, network, region, spatial pattern, shape, and distributional characteristics such as density and dispersion.

Researchers examined questions of how information was acquired and turned into knowledge. An important part of Piagetian developmental theories, as seen in the 1967 publication by Piaget and Inhelder, focused on the age-based acquisition of spatial knowledge by children. Exploring adult spatial knowledge acquisition led Siegel and White to offer a theory (called

macrogenesis) in 1975 that suggested that spatial knowledge acquisition began by noticing landmarks and then developing route knowledge as a result of experiential activity and, finally, developing relational or configurational or layout comprehension of an environment. This model influenced research on spatial cognition for more than two decades, producing theoretical and empirical models such as image theory by Kosslyn in 1975, propositional theories by Pylyshyn in 1984, dual coding theory by Pavio in 1971, anchor point theory by Golledge in 1978, and a theory of quantitative accumulation and refinement of metric knowledge by Montello in 1997. Other research, such as that by Allen in 1999, focused on an individual differences approach. Spatial knowledge acquisition thus became the building block for further developments in spatial cognition.

Macrogenesis assumes that there is development from egocentric to exocentric understanding of environments; that there is a transition in the geometries needed to comprehend space from topology to projective to Euclidean forms; and that landmark knowledge dominates in the early phases, followed by route-based knowledge with survey or configurational knowledge emerging as the ultimate abstract and complex knowledge structure. Critics of this macrogenetic model have argued that it fails to take account of knowledge acquisition and loss across the life span as aging occurs, that Piaget's stage of development theory fails to adequately consider individual differences, and that it is not an appropriate model on which to base explanations of adult learning of new environments (because adults already have the ability to think metrically and abstractly about spatial relations). Moreover, there has been constant discussion about the relative importance of landmark vs route-based knowledge acquisition. Some argue that landmark recognition must occur first for locations, anchor routes, and configurations, whereas others argue that most of our knowledge is gained from experiencing space via movement and, therefore, route knowledge must necessarily be the initial and primary source of spatial knowledge acquisition. No consensus exists.

All points of view, however, acknowledge the importance of a frame of reference in the process of acquiring spatial information. It is generally accepted that there are two dominant modes: egocentric, in which environmental information is related to the location of the knowledge seeker, and exocentric, consisting of the spatial relations embedded in the layout of objects regardless of the presence or location of a knowledge seeker.

Attaching nonspatial information to some type of locational or, more generally, georeferenced place is an example of spatialization. This is the process of making nonspatial things spatial for ease of representation, interpretation, and use (e.g., plotting age and income data on a graph).

### 3. SPATIAL COGNITION AND SPATIAL REPRESENTATION: TWO SIDES OF THE SAME COIN

A significant focus within the realm of spatial cognition is the process of cognitive mapping. Evidence of the presence of a cognitive map includes the ability to perform wayfinding activities by following new routes that can include detours and shortcuts. The existence of a cognitive map implies that wayfinding can take place by piloting or by the homing vector activity defined in path integration. Thus, given a current location, a new route can be deduced from memorized spatial relationships between the current location and other locations (egocentric organization) or between different locations (exocentric processing, often associated with imagination conditions).

Research indicates that spatial knowledge only reaches the required degree of coherence to be called a cognitive map in primates. The less loaded term spatial representation is often used for species other than primates. In the latter case, spatial representation is often described in terms of locomotor space, in which an organism evinces an ability to change the location or orientation of its entire body with respect to the surrounding environment. The concept of spatial representation has been tested extensively by examining location behavior with respect to biological needs. Locomotion behavior relies on developing correspondence between information stored in spatial memory and information perceived and encoded about the current task situation.

A cognitive map consists of declarative knowledge. This knowledge is organized and manipulated by cognitive processes during thinking and reasoning. Bits of information relative to some decision-making, choice, or problem-solving situation are manipulated via procedural rules, propositional structures, image schemata, and logic and other modes of thinking and reasoning.

Cognitive maps or spatial representations have several functions. First, they serve as a reference for

comparing current inputs with data stored in memory to detect matching or mismatching of schemata (such as might occur in changed environmental settings). Second, they contribute to the formation of structures or abstract representations that may involve general spatial properties common to a variety of situations. Finally, they facilitate or enable computation and planning of oriented trajectories, such as wayfinding and search within the external physical world. Thus, spatial representations may be regarded as occupying the interface between memory and action.

#### 4. HUMAN SPATIAL COGNITION

Humans are born into a three-dimensional spatial world. In adapting to and living within this world, three spatial reference frames are used: body-centered (egocentric), object-centered (exocentric), and environment-centered (landmark-based). For humans, who constantly perceive the spatial environment as a layout of landscape features, it is possible that our general representation of locomotor space owes as much to our apprehension of space through experience, internal feedback, and optic flow during locomotion as it does to the direct perception of spatial objects scattered throughout the environment. To determine location at any particular moment, spatial updating is essential. This takes place with respect to objects in the environment and with respect to a related frame of reference.

The following are basic assumptions of human spatial thinking:

- Geographic knowledge representations in the human mind are constructed on demand.
- A representation is constructed based on knowledge stored in long-term memory.
- Information in long-term memory is fragmentary and hierarchically structured.
- A representation constructed in working memory can be visual, propositional, or spatialized.

Egocentric relations are related to a body framework, such that front–back and left–right axes are important, as is the up–down axis involved in gravitational force on the body. Exocentric frames are based on object-to-object relations that are often specified in terms of arbitrary coordinate systems, interpoint distances and directions, layout patterns or configurational structure, and layout or object-by-object relations to a superimposed, external frame of reference (e.g., a street system or the latitude and longitude coordinate system). As a

body experiences an environment, sensory input is included with respect to one or another of these frames of reference. There is ongoing discussion as to whether spatial information is orientation free or orientation dependent, whether (or to what extent) alignment forces influence encoding processes, and the extent to which information is view (or perspective) dependent. One pervasive question concerns how information coded in separate environments is integrated by the mind to form spatial relational structures (e.g., networks from individual routes). There appears to be strong evidence that spatial information is stored in memory in a hierarchical fashion that facilitates recall, decoding, and use. There is also evidence of a regionalization effect for spatial information, in that distance estimates between pairs of points are judged smaller for distances within a region than similar interpoint distances between places in different regions. Smaller distances tend to be overestimated and longer distances underestimated (regression toward the mean), and interpoint distances closer to a viewpoint are overestimated relative to interpoint distances far from a viewpoint. Research findings include that distances toward landmarks are often underestimated and distances from landmarks are overestimated, and that distances to landmarks are underestimated in comparison to distances to nonlandmark features. This throws doubt on the validity of a metric assumption (i.e., that distances between points are symmetric) in the world of spatial cognition.

To help us think about and communicate about spatial information, we use two kinds of tools to assist us: visualizations of phenomena (that can be represented on maps and pictures and so on) and spatializations of information (that are more metaphorically visual, including graphics, charts, and diagrams).

#### 5. SPATIAL COGNITION IN ANIMALS

Exploratory behaviors in mammals are reactions to novel settings. This is easily observable in a variety of animals, but most experimental work in psychology, for example, has focused on rodents or primates. Both engage in wayfinding activities and reveal specific patterns of movement-related behaviors. In rodents, exploration usually follows a short freezing phase in which the various senses are used to intake available information and evaluate and look for danger. Exploration follows, and multiple contacts with objects

located in the environment take place via the phenomenon of habituation. This is part of a more general process of knowledge acquisition. In settings previously encountered, reexploration or dishabituation occurs. Thus, novelty in an environment exists normally through reference to stored memories of previously explored or familiar places. The key here is that if novelty is detected and exploration behaviors are triggered, this implies that some representation of the initial situation has been constructed. Such exploratory behavior is assumed by neuropsychologists to be an updating activity of the currently stored cognitive map or spatial representation. When experiencing a new environment, a mismatch between what is stored in the cognitive map or spatial representation and what is observed triggers exploration. In this context, spatial cognitive abilities in animals are often evaluated in a reaction-to-change environment, when repeated exposures to a given spatial configuration are followed by habituation or exploration to fix spatial relationships between environmental objects. If an object is displaced, then the spatial arrangement changes. New exploratory activity provides evidence of a mismatch between the new representation or arrangement and a representation (or cognitive map) of the original situation. Spatial processing and learning in animals are often examined via shortcut and detour tasks (e.g., when a learned path becomes blocked and requires some alternate route to be selected to reach a goal).

Much animal wayfinding activity is said to be governed by the process of path integration. This type of activity has often been observed in the behavior of ants, bees, wasps, or other nesting or home-based organisms. Thus, when foraging behavior occurs and food is found, the organism can return in a direct line or shortcut to the home base. Often called dead reckoning, this action is taken to mean that spatial representations occur across species from insects and other invertebrates to mammals. In insects, the spatial representation may be produced as a retinoptic image (or snapshot) compared to the constructed representations of mammals.

Homing has been observed in a wide variety of species, including ants, spiders, dogs, bees, wasps, and humans, suggesting that it may involve a similar algorithm for processing route-based information across these different species. This appears to be true even when the input data vary among the different tasks. For example, day roving species may use visually detected landmarks to correct minor errors in homing. However, when visual cues are not directly available

(e.g., in night active rodents), when deprived of any directional reference, or when animals are prevented from using the sun's magnetic field (as when a reflective surface is interposed), they still perform homing tasks successfully, implying that they assess distance and direction through inertial signals from the vestibular system, somatosensory feedback, and motor signals that specify speed of locomotion and change of direction. Vestibular signals are presumed to play a key role in the estimation of rotations but not translations. For day foraging rodents, it has been suggested that dead reckoning is mainly a backup system that allows an animal to explore novel or unfamiliar parts of an environment when access to familiar landmarks is not available.

## 6. THE NEUROBIOLOGY OF SPATIAL COGNITION

The study of spatial cognition in cognitive neuroscience focuses on identifying the neural and computational mechanisms associated with spatial activities such as path integration, wayfinding, exploring imaginary environments, or locating invisible targets from memory. Cognitive neuroscientists have investigated the possible existence of place cells for providing location-based object recognition, language production and understanding of location- and direction-based verbal statements, navigation, and orientations in space. Cognitive psychologists suggest that the occipital lobe, the parietal lobes, the dorsal system in general, and the hippocampus are essential to the encoding, storage, processing, and decoding of spatial information.

It has been proposed that spatial information is stored in the hippocampus and neocortex, where spatial learning and memory take place. This hypothesis was popularized by O'Keefe and Nadel in 1978. The hypothesis results from examining the time course of the respective involvement of the hippocampus and the neocortical areas in learning and memory tasks. The argument proposes that spatial memories are initially stored in the hippocampus and become gradually incorporated into the neocortical system as structures or abstract forms. To prevent interference with the previously stored memories in the neocortical system, knowledge structures need to build up slowly, along with experience. The term structure is essentially equated to the term cognitive map. Thus, cognitive mapping can be readily seen to be a process for orienting and for active information-organizing activities in the brain.

The seminal work by O'Keefe and Nadel helped formulate the notion of place cells as the fundamental elements of a hippocampal spatial mapping system. The importance of the hippocampus in spatial thinking and reasoning has been borne out by research on rats with the hippocampus removed and by some research on brain-damaged humans. Given a task that can only be solved using a place strategy, consequent hippocampal dysfunction makes spatial learning impossible. Other research has produced evidence that animals without a functional hippocampus solve spatial tasks predominantly with response or guidance strategies rather than place strategies. Recent research appears to have borne out these predictions, particularly research involving the use of the Morris water maze.

## 7. AGE-RELATED DIFFERENCES IN SPATIAL COGNITION

The study of the development of spatial knowledge structures in children, known as developmental theory, has a long tradition of Piagetian influence. Much less research has been undertaken on spatial cognition in the aged, which has been limited to experiments examining the effect of declining information processing speed on visuospatial abilities. However, with the increasing aging population of the post-World War II baby boomers and with more attention being paid to memory loss (Alzheimer's disease), the behavior of the elderly and concerns about the general quality of life of aged people have become a more popular and pressing applied area for spatial cognition research.

The aged population (generally defined as those people 65 years of age or older) comprises approximately 15% of the total U.S. population and will grow to approximately 20% in the next few decades. A high proportion of this population lives independently, navigates, and drives to work, to go shopping, and for leisure, health, or other needs. There is legitimate concern about gaining information relating to continued spatial competence. In addition, a very high proportion of the aged population has severe vision problems (approximately 80% of the vision-impaired population in the United States is aged 65 or older). This again emphasizes the need to know the state of age-based spatial competencies (skills and abilities) required for safe and effective life in the spatial domain. Research needs to examine questions of how elderly people interact with spatial situations and how individual differences

mitigate the circumstances of such interactions. This requires knowledge about information processing abilities of the aged, the variety and extent of physical abilities present in this group, neurological states, memory loss, and information retrieval and response times. For example, in learning about places, older adults tend to recall significantly fewer landmarks than younger adults, and their memory of most routes (measured in terms of retrace abilities or route sketching abilities) is generally less than that of younger populations. One interesting research finding is that there is an increased importance of landmarks in wayfinding and in spatial cognition generally among the aged. Apparently, elderly people perform wayfinding tasks more effectively if they have prior information about the area to be explored (e.g., by developing a travel plan). However, in general, research shows that route learning ability declines with age. This means that route retracing, landmark recognition, and sequential ordering of landmarks are tasks that are performed more poorly by the elderly than by younger participants.

The differences between aged and younger people are most marked when dealing with unfamiliar environments. In familiar environments, many spatial information-processing tasks are performed quite well by the elderly, and the difference between the aged and the young diminishes significantly. Evidence of spatial cognitive abilities (e.g., distance estimations and judgments of spatial extent) shows systematic age-related distortions consistent with the effort that is perceived to be required to interact physically with the test area.

## 8. SEX-RELATED DIFFERENCES IN SPATIAL COGNITION

Evolutionary psychologists and anthropologists agree that from the earliest times both sexes have been required to be spatially aware and able to act with reflex speed in certain (life-preserving) spatial settings. Sex-based differences in spatial abilities may have evolved over time as a function of a historical division of labor. This "hunting" versus "food gathering and home care" dichotomy is said to be reflected in substantial differences in the diameter of male and female home ranges. Because of the many dangerous situations in which they were involved as hunters, it is argued that males developed an ability to integrate novel incoming spatial information quickly and tended to develop reference frames that extended beyond the visual domain. Females are

assumed to have developed more localized spatial skills associated with foraging for food. These skills involve recognition and recall of landmarks, noticing and storing the details of local environmental settings, understanding configural relations among local environmental features, and developing incidental memory of the location of objects. It is also argued that females are able to better record details of their immediate environment through an unconscious imaging process.

There have been significant differences in the form of testing for sex-based differences in males and females. Some testing has relied exclusively on psychometric spatial ability tests (e.g., the Water Level Test, Primary Mental Abilities Test, Shepard–Metzler Rotation Test, Vandenberg–Kuse Test, and Silverman and Eals Test), which have produced results that appear to indicate that males almost universally outperform females on three-dimensional rotation tests and sometimes on other spatial visualization or surface deployment tests. Contrary results (except for the rotation tests) are often found in the literature, and reports indicate that on some spatial tests (e.g., Silverman and Eals), women tend to outperform males.

It is traditionally believed that spatial abilities are higher in men than in women. As societies and cultures have evolved, as equal opportunity in all facets of social, economic, and recreational life has become more the norm, and as unisex participation in spatially related activities (e.g., sports) has occurred, the gap between male and female scores on spatial ability tests has narrowed. It is now believed that more effacing self-stereotyping by females has played a role in influencing performance on spatial tasks (whether laboratory or field oriented).

Wayfinding research has indicated no significant sex-based differences on tasks such as accuracy of pointing to landmarks, on wayfinding by visitors to an office building, and on wayfinding by the blind. However, males have been found to be more accurate than women with regard to the geometric (coordinate) placement of objects on a map, defining the direction of landmarks or other objects in environmental settings, estimating travel distance, and using cardinal reference points to give directions. Researchers have also found that males make fewer errors and require fewer trials to learn a novel route and perform better (fewer errors) in computer simulations of wayfinding. On the other hand, women have been found to more likely refer to landmarks when giving travel directions and are more accurate in the recall of specific landmarks or objects referenced to landmarks. It has been

suggested that males more readily keep track of current location with respect to distant reference points (implying the use of configurational understanding), whereas women prefer landmark-based, step-by-step instructions (often called piloting) to direct them along a route or to recall it. Studies of the use of technology have indicated a trend toward masculine dominance in the use of wayfinding technologies and that spatially based video games are less popular with females than males. When maze-based computer games are emphasized, males solve the maze problems significantly faster than do females. In general, males outperform females on the Vandenberg Mental Rotation Test, the Shepard–Metzler Rotation Test, the Guilford–Zimmerman Spatial Orientation Test, and the Money Roadmap Test. However, there is evidence that male/female differences are observed more when encoding involves only a single task; where simultaneous encoding of dual tasks is required, sex-based differences are more difficult to observe.

Females apparently notice more details in a given environmental setting than do males. Results from the Silverman and Eals Test support this finding, as do results from other tests that involve recalling the number and location of multiple objects on a computer screen. This applies to both intentional and incidental learning scenarios. Males tend to outperform females on tests of general geographic knowledge. This is supported by evidence that many more males than females make it to the finals of the National Geography Bee, and that male scores on geographically relevant questions on college testing exams are higher on average than female scores.

## 9. SPATIAL COGNITION WITHOUT THE BENEFIT OF SIGHT

Without vision, a person has no visual memories of particular places and has to substitute other senses for direct visual input while experiencing environments. Also, such persons must cope with the fact that fundamental distance, direction, orientation, and reference frame information that is almost automatically provided by vision is absent and must be substituted for. At times, spatial experience comes from hearing; at other times, it comes from touch and motion. Vision provides relatively simultaneous perception of the macrospace—simultaneous with experience. Any objects, their locational positions, and the spatial relations among them can be absorbed in a few glances.

Even when attention wanders, sight continues to operate and incidental information is noticed and encoded. Important cues such as size, dominance of form, color, or specific function are easy to discern with sight and very difficult to discern without it. Sight allows a precise geometry to be used to help define spatial relations, and sight can provide access to information about far more distant places than can be obtained from the other senses.

In small-scale (local body) space, haptic exploration using the hands and arms provides a way to explore object location. The body to which hands and arms are connected provides a stable egocentric frame of reference, even when moved around a tabletop setting, for example. There is evidence that those who have lost their sight later in life and have different periods of sighted experience can more easily comprehend spatial concepts and, consequently, their performance on a variety of spatial tasks is quite similar to those of sighted people.

Evidence of spatial directional knowledge is usually obtained by pointing. Pointing can take place with a wand, a finger, or an arm, or the entire body can be rotated to point in the direction of an object. This process does not require sight and is often used in studies of spatial updating. Spatial updating in the sighted and the adventitiously (late) blind group has been shown to be reasonably accurate in locomotion conditions but less accurate in imagination conditions. In contrast, early (congenitally) blind subjects seem to perform similarly in both conditions.

## 10. NEW FRONTIERS IN SPATIAL COGNITION RESEARCH

Applied areas in which spatial cognition is of increasing importance include wayfinding and navigation, the development of Geographic Information Systems, remote control of robots, teleoperation, location-based services, in-vehicle navigation, and building smart environments—particularly using visualization techniques for presenting relevant spatial information in a decision support system context.

Recent research in spatial cognition has emphasized the impact of alignment effects on spatial recognition and information use, suggesting a major distinction between egocentric heading or “track up” perspective in map displays for in-vehicle navigation systems and the exocentric display of such information in orientation-fixed (“north up”) displays. In general, research

suggests that when there is a matching perceptual alignment between the intrinsic geometrical structure of an environment and the orientation of a representation, it leads to more accurate, more easily accessible mental representation. Such results are impacting how assistive navigation technology is being installed in private and public vehicles for transit and transportation purposes.

The need for communication between people about places became more pervasive as the technological revolution emerged and accelerated. It produced the need to discover the nature of spatial information as communicated in spatial language, to comprehend the value of the processes of spatialization (e.g., graphing or mapping) or making representations (e.g., photos or images), and to understand layouts (such as can be captured by all or any of these representational processes). Interest has been increasing at an exponential rate.

In the future, it may be feasible to conduct more extensive research on relationships between aging and spatial cognition using virtual environments and magnetic resonance measuring techniques for examining the effect of brain cell loss on human spatial performance. Obviously, as the aged group increases in size, the critical question of accessibility in the general spatial–geographic domain will become even more important. Examining the extent of spatial experiences of the aged should be a focus of socially relevant applied psychology in the future.

Spatial cognition has provided underlying algorithms for the incremental generalization of routes that are independently traveled by a mobile robot, which then can be integrated into a spatial representation (or network) of these routes. In robot-related artificial intelligence, a route is defined as a sequence of decision points that are connected by line segments along which a robot can travel. Robot navigation consists of a series of spatial language instructions that, when followed in sequence, allow the robot to select links, make turns at specific angles, and proceed to a destination.

Spatial cognition occupies a place of growing importance in the general area of location-based services and mobile computer applications. This includes design and use of next-generation smart phones, personal digital assistants, wearable computers, and modifiable portable computers such as Palm Pilots. Design of these technologies in such a way as to facilitate use by naive observers has meant an increasing emphasis on the use of concepts and theories related to spatial cognition. In particular, one needs to know what cognitive processes are involved in using them and how to

minimize this cognitive effort to facilitate naive usage. Concerns range from how to present location information to a potential user to how to represent environmental layout information in an interpretable way on the small LCD screens typically built into such portable devices. For example, if maps are to be a primary communicative device, then questions of scale, simplification, generalizability, alignment, and symbology will become important issues.

### See Also the Following Articles

Cognition and Culture ■ Cognitive Aging ■ Motives and Goals

### Further Reading

- Allen, G. L. (1999). Spatial abilities, cognitive maps, and wayfinding: Bases for individual differences in spatial cognition and behavior. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 46–80). Baltimore, MD: Johns Hopkins University Press.
- Burgess, N., Jeffery, K. J., & O'Keefe, J. (Eds.). (1999). *The hippocampal and parietal foundations of spatial cognition*. New York: Oxford University Press.
- Devlin, A. S. (2001). *Mind and maze: Spatial cognition and environmental behavior*. Westport, CT: Praeger.
- Freksa, C., Brauer, W., Habel, C., & Wender, K. F. (Eds.). (2000). *Spatial cognition II: Integrating abstract theories, empirical studies, formal methods, and practical applications*. Berlin; New York: Springer.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gärbling, T., & Golledge, R. G. (Eds.). (1993). *Behavior and environment: Psychological and geographical approaches*. Amsterdam: Elsevier/North Holland.
- Golledge, R. G. (1978). Learning about urban environments. In T. Carlstein, D. Parkes, & N. Thrift (Eds.), *Timing space and spacing time, Volume I: Making sense of time* (pp. 76–98). London: Arnold.
- Golledge, R. G. (Ed.). (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes*. Baltimore, MD: Johns Hopkins University Press.
- Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.
- Kitchin, R., & Freundschuh, S. (Eds.). (2000). *Cognitive mapping: Past, present and future*. London: Routledge.
- Lloyd, R. E. (1997). *Spatial cognition: Geographic environments*. Dordrecht, The Netherlands: Kluwer.
- Montello, D. R. (1997). The perception and cognition of environmental knowledge: Direct sources of information. In S. C. Hirtle, & A. U. Frank (Eds.), *Spatial information theory: A theoretical basis for GIS* (pp. 273–311). Berlin: Springer-Verlag.
- Morris, R. G. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, 12, 239–260.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, UK: Clarendon.
- Passini, R. (1992). *Wayfinding in architecture*. New York: Van Nostrand Reinhold.
- Portugali, J. (Ed.). (1996). *The construction of cognitive maps*. Dordrecht, The Netherlands: Kluwer.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Talmy, L. (2000). *Towards a cognitive semantics* (Vol. 1). Boston: MIT Press.
- Talmy, L. (2003). *Towards a cognitive semantics* (Vol. 2). Boston: MIT Press.



# Special Education

Stephen D. Truscott, Ruth A. W. Berry, and Kelvin Lee

University at Buffalo, State University of New York, Buffalo, New York, USA

1. Introduction
  2. History and Theoretical Foundations: From Institutionalization to Inclusion
  3. Current Special Education Practices
  4. Medical and Technological Progress
  5. Multicultural and Racial Issues in Special Education
  6. International Perspectives
- Further Reading

needs; (c) identifies any special services, testing modifications, and/or equipment needs; and (d) provides annual goals and objectives.

**least restrictive environment (LRE)** Requirement that special education must be provided to the maximum extent possible in or near the general education placement.

**special education** Supplemental or replacement educational services designed to address the educational, social, physical, behavioral, and preoccupational needs of students with recognized educational disabilities.

## GLOSSARY

**continuum of services** Special education services ranging from limited support (e.g., physical therapy, counseling) to full-time, out-of-state residential programs.

**educational disabilities** Any of the 13 disability categories identified by U.S. law as eligible for special education (e.g., learning disability, mental retardation, emotionally disturbed, deaf, other health impaired).

**free and appropriate public education (FAPE)** Education provided as a federal civil right to all students between 3 and 21 years of age regardless of disability; education must be free to the students, must meet identified needs (i.e., potentially beneficial) appropriately, and must be as near to general education as possible.

**general education** Public education provided to all students.

**inclusion** Special education services provided in the general education setting; philosophically, inclusion exceeds least restrictive environment (LRE) requirements.

**individualized educational program (IEP)** Program designed for all students in special education; it (a) documents current levels of performance, learning, social, physical, and behavioral needs; (b) outlines a program to meet these

Special education is a service provided to students with educational disabilities. This article provides a general overview of the history, current practices, and contemporary issues regarding special education. It focuses on special education practices in the United States, although other developed nations implement comparable practices and share similar concerns. It also provides a brief review of international issues in special education.

## 1. INTRODUCTION

In 1896, Lightner Witmer opened one of the first psychological clinics in the United States. His first client was a child referred for spelling difficulties by a public school teacher. In 1907, Witmer wrote that the clinic served children who were "conspicuous because of an inability to progress in school work . . . or [who had] . . . moral defects which rendered them difficult to manage." The

partnership between applied psychology and special education began early and continues to this day. Over the years, psychologists have often led efforts to find effective educational, behavioral, and psychoeducational interventions to address the needs of students with disabilities. Most recently, 5 of the 19 members of the 2002 President's Commission on Excellence in Special Education (PCESE) were psychologists. This article provides a general overview of the history, current practices, and contemporary issues regarding special education and its relationship with applied psychology.

## **2. HISTORY AND THEORETICAL FOUNDATIONS: FROM INSTITUTIONALIZATION TO INCLUSION**

People with disabilities have always been part of society. What has changed is how society defines and explains these differences and how people with special needs are treated. Early explanations of disabilities were grounded in superstitious belief systems, and treatments included abandonment and extermination. By the 1800s, these gave way to quasi-scientific explanations and institutionalization of people with disabilities. Early efforts to educate children with disabilities were often conducted by physicians (e.g., Itard, Seguin, Howe, Gallaudet) whose work established the foundation for current practices such as individualized instruction and an emphasis on functional skills. Because societal demands were different (e.g., for literacy), early efforts to provide special education were focused on people with severe disabilities such as mental retardation and psychosis. During this period, asylums for the treatment of individuals with disabilities began to appear. Such institutions were prevalent in the United States until the relatively recent past and still exist widely in many other countries. Although asylums were originally intended to protect people with disabilities from cruel treatment in the outside world, patients were sometimes abused and neglected.

### **2.1. The Child Study Clinic Movement**

Witmer's psychological clinic was part of a movement that featured the application of scientific methods to problems of living, an emphasis on public health, and attention to social justice. At the turn of the 20th

century, such "child study clinics," often associated with public schools, sprouted up in metropolitan areas throughout the United States. These clinics focused on using the new science of psychology to address the problems of children. The clinics served students with a range of disabilities, including mental retardation, learning problems, and behavior disorders. In addition to prescribing treatments and educational interventions, these pioneering applied psychologists viewed schools as places to provide things such as nutrition, parent training, and healthy environments when these were missing at home. It is no coincidence that child study clinics were established at approximately the same time as the rapid expansion of public education and the implementation of child labor laws.

The clinics focused on applying the new science of psychology to the needs of individual children, so it is not surprising that the Binet-Simon test (translated in 1910) found widespread acceptance among early clinicians. The new test provided an objective and scientific way in which to classify students and played a significant role in the rise of special programs for students with disabilities in metropolitan areas served by clinics. During the 1920s, special programs for students with disabilities (notably mental retardation) were widespread in cities but nearly absent in rural America. The Binet test, administered by clinicians in child study clinics, played a critical role in these placement decisions.

### **2.2. The Rise of the Medical Model and the "Within Child" Deficit Approach to Special Education**

Throughout the 20th century, psychological and medical technologies were increasingly evident in special education diagnoses and programming. "Within child" and biological explanations of disabilities (e.g., minimal brain dysfunction, dyslexia, Down syndrome) became more and more dominant after 1950. Following the medical model, disabilities were thought to originate within the child and were most often attributed to manifestations of underlying biological problems. Special education treatment followed along with these explanations, and the idea of diagnostic prescriptive teaching came to the forefront. This model is predicated on the idea that particular diagnoses can be addressed best by prescribing specific treatments. In special education, the individualized educational program (IEP) exemplifies this approach.

Concurrently, after the 1954 *Brown v. Board of Education* U.S. Supreme Court decision ruled against

public school segregation by race, a series of court cases and legislation provided U.S. constitutional rights (e.g., equal protection) to people with disabilities. Special education rights grew rapidly, culminating in 1975 with the enactment of Public Law (PL) 94-142, a federal law that granted access to a free and appropriate public education (FAPE) to all students regardless of disability. The rapid expansion of special education services after PL 94-142 has continued to date.

### **2.3. Current Conceptualization: Behavioral, Ecological, and Sociocultural Models**

Currently, behavioral, sociocultural, and ecological approaches mark a fundamental transition in special education in the United States. In contrast to the medical model, these contemporary approaches attribute the manifestation of a “disability” to the transaction between the demands of the environment and the behavior of the individual rather than to the expressions of internal biological deficits. Fueling the transition is the understanding that the widespread goal of recognition and inclusion of people with disabilities in the educational system has been reached but that less attention has been focused on the learning, social, behavioral, and occupational outcomes of special education placement. In addition, considerable growth has occurred in the percentage of students classified in the mild or “judgmental” categories of disability (i.e., those disabilities that are determined primarily by results of psychometric testing such as learning disability [LD], speech/language impaired, and emotionally disturbed). Such students now comprise 85 to 90% of the special education population in the United States. The percentage of students in disability categories with a known etiology or physical problem has remained much more stable over time.

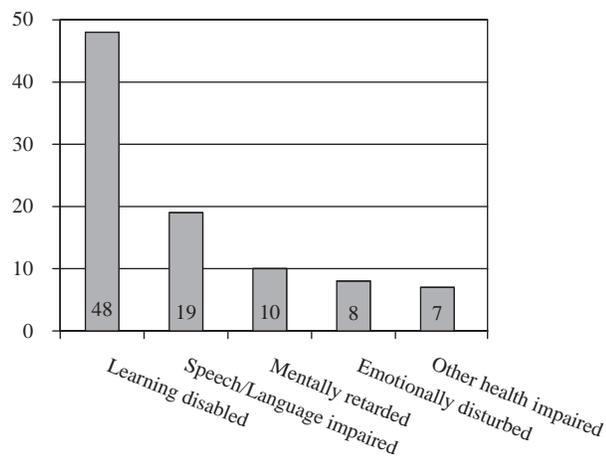
Indeed, in the United States today, most students in special education have neither clearly defined disabilities nor known etiologies. Wide differences among the characteristics of students within the same disability classification in various settings support the idea that most educational disabilities are primarily social constructions. For example, in 1994, Gottlieb and colleagues reported that the average intelligence quotient (IQ) of students labeled as LD in urban schools was 81, compared with an average of 102 for students labeled as LD in suburban communities. Examination of such differences has resulted in the serious ongoing reexamination of special education definitions and

services exemplified by the PCESE’s finding in 2002 that many students in special education are essentially “instructional casualties” rather than students with educational disabilities. As programs based on the medical model (e.g., neurological models of LD) have failed to result in acceptable outcomes for the majority of students in special education, new interest in functional assessment of behavior, direct assessment of academic skills (e.g., curriculum-based measurement), attention to observable intervention outcomes, and direct instruction of academic, social, and behavioral skills has become an important element of special education reform in the United States. These practices, which are derived from behavioral psychology and ecological developmental theory, are based on the idea that the learning problems experienced by most children labeled as educationally disabled arise from a mismatch between the learning needs of the students and the instructional or other environments.

## **3. CURRENT SPECIAL EDUCATION PRACTICES**

### **3.1. Classification and Labeling**

In the United States, approximately 13% of public school students were enrolled in special education programs during the 2001–2002 school year. The largest disability group, nearly half of the total, is composed of students identified as having specific learning disabilities (SLDs) (Fig. 1). The next largest subgroup is composed of students with speech or language impairments. Although the vast majority of students in special education have mild difficulties, educational disabilities can range in severity from mild to profound and can range in scope from narrow to broadly pervasive. For example, a learning disability may affect only a student’s ability to complete mathematical calculations, whereas a child labeled as multiply disabled may have severe cognitive, physical, social, and behavioral difficulties. Both of these students may require special education, yet the scope of the disability, required interventions, and educational goals for these two children would be quite different. Educational disabilities are caused by a range of factors that can be located on a continuum from environmental deficiencies (e.g., lack of effective reading instruction) to within child physical factors (e.g., genetic defects). In general, the more severe the disability, the more likely it is that the etiology can be identified.



**FIGURE 1** Approximate percentages of U.S. special education students in the 5 major disability categories (2002 data). Approximately 92% of all U.S. students in special education are classified in these 5 of the 13 educational disability categories.

In the United States, definitions of educational disabilities are specific to special education and are separate from “disabilities” or diagnoses used in other legislation (e.g., Americans with Disabilities Act) or systems (e.g., *Diagnostic and Statistical Manual of Mental Disorders*). Special education definitions require two primary components. First, it must be determined that a student has a need for special education. That is, the suspected disability must negatively affect the child’s education. Second, the suspected disability must be in one of the following general categories: (a) sensory disabilities such as visual impairments, hearing impairments, and deafness–blindness; (b) physical and neurological disabilities such as orthopedic impairments, other health impairments, traumatic brain injury, multiple disabilities, and autism; and (c) developmental disabilities such as SLD, speech or language impairments, emotional disturbance, mental retardation, and developmental delay. Children up to 9 years of age can be identified as having a developmental delay, reserving the identification of specific disabilities until they are older and the disability characteristics become more distinct.

### 3.2. Identification Processes

In most U.S. schools, determining a student’s need for special education begins with a teacher referral and prereferral intervention. Prereferral intervention most

often involves a team of administrators, specialists, and teachers that examines the student’s educational difficulties and recommends interventions or services to address the student’s needs in the general education setting prior to special education referral. Ideally, such a team implements a problem-solving approach by collecting relevant data, recommending targeted general education interventions or services, and monitoring the results of these recommendations. Over time, if the team concludes that the child has not benefited sufficiently from the recommended interventions, the child is referred formally for a special education evaluation.

The formal determination of an educational disability is made by a multidisciplinary team composed of parents, general and special educators, an educational administrator, an individual with expertise in interpreting assessments (usually a school psychologist), other service providers, others with special knowledge of the child, and the child (if appropriate). Parental consent is required at several points in the process, including at the initiation of the formal evaluation process. After the child is determined to have an educational disability, an IEP is developed. Once the IEP is implemented, the student’s progress must be reviewed annually. Every 3 years, the student’s continued eligibility for special education services is reevaluated.

Identification processes vary according to disability category. Sensory, physical, and neurological disabilities typically involve medical diagnoses, whereas developmental disabilities rely on multiple psychoeducational assessments to identify traits of the specific disabilities. Psychoeducational assessments are a function of schools and typically involve an individually administered IQ test, an educational achievement test, behavior rating scales, personality assessment, psychological processing tests (e.g., memory), social and developmental histories, and classroom observation. Assessment is conducted both to identify potential eligibility criteria and to rule out potential exclusionary factors (e.g., environmental and physical causes). For example, a student who is frequently absent from school might not be eligible for special education despite poor academic achievement unless it is also determined that the child has an educational disability over and above the problems associated with infrequent attendance. Most often, a school psychologist administers and interprets the psychological assessments. Frequently, a school social worker, occupational therapist, speech/language pathologist, and special education teacher are also involved in the assessment process.

Establishing reliable and valid criteria for the identification of learning disabilities, the largest special education category, has been the single most controversial issue in contemporary special education. The central criterion of the current definition of LD, an aptitude–achievement discrepancy, is particularly controversial because it has not proved to be a reliable way by which to determine the presence or absence of learning disabilities, distinguish students with learning disabilities from other low-performing students, or predict who will benefit from remediation. Furthermore, although the importance of early intervention is well established, the practical effect of using discrepancy formulas is that students must wait to fail because an aptitude–achievement gap requires time to develop. For these reasons, researchers and members of the PCESE have called for the development and use of responsiveness-to-instruction or responsiveness-to-intervention (RTI) models for identifying students with learning disabilities. RTI models avoid the “wait to fail” problems associated with current special education identification by providing progressively more specialized instruction to all students performing below local achievement norms prior to consideration for special education eligibility. Many students will respond to this increasingly intense instruction and can leave the RTI program. Those who do not respond to these multiple levels of intervention are subsequently considered for special education.

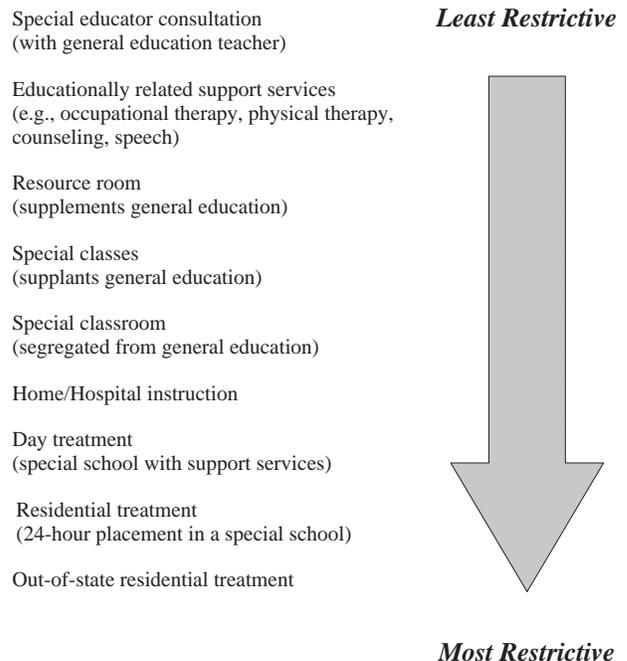
### 3.3. Placement and the Continuum of Services

Until recently, students with educational disabilities have been served primarily by special education teachers using special materials and instruction delivered in classes and programs designed for specific disabilities (e.g., all students with mental retardation receive instruction in a special classroom designed for students labeled as mentally retarded). Two recent trends have altered conceptualizations about how placement decisions are made and where special education services are provided. First, researchers have not found consistent or compelling evidence that placement is the decisive factor in students’ academic progress, particularly for students with mild or moderate disabilities. Second, many parents and educators believe that inclusion in the general education setting is the preferred context for special education services.

In the United States, contemporary special education is best conceptualized as services provided to students

regardless of location rather than as placement in specific programs (Fig. 2). Special education services are selected to meet student needs identified on the IEP. Although there are many variations, the large majority of students receive some combination of the following: (a) a resource room provided to supplement general education instruction, usually for a limited time of time (e.g., one class period) each day; (b) a special class provided to replace general education instruction for anything ranging from a single subject (e.g., reading) to the full day; and (c) a general education class with special education support ranging from consultation to in-class instruction by special educators. In addition to instructional services, many students in special education receive related services such as counseling, physical therapy, occupational therapy, and speech/language therapy.

Inclusion, the practice of educating children with disabilities in general education contexts, has a history of controversy. Early disputes were often ideologically based, positioning those who saw dedicated special education contexts as segregated environments with negative connotations against those who feared that students’ individual needs would remain unmet in the general education context and that some special education students would disrupt general education learning environments. In most places, the inclusion



**FIGURE 2** Typical continuum of special education services.

debate, benefiting from increasing experience and research, has moved to considerations of how to be successful in including all students in general education classrooms. Questions relating to providing special education services in general education contexts now address issues of teacher attitudes, training, and collaboration; supports and accommodations for students with disabilities, including materials and technology; effect of inclusion on other students (e.g., disruptive behavior, changes in curriculum or instructional materials); and costs (e.g., smaller class sizes, increased planning time for teachers, technology).

Most special educators now believe that placement decisions should be made on an individual basis rather than on a categorical basis because different educational contexts may address some learning and management needs better than do others. Where socialization or learning subject content information is the primary goal, a general education context might be most appropriate. Where a student's learning rate or level is highly unlike that of peers, a separate setting providing small group intensive instruction might be best. In any setting, effective teaching and consistently implemented instructional plans are critical for student achievement. Collaborative consultation between general educators and special educators, cooperative teaching arrangements, cooperative learning, peer-mediated instruction, peer tutoring, and direct instruction of skills all have been identified as effective educational strategies for children with disabilities in inclusion contexts. However, even when most instruction occurs in a special education setting, some degree of general education participation enables students with disabilities to be exposed to a wide range of academic content and to interact socially with nondisabled peers (e.g., having students with disabilities participate in all classroom activities but on a reduced basis with appropriate accommodations and modifications).

### **3.4. Instructional Strategies/ Effective Interventions**

Many instructional strategies are available to meet the needs of students who require special education services. Some are effective for all students and can be used successfully in general education classrooms, whereas others have been developed to address unique learning needs of students with disabilities and are not likely to be used to address general education needs. In general, special education interventions fall into one of

two broad categories: (a) interventions designed to remediate the disability or (b) interventions designed to compensate for the disability. For example, a student with a reading disability can (a) receive extra or different instruction in reading or (b) be provided with audio recordings of books. Both approaches have support, depending on the goal of the instruction. The following subsections briefly describe effective educational interventions for students with (a) sensory disabilities, (b) physical and neurological disabilities, and (c) developmental disabilities.

#### **3.4.1. Sensory Disabilities**

This category includes children with labels such as visually impaired, hearing impaired, and deaf/blind. Degrees of sensory impairment range from mild to profound. Thus, educational interventions include diverse approaches from adjusting classroom conditions (e.g., reducing environmental noise for students with impaired hearing) to providing specialized instruction and remediating secondary effects (e.g., teaching vocabulary and idioms to children who are deaf). Visual impairment refers to people who cannot see well even with correction. Low-vision devices and technology are used to support academic learning for some children with this disability. Educational interventions for students with more severe vision loss may include orientation and mobility skills. In general, students with sensory disabilities need to compensate for their diminished sensory input and might need to develop communication, vocational, self-help, advocacy, and recreational skills.

#### **3.4.2. Physical and Neurological Disabilities**

This category includes children with orthopedic impairments, other health impairments, traumatic brain injury, multiple disabilities, and autism. Students in this category have widely diverse educational needs. Some might lack the ability to walk or talk, whereas others might have a "hidden" disability such as a brain injury or another health condition (e.g., diabetes, asthma). These students might experience limitations due to the direct results of the disability or its effects (e.g., side effects of prescribed medications). In addition to modifications and adaptations in the learning context, teachers might need to monitor the physical or health conditions of these students. Students with physical and neurological disabilities, as well as those with sensory disabilities, often benefit from assistive technology to achieve academic success.

Historically, educational interventions for students with severe physical or cognitive disabilities have focused on social and functional skills required for survival in general society (e.g., communication, motor skills, decision making, functional reading, math) across multiple life domains, including vocational, leisure, home, and community settings. Instruction often focuses initially on discrete skills that are then linked to form complex behaviors. For example, a child might first learn to hold a toothbrush, then apply toothpaste, and then raise it to his or her mouth as part of learning the complex behavior of brushing one's teeth. Typical teaching strategies include verbal and physical prompting with eventual fading of prompts, praise for correct responses, instructive error correction, augmentative and alternative communication, and numerous opportunities to practice the new skill across multiple contexts.

### **3.4.3. Developmental Disabilities**

Learning disabilities, speech and language disorders, and emotional and behavioral disorders are the three largest groups in this category. Learning disabilities are defined in the Individuals with Disabilities Education Act (IDEA) as “disorders in one or more of the basic psychological processes” that are evident in an “imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations.” Some intervention strategies seek to improve academic outcomes by targeting underlying deficits in basic psychological processes. For example, a student may receive training in visual processing. However, to date, such indirect approaches have demonstrated little or no effect on academic achievement. Although it is possible that such approaches have merit, further development of accurate measures of basic psychological processes, as well as interventions that are designed to address these issues, will be required before such instruction is likely to be effective. In addition, although the modality-matched approach (e.g., programs for “visual learners”) and the multisensory approach (e.g., using visual, auditory, and/or kinesthetic cues in instruction) are popular with many teachers, they have not been supported empirically.

In contrast, explicit, intensive, and supportive academic interventions have received substantive empirical support as successful methods for addressing academic failure. Effective instructional approaches include controlling task difficulty, using small interactive groups, teaching metacognitive strategies and problem-solving skills, using direct instruction, improving reading skills (e.g., phonemic awareness) and writing skills (e.g.,

organization and mechanics), and providing ongoing and systematic feedback. Such approaches improve academic achievement with students in both inclusion and special education classrooms, with students in a variety of disability categories, and with at-risk general education students.

Students with speech and language disorders, often referred to as communication disorders, may exhibit deficits in any of the following areas: receptive and expressive language, language development, articulation, fluency, and voice features. Speech and language pathologists provide a range of services in either private or classroom contexts to remediate these difficulties. Instruction may be targeted solely at learning the physical movements required for clear articulation (speech) or at developing language-related elements required for communication such as vocabulary and grammar.

Students with emotional or behavioral disorders may display inappropriately excessive or deficient behavior, possess unsatisfactory interpersonal relationships, and demonstrate academic learning problems. Intervention approaches used with these students include various psychotherapies, behavioral contingencies, family therapies, and segregation in residential programs. Behavioral and cognitive-behavioral approaches that include examining and altering antecedents and consequences associated with inappropriate behavior (e.g., functional behavioral assessments [FBAs]) and training in social skills have been shown to be effective in decreasing inappropriate behavior and increasing appropriate behavior. Because behavioral issues are central for most students labeled as emotionally disturbed, less research has been directed to identifying academic (vs behavioral) interventions specifically for students with emotional or behavioral disorders. However, instructional strategies that promote academic engagement and responding along with self-monitoring procedures have been shown to be effective. Because interventions for these students often require highly structured environments and more intensive instruction than is available in general education, this group of students is likely to receive direct services in a special classroom or school.

Students with emotional and behavioral disorders often become involved with the school disciplinary process. This can result in a conflict between the legitimate disciplinary needs of the school and the student's right to FAPE, particularly when the relevant behavior is severe or dangerous. In such cases, the school must conduct a “manifestation determination” to determine whether the behavior is a manifestation of the child's behavioral disability. If it is, then the child should not

be punished for manifestations of the disability. Instead, an FBA should be conducted to determine the situational and behavioral factors underlying the misconduct, and the student's IEP and/or placement should be changed to better meet the identified behavior management needs. If the misconduct is not a function of the student's disability, the student is subject to the normal disciplinary procedures.

### **3.5. Early Intervention and Adult Transition Services**

Many educational disabilities can be addressed most effectively before children reach school age. Other disabilities are so severe that ongoing services are required beyond 21 years of age. U.S. federal legislation requires that states and local educational agencies provide special education and planning services to children who need intervention before 3 years of age and provide transition planning services to children who need adult services after 21 years of age. Such services provide early family-based interventions for young children and a planned transition to adult services for students who require support beyond the years covered by FAPE. Most often, early intervention and adult transition services are provided to children with moderate to severe sensory, physical/neurological, and developmental disabilities.

#### **3.5.1. Early Intervention**

Early identification of disabilities and subsequent intervention are crucial for enhancing development and learning, preventing the development of additional problems, and helping families adjust to the needs of children with disabilities. Early intervention is mandated by IDEA for very young children, from birth through 2 years of age, and refers to a comprehensive array of services that include educational interventions, health care, and social services (e.g., family counseling, parent classes) for families as well as for the affected children. A focus on the family and home environment is intended to (a) equip parents to participate in all aspects of decision making and administer interventions for their children and (b) locate instruction involving movement, language, self-care, and cognition skills in natural settings. To plan for and document interventions, an individualized family service plan (IFSP) is developed. The IFSP is similar to an IEP for school-age children, but the IFSP provides for more holistic goals, includes family

needs in the intervention plan, and designs instruction around naturally occurring routines (e.g., bathing).

#### **3.5.2. Transition to Adult Services**

Transition to postschool settings and services is an important service provided by schools for the many students with disabilities who require supportive services after public schooling ends (i.e., 21 years of age). In the United States, transition planning begins a few years before a student's public education ends and continues until the transition is actually made. The transition planning process, which addresses academic, occupational, living, and social skill needs, is best conceptualized as an approach built on abilities, options, and self-determination rather than as a disability-focused process. Vocational education, work experience, tutoring, extracurricular group activities, parental support, and collaborations with community businesses, organizations, and agencies are important elements to address in successful transition planning.

## **4. MEDICAL AND TECHNOLOGICAL PROGRESS**

Current and emerging information and medical technologies provide considerable hope for intervention for some disabilities and create considerable concern about ethical implications for others. Schools are required to provide assistive technology for students who need such devices to participate in and benefit from FAPE. Assistive technology refers to devices and services that enhance the performance of individuals with disabilities by enabling them to complete tasks more effectively, efficiently, and independently than would otherwise be possible. Assistive technology is used principally for aiding persons with physical and sensory impairments and includes items such as alternative keyboards, switches, Braille printers, mobility devices, adaptive feeding instruments, voice synthesizers to aid speech, and eye-controlled devices to help people without the use of their hands to accomplish everyday tasks. Assistive technology can also be used for students with high-incidence disabilities who require technologies such as speech recognition software to help compensate for writing difficulties, pagers or personal digital assistants to help with organizational skills, computers that translate text to speech, and calculators to assist with math learning.

Other technologies offer the promise of substantially augmenting current skills and providing alternative communication strategies to people with disabilities that have not been treatable previously. For example, electronic cochlear implants allow some deaf children to hear, and emerging electronic vision technologies may soon allow blind people to see. Some new medical technologies are controversial. For example, widespread use of cochlear implants could result in the disappearance of the vibrant and rich deaf culture that has its own language, customs, and values. Widespread implementation of genetic counseling and other medical techniques to screen for disabilities may allow for the possibility of correcting some disabilities before birth (in utero) but also have ethical implications ranging from whether parents will choose to have less than “perfect” children to judgments about the value of people with disabilities as contributing members of society.

## **5. MULTICULTURAL AND RACIAL ISSUES IN SPECIAL EDUCATION**

Special education in the United States is grounded in civil rights legislation. The current system and legislation were constructed to ensure that all students have equal access to FAPE regardless of disability. However, issues of minority overrepresentation in special education arose almost immediately after the widespread implementation of special education services. Some advocates have seen special education as an inferior educational track that has been used to isolate and segregate students from minority cultures in substandard classrooms. Two recent trends make the already critical concerns about minority (racial and cultural) overrepresentation in special education even more imperative. First, the demographic profile of the United States is changing rapidly. By some predictions, the public school population will be more than 50% non-White by 2050, as it already is in six states. Second, academic achievement for students of color has been identified as a national concern. In general, public education in the United States has not served students from minority cultures well, particularly when such children are poor and live in urban and rural communities. However, the basic problem of poor school achievement for children from minority cultures is only tangentially related to special education. Poverty and deficient educational programs (general education and special education) in many urban and rural

schools are issues for students from minority cultures primarily because these students are disproportionately represented in low-income groups. Placing large numbers of non-White students in special education cannot possibly address the fundamental issues of poor achievement. Interestingly, the issues of minority overrepresentation in special education are limited primarily to the judgmental educational disability categories. The incidence of severe disabilities is relatively stable across socioeconomic groups and cultures.

## **6. INTERNATIONAL PERSPECTIVES**

Although special education is established in most developed nations, acceptance of people with disabilities and awareness of special education needs vary widely in other parts of the world. In 2001, the United Nations Education and Scientific Cultural Organization (UNESCO) estimated that 98% of children with disabilities in developing countries do not attend schools. Obviously, special education is not a priority in countries that do not provide general public education to their citizens. However, intergovernmental organizations such as UNESCO have advocated for education for all children. In 1994, the Salamanca Statement and Framework for Action on Special Needs Education was signed by delegates from 92 governments and 25 international organizations. The statement acknowledged that children with disabilities should be included in regular schools and that schools should make changes to “accommodate [the children] within a child-centered pedagogy capable of meeting [their] needs.” As a result, some efforts to include children with disabilities in regular schools occurred worldwide.

Still, the provision of special education varies widely between and within nations. Countries such as Britain, Denmark, Ecuador, India, South Africa, and China have adopted changes in their education systems to include children with disabilities into regular classrooms. Although educational systems in some developing countries are moving toward inclusion, inclusive education is still viewed a luxury of First World countries. Currently, special schools in developing countries (e.g., China) are often affiliated with religious or private philanthropic organizations. In such cases, education for students with special needs depends on the availability of private services, the ability to access such services, and knowledge of their existence. However, in many rural areas of developing countries, livelihood is still dependent on manual labor. In these cases, literacy is not considered

to be essential and children who might be classified as having mild or moderate disabilities in other settings can lead normal lives and might not be diagnosed.

### See Also the Following Articles

Behavioral Assessment in Schools ■ Learning ■ Learning Disabilities ■ Learning Styles and Approaches to Studying

### Further Reading

- Cook, B. G., & Schirmer, B. R. (Eds.). (2003). What is special about special education? [special issue]. *Journal of Special Education, 37*(3).
- Friend, M., & Bursuck, W. D. (2002). *Including students with special needs: A practical guide for classroom teachers* (3rd ed.). Boston: Allyn & Bacon.
- Hallahan, D. P., & Kauffman, J. M. (2003). *Exceptional learners: Introduction to special education* (9th ed.). Boston: Allyn & Bacon.
- Kohler, P. D., & Field, S. (2003). Transition-focused education: Foundation for the future. *Journal of Special Education, 37*, 174–183.
- MacMillan, D. L., & Reschly, D. J. (1998). Overrepresentation of minority students: The case for greater specificity or reconsideration of the variables examined. *Journal of Special Education, 32*, 15–24.
- Odom, S. L., & Woolery, M. (2003). A unified theory of practice in early intervention/early childhood special education. *Journal of Special Education, 37*, 164–173.
- President's Commission on Excellence in Special Education. (2002). *New era: Revitalizing special education for children and their families*. Washington DC: U.S. Department of Education.
- Vaughn, S., & Fuchs, L. S. (Eds.). (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems [special issue]. *Learning Disabilities Research & Practice, 18*(3).
- United Nations Education and Scientific Cultural Organization. (2001). *The flagship on education for all and the right to education for persons with disabilities: Towards inclusion*. [www.unesco.org](http://www.unesco.org).
- U.S. Congress. (1997). *Individuals with Disabilities Education Act amendments of 1997*. Washington, DC: U.S. Government Printing Office.
- Witmer, L. (1907). Clinical psychology. *The Psychological Clinic, 1*, 1–9.



# Sport Psychology, Overview

Yuri L. Hanin

KIHU – Research Institute for Olympic Sports, Jyväskylä, Finland

Natalia B. Stambulova

Halmstad University, Halmstad, Sweden

1. Introduction
  2. Competitive Sport as a High Achievement Setting
  3. Performance Enhancement
  4. Athletic Excellence from a Developmental Perspective
  5. Conclusion: Future Directions in Sport Psychology
- Further Reading

Sport psychology is a subdiscipline of psychology applied to a competitive sport as a specific context of organized physical (motor) activity. Competitive sport is focused on high achievement and consistent excellence, in contrast to other settings in which exercise is used for physical education, leisure, or rehabilitation. The major emphasis in sport psychology is on the study and application of psychological factors enhancing athletic performance and on the impact of sport participation on a person's (or team's) development.

## GLOSSARY

**athletic excellence** An athlete's (or team's) exceptionally good performance process and outcomes compared with the previously achieved standards (self-referenced or normative). Athletic excellence reflects an athlete's (or team's) ability to perform consistently up to his or her potential by recruiting and utilizing effectively available resources to match the demands of a task at hand.

**experiences** The totality of past and present characteristics that makes up the particular quality of a person's performance. Performance-related experiences are reflected in situational states, relatively stable patterns (traits, qualities), and meta-experience (self-knowledge, attitudes, preferences/rejections of one's experiences).

**resources** Psychobiosocial (cognitive, affective, motivational, bodily, behavioral, operational, communicative) assets determining an athlete's ability to performance consistently up to his or her potential.

**sport psychology** A subdiscipline of psychology that deals with "psychological foundations, processes, and consequences of the psychological regulation of sport-related activities of one or several persons acting as the subject(s) of the activity" (FEPSAC Position Stand, 1995).

## 1. INTRODUCTION

### 1.1. Defining Sport Psychology

What is sport psychology? How is it different from other subdisciplines of psychology? How is it related to sport sciences? Although many definitions of sport psychology have been suggested, there has been no comprehensive and internationally accepted definition of sport psychology. In its Position Stand # 1 (1995), the European Federation of Sport Psychology (FEPSAC) proposed that "sport psychology is concerned with the psychological foundations, processes, and consequences of the psychological regulation of sport-related activities of one or several persons acting as the subject(s) of the activity" (p. 4). This definition indicates that sport psychology attempts to improve athletic performance and help athletes to concentrate better, deal effectively with competitive stress, and to

practice more efficiently. Moreover, sport psychology also attempts to understand the impact of long-term sport participation on development of personal resources of athletes in the setting of organized competitive sport. The term “sport” is used as an umbrella term that includes different kinds of sport, exercise, and other physically active pursuits. These types of physical activity are also used in other settings such as organized physical education, leisure, and rehabilitation (healing). Another important feature of sport psychology is its double nature. On the one hand, it is a part of psychology; on the other hand, its knowledge base is related to sport sciences focused on understanding human activity in this particular context. Thus, in applications, these two sources of knowledge help to better understand a person, the environment, and the key aspects of the sporting activity.

The major focus of this article is the context of competitive (high-achievement) sport. From this perspective, sport psychology examines mainly the short- and long-term impact of psychological factors on athletic performance and the potential effects of systematic participation (involvement) in sport. Applied sport psychology attempts to solve specific practical problems by improving athletic performance and thus helping athletes to develop their potential in the sport setting.

This article briefly reviews selected aspects of applied sport psychology within the framework of three basic constructs: athletic excellence, performance-related subjective experiences, and individual resources (psychological strengths). The key aspects of athletic performance are examined from the short-term (readiness for competition and performance excellence) and long-term (consistent excellence, career development) perspectives.

## 1.2. Major Focus and Trends in Sport Psychology

What are the major focuses in sport psychology research? What are the main trends in applied psychological work with athletes, teams, and coaches? Noteworthy are two major focuses in sport psychology research, with two corresponding trends in applied work. The first is understanding the psychological factors that affect athletic performance and how athletes realize their potential in sport. Applied aspects here include high-quality practices, optimal performance, and adequate recovery at the level of an

individual athlete and team. The second important objective of sport psychology is to understand how athletes develop in sport and what are the “benefits” and “costs” of their multiyear sport participation. Applied aspects here include the need to help athletes cope successfully with career transitions and find a balance between sport and other spheres of life. In team sports, this also involves dealing with team-building issues and helping individual athletes find a balance between individual and team interests and values.

In competitive sport, applied psychologists deal with healthy, motivated, and high achievement-oriented people striving for consistent excellence, performance up to their potential, and continuous self-development. Thus, the focus on enhancement of athletic performance and empowering approach reflect a positive, proactive, and constructive nature of applied sport psychology. Interestingly, Seligman and Csikszentmihalyi (2000) called upon applied psychologists to move beyond studying psychological disorders and problems and spend greater efforts studying positive psychology that can be used to facilitate and enhance human functioning. This emphasis on positive psychology, or psychology of human resources and strengths, is not new, but has been occurring for the last 25 years. However, there is still an urgent need to attend to current concerns of athletes and coaches and examine more closely their successful experiences by bridging the gap between group-oriented and individualized approaches. Therefore, it is argued that sport psychology is the psychology of personal and athletic excellence and as such, from the very beginning was oriented to identifying a person’s resources (strengths) to facilitate consistently successful performance up to the person’s potential.

## 1.3. How Sport Psychologists Work

Who are applied sport psychologists? What do they do, and how and why do they work with athletes, teams, and coaches? These questions are important for an understanding of what sport psychologists can and can not do in the field of competitive sport.

First, sport psychologists as a professional group are experts with different backgrounds. They may be clinically oriented consultants, educationally oriented consultants, mental trainers, applied researchers specializing in performance enhancement, or social or personality psychologists. However, whatever their specialization, applied sport psychologists are usually required to be well versed not only in psychology

but also in sport and sport sciences. This helps them to establish and develop working relationships with individual athletes, teams, coaches, parents, managers, etc.

Second, it is well known that the science of coaching focuses on the use of general principles. Per Weinberg and Gould (1999), “the art of coaching is recognizing when and how to individualize these general principles” (p. 15). As with coaching, the practice of applied sport psychology is both a science and an art. As a science, it is based on various theoretical models and results of empirical studies describing what is typical for athletes in particular sport situations. As an art, sport psychology is grounded in the personality as well as personal and professional experiences of the consultant, and it is expressed in his or her ability to understand the particular athlete within a psychological context and to choose the most effective applied approach or intervention. That is why different consultants may work differently with the same athlete yet be equally successful.

Art and science aspects are sport psychologists’ tools to help athletes and coaches, who often focus mainly on the symptoms or consequences of psychological problems, deal with real causes of the problems (challenges, task demands).

Third, there are certain organizational working models, assessment technologies, and interventions based on specific ethical norms that characterize how sport psychologists work. For instance, sport psychology research and effective delivery of psychological services to elite athletes and coaches usually focuses on two closely related aspects: (1) performance enhancement in practices and competitions, and (2) optimization of interpersonal and intragroup communication, creating optimal team climate and effective management. Sport psychologists use several guidelines or principles to enhance their work, including action and growth-orientation; an emphasis on developing individualized strengths rather than on repairs of deficiencies; empowering athletes, coaches, and teams rather than developing over-dependency on outside experts; and enhancing active participation, partnership, and cooperation between sport psychologists, athletes, and coaches.

In brief, working with an elite athlete or coach usually includes several action-oriented steps: (1) listening to the coach and athlete’s account of the current situation and past performance history to identify their concerns that need to be addressed; (2) providing a general summary of how similar situations are usually handled in sport and suggesting a tentative plan of

joint work on the problem at hand; (3) collecting the data and providing a detailed feedback with the interpretation of results using context-related language clear to the athlete and coach; (4) preparing an action plan for further analysis, change, and monitoring of the key parameters involved; (5) evaluating the effectiveness of the initial steps and developing an individualized intervention program with clear criteria to assess the athlete’s progress on a daily, weekly, monthly, or seasonal basis; and (6) contacting (by phone, e-mail, or fax) the athlete and coach systematically, which is an important part of their work during the entire season. A wrap-up, “lessons learned” session is also a good way to summarize experiences of all participants by the end of the season. It is important to realize that this approach is different from the traditional role of an outside expert telling the client what to do or not to do. Sport psychologists’ main task is to empower athletes and coaches via an individualized approach focusing on their strengths and successful experiences rather than on deficiencies and limitations.

## **2. COMPETITIVE SPORT AS A HIGH ACHIEVEMENT SETTING**

### **2.1. Sport and Competition**

What is an athletic competition? Why is it so important in sport? These questions relate to the psychological characteristics of sport and to social psychological contexts of competitions. Sport as a part of the world and national cultures is also a human activity in which people find, realize, and further develop their individual potentials. Organized competitive sport is characterized by a clear focus on high achievement, exceptional level of skills, enhanced working capacity supplemented by health, well-being concerns, and prevention of injuries. However, the key aspect in understanding the psychology of high achievement sport is competition as a social comparison process.

The essence of sport competition is an evaluation and a social comparison of athletes according to the specially developed and approved rules of that sport. Observable competitive performance is a process of delivering sport results by athletes or teams. Usually, judges measure results in competitions, and on the basis of the comparison between the participants, rank each athlete. In addition, athletes often use self-referenced (process- and/or outcome-oriented) criteria to interpret their results in terms of personal success or

failure. To demonstrate athletic excellence in competitions, athletes have to practice regularly and continuously develop their resources. Both practices and competitions contribute to the development of athletes' physical and mental competencies and skills required in their chosen sport event. However, only participation in competitions allows the athlete to demonstrate her excellence in public and thus to win social recognition and prestige.

Rules of competition in different sports create three distinct psychological contexts for competing athletes: (1) "one-by-one" performances (with no physical or psychological contact between opponents during performance); (2) "one-near-one" performances (with only psychological contact between opponents); and (3) "face-to-face" performances (with both physical and psychological contacts between the opponents during performance). Each of these contexts creates specific challenges for athletes and requires specific resources to cope with task demands. Moreover, a competitor can be either an individual athlete or a team. A sport team has specific structural and dynamic features, such as values, cohesion, communication, and leadership. Group processes can either expand or drain the individual resources of team members, thus affecting the quality of practices and achievement level in competition.

Since the beginning of the modern Olympic movement in 1896, sport has been developing immensely. Contemporary sport has become an international phenomenon and also a part of the world business. Increased mass media involvement has turned international competitions into prestigious social events where athletes often feel extremely high pressure from the social environment. The intensity of competition in high achievement and professional sports has increased dramatically, and in many sports the current level of results is close to the natural limits of human abilities. All this explains the increased role of psychological factors in contemporary sport, creating a challenge for applied sport psychologists to develop effective approaches in helping sport participants.

## **2.2. Individual and Team Excellence**

What is athletic excellence? How is it related to the individual athlete and to the team? Is team excellence simply the sum of individual excellences? How do athletes in team sports find a balance between individual and team goals?

Athletic excellence is defined as an athlete's exceptionally good performance compared with the previously achieved standards. The standards of performance can be self-referenced, i.e., based on a particular athlete's record of achievements and performance history. In contrast, normative standards reflect performance levels of other top performers in a particular sport event. In both cases, the indicators of athletic excellence are results (outcomes) achieved and the quality of performance process (task execution). Athletic excellence is an indicator of athletes' ability to perform consistently up to their potential by recruiting and using effectively the available resources matching the task demands. On the other hand, the notion of personal excellence reflects a high level of ability to function effectively as a human being in different settings, including sport.

Depending on the type of sport activity, athletic excellence can be divided into individual excellence (demonstrated by an athlete) or team (or collective) excellence (achieved by a team). Although team excellence depends on individual contributions, it is often not equal to the sum of the individual performances. Therefore, a team composed of "star" athletes does not always demonstrate team excellence, whereas average players working for the team and sharing team values and high work morale can achieve outstanding team excellence. Research shows that team excellence requires not only individually outstanding performances but also adequate interpersonal and intragroup communication. These communication processes reflected in the team's values, norms, and leadership processes can provide substantial support for the unique resources of team members and compensate for the lack of other resources.

To achieve a collective excellence, it is important to find an adequate balance between the athletes' individual goals and the team goals. These goals usually overlap, but they often do not perfectly match. However, a coach should realize that the degree of this match or mismatch between individual and team goals can result in a balance or imbalance between cooperation and competition processes in the team. Specifically, higher overlap (a match) between individual and group goals leads to better cooperation between teammates, whereas a lower overlap (a mismatch) can result in competitive behavior among the players (e.g., competing for starting positions, playing time, etc.). To find an adequate balance between stimulating athletes to develop their individual excellences and encouraging them to contribute maximally to the team is one of the key issues for coaches.

Another important factor in developing a collective excellence is to identify individual resources and strengths of the players in order to give the players clearly formulated and interrelated roles as the components of specific tasks. Each task may be perceived as a challenge, a routine, or a risk, depending on the perceived relationships between the demands of the task and the available resources (individual and team). When resources and task demands match each other, the team has a set of challenges. Successful coping with challenges results in the development of available resources. However, if available resources exceed the task demands, the task may be perceived as too easy or routine, not requiring recruitment and effective use of resources. This may lead to boredom and low task involvement. Finally, if the task demands exceed available resources, then the task is perceived as a threat and a risk (of failure).

The distinction between challenges, routines, and risks is important for understanding the players' (and team's) development. Per Hendry and Kloep (2002), the lifespan model of developmental challenge states that development occurs when the "pool" of potential resources is added to and resources are strengthened. By contrast, stagnation describes a condition in which no new resources are added to the pool, or they are not strengthened. Finally, a developmental decay in an individual or a team performance is expected if the task demands exceed the potential (available) resources and thus drain the pool ceaselessly. The task of the coach is then to create specific challenges for an athlete (or a team) that will stimulate the effective recruitment, use, and development of existing resources (strengths).

### 2.3. High-Quality Practice

How much time do athletes have to spend in practice in order to achieve athletic excellence? What is the difference between high- and low-quality practice? The major focus in sport psychology since the late 1960s was on successful and poor performances in competitions. Although competitive stress is still a popular topic of research, it is clear that excellence in competitions depends on how much and how well athletes practice. Research shows that top performers typically engaged in 10,000 hours, or 10 years, of deliberate (effortful) and sometimes non-enjoyable preparation to become experts in their domain. Although becoming an expert does require a lot of work, practices must also be high-quality. Moreover, it is important to realize that there is a time limit of what is possible

to achieve in quantity-oriented practice, whereas quality-oriented practice is limitless.

High-quality practices have several important features. First, they require an athlete to be very aware of his or her individual strengths and limitations, optimal emotional states, and bodily signals. An athlete should know how to recognize and monitor this working state during the entire practice and how to recover effectively. Moreover, each training session should have a special meaning for the athlete in relation to a long-term perspective of the season goals and specific tasks. One of the top Finnish alpine Skiers, Tanja Poutiainen, explained in a television interview the "secret" of her successful performances in the World Cup by emphasizing the role of high-quality practices. Specifically, she said, "Now I train differently. I focus thoroughly for each down hill race in practices. I know exactly what I want to achieve and I know what I am working on. It makes much more sense in what I'm doing now. Before I just did it, too often mechanically practicing different movement patterns." This athlete created a mind-set for a task that matched her resources, and these challenges helped her develop as an athlete and a competitor. She was able to learn more about herself and how to use her resources (skills) under different conditions from every practice.

Another important feature of high-quality practices is a simulation of specific competition conditions (e.g., time, competition rhythm, organization, track profile). Learning to focus on one's own game is another important characteristic of effective pre-competition simulations. If practices during the competitive season are more directly related to competition tasks, they serve as a more focused preparation for competitions. On the other hand, lessons learned in competitions provide useful ideas for more effective practices. Especially important are high-quality practices during a competitive season (training between and during several competitions). Basically, the focus of high quality practice is on recuperating, improving, and further developing one's physical, technical, tactical, and psychological resources. Such an approach is especially relevant in professional sport. For example, NHL ice hockey players usually play over 80 games during the season. The players do not have time for much practice, and it is not uncommon that the skills of these talented performers begin to deteriorate. Thus, consistent athletic excellence requires the conservation of available resources (physical, technical, tactical, and psychological strengths) through their recruitment, use, recuperation, and ongoing development.

### 3. PERFORMANCE ENHANCEMENT

#### 3.1. Performance Related Experiences and Athletic Excellence

What is the difference between peak, optimal, and sub-standard performance? What are the optimal and dysfunctional experiences accompanying athletic performance? How do athletes develop competitive experiences?

As mentioned previously, athletic excellence is an extended period of exceptionally good performance by an athlete or a team that exceeds previously established or situationally acceptable self-referenced standards. The usual level of performance provides the frame of reference for defining individually successful (optimal, peak), less than successful (sub-standard, below average, plateaus), and poor (choking, slumps) performances. Peak performance describes an ideal (outstanding, desired) performance. In contrast, optimal performance is the greatest degree attained (or attainable) under implied or specified conditions (e.g., skill level, health status, opponents, weather conditions, competition site). Optimal performance is evaluated using the individualized (self-referenced) criteria based on an athlete's past performance history and present performance status. From this perspective, any athlete can attain an optimal performance, whatever her skill level.

Athletes' behaviors and subjective experiences accompany successful and less than successful performances. Pre-event emotional experiences affect performance, whereas ongoing performance affects the dynamics of mid- and post-event emotional experiences. There are three interdependent levels of human experiences related to and induced by athletic performance: (1) situational transitory emotional experiences (psychobiosocial states) such as anxiety, anger, joy, or excitement, (2) relatively stable patterns of experience (traits, dispositions), and (3) meta-experiences (experiences about experiences). For instance, an athlete can experience a high level of anxiety prior to a competition. This situational state manifests itself in negative thoughts and expectations, such as feeling nervous, worried, and apprehensive. This experience is very individual (idiosyncratic), and for different athletes it can be harmful, can be helpful, or may not affect athletic performance in a particular competition. If anxiety is experienced repeatedly, a consistent pattern of experiences or a typical response disposition (trait anxiety) is formed. However, the athlete often reflects on significant emotional experiences in

particular situations and their effects upon athletic performance. As a result, meta-experiences are formed, and this self-knowledge, beliefs, and attitudes can strongly affect the athlete's interpretations of different performance situations and the choice of adequate (or inadequate) coping strategy.

For instance, Michael Johnson is often quoted as saying that "he was really nervous when he was not nervous prior to an important race." From previous experiences, he knew that high situational anxiety was an optimal experience for his performance. Specific meta-experiences usually trigger corresponding self-empowering or self-defeating thoughts and self-statements and thus determine the beneficial or detrimental impact of emotional state upon performance. Therefore, there is a special need for psychological help for athletes who are unaware of their optimal experiences or whose meta-experiences are less than effective (self-defeating).

There is ample research examining situational emotional states accompanying individually optimal (successful) and less than successful (poor) performances in different athletes across different and similar sports. For instance, the individual zones of optimal functioning (IZOF) model as an individual- and action-oriented framework developed in high achievement setting focuses on optimal and dysfunctional situational experiences accompanying both successful and poor performances. This individualized approach to description, prediction, and explanation of emotion-performance relationships employs a multidimensional conceptualization of emotion as a component of psychobiosocial state. The model predicts interindividual variability of emotion content and intensity and their effects on individual athletic performance based on the "in/out of the zone" principle. It is argued that different forms of psychobiosocial state (cognitive, affective, motivational, bodily, motor-behavioral, operational, and communicative) reflect availability of (or a lack of) resources, their recruitment and utilization, and a need for recovery (recuperation).

Briefly described, these findings indicate that (1) negative situational emotional experiences (such as anxiety or anger) are not always harmful for individual performance; (2) positive emotional experiences are not always helpful or optimal for performance; (3) optimal and dysfunctional emotional experiences are highly individual (idiosyncratic). An optimal emotional performance state is the one most favorable for a particular individual (or a team) under specified conditions, and usually results in an individually successful performance that is equal to or better than realistically expected.

Research also shows that, in contrast to an ideal performance state (flow state) triggered by outstanding performance, optimal emotional states can be positive and negative prior to, during, and after performance. Positive optimal states are experienced when an athlete's resources match well with current task demands; positive dysfunctional states reflect a routine performance situation in which resources are available but are neither recruited nor used properly. The task is perceived as too easy, which results in excessive (demotivational) satisfaction (leading to complacency and less involvement in the task) and even boredom. Negative optimal states (anger, anxiety) reflect a threat (or a risk) situation (task demands exceeding available resources) in which an athlete attempts to actively cope with this imbalance. Finally, negative dysfunctional states reflect a situation when an athlete is unable to cope, with task demands exceeding currently available resources. Repeated experiences related to unsuccessful performance (slumps) and a failure to recuperate existing resources could result in chronic staleness, overtraining, and burnout.

Sport psychology describes different aspects of performance-related situational experiences that actually characterize a state of readiness for competition. These include self-confidence (state and trait), attention and concentration, experiential and behavioral manifestations of intrinsic and extrinsic motivation at the situational and dispositional levels, and individually optimal levels of anxiety. This research indicates that the intensity of anxiety (as well as other emotions) associated with optimal sport performance varies considerably across athletes, even for those competing in the same event. It also indicates that a substantial percentage of athletes actually benefits from elevated anxiety; in these cases, interventions aimed at reducing anxiety may be counterproductive. These findings illustrate the notion that optimal anxiety reflects attempts to situationally compensate for an apparent lack of resources as related to task demands as a person-specific coping strategy. The effect is further enhanced by an optimal self-empowering meta-experience, i.e., the athlete knowing that such a level of anxiety is useful and helpful for him. The previously cited example of Michael Johnson illustrates this point.

On the other hand, athletic excellence requires an optimal and sustained effort, and the athlete's body must be appropriately energized, with physiological and psychological resources prepared for the stresses and physical demands of competition. Arousal as a component of the psychobiosocial state manifests itself

in physiological reactivity and physical energy. It is also paired with varying levels of concomitant cognitive, affective-volitional, and motivational activity, and behavioral display.

Typically, the physiological component of arousal is measured through muscle tension, cortical activity, electro-dermal activity, respiration, and biochemical markers such as epinephrine and cortisol. However, in recent years, there have been numerous attempts to identify idiosyncratic markers of perceived subjective bodily response to competitive stress both in best and worst competitions. This line of research has a great potential for the practice of sport psychology by providing a tool to enhance an athlete's bodily awareness.

### 3.2. Resources as Performance Enhancement Strategies

What are the internal and external resources that can enhance athletic performance? How can athletes and teams use their resources more effectively?

The construct of internal and external resources proposed here is not entirely new. For example, it is used in the conservation of resources (COR) model proposed by Hobfoll to define and explain psychological stress. Examples of broadly defined resources include not only personal characteristics (self-esteem, mastery, well-being) but also interpersonal, material, and work-related resources. The basic tenet of the COR model is that people strive to retain, protect, and build resources because the potential or actual loss of these resources is a threat and a source of psychological stress. Then psychological stress is defined as a reaction to the environment in which there is (1) the threat of a net loss of resources, (2) the net loss of resources, or (3) a lack of resource gain following the investment of resources.

Hendry and Kloep proposed the lifespan model of developmental challenge, which employs the constructs of resources and challenges to explain the processes of human growth. Examples of potential resources include biological dispositions (health, personality, talents, intelligence, body shape, attractiveness); social resources (trust, attachment, size and quality of network); skills (basic, learning, social, psycho-motor, etc.); self-efficiency (self-efficacy appraisals, experience with success, assurance from others, locus of control); and structural resources (country, race, class, family, income, gender).

In competitive sport, resources are defined as psychobiosocial assets that determine athletes' ability to

consistently perform up to their potential. Here the emphasis is on how available resources are identified and then systematically and effectively recruited, used, recuperated, and further developed. There are four closely related approaches to the enhancement of athletic performance: situational and individually optimal states, relatively stable experience patterns (dispositions, personality traits, and sport-specific qualities), psychological skills, and group dynamics factors. The strategies used in each of these four approaches to performance enhancement are actually different groups of internal and external resources. There is considerable overlap between these four groups of resources, and they also concur well with the COR model and the lifespan model of developmental challenge.

Earlier discussion of individually optimal situational states and relatively stable experience patterns indicates that an athlete's awareness of her optimal states and adequate meta-experiences can be a very effective internal resource. Furthermore, personality characteristics and sports-specific qualities (e.g., winner's profiles, a wheel of excellence) are also important potential resources for achieving athletic excellence. Although personality characteristics do not directly predict situational performance, they could be instrumental in predicting long-term effects of sport participation, for example the mental health model proposed by Morgan (1985). However, it is important to realize that athletes can achieve success even if they lack certain personality traits and qualities. The implication for a consultant working with an athlete (or a team) is clear: he or she should focus on developing available individual strengths rather than repair apparent deficiencies.

Psychological skills as a set of techniques and coping strategies aiming to produce an optimal state of readiness are valuable resources that successful athletes learn and use systematically to achieve consistent excellence. These resources are usually targeted at some specific modality. Therefore, the classifications of psychological skills usually include implicit or explicit reference to some form of human functioning (e.g., cognitive, affective, motivational, bodily, motor-behavioral, operational, communicative). At the same time, these different forms of functioning as components of psychobiosocial state can also be employed to describe different task demands.

Group dynamics and environmental factors are also important personal and team resources. These potential resources include cohesion, psychological climate in the team, patterns of interpersonal communication

among the teammates and between the coach and the players, leadership style, and group norms and values reflecting sport subculture. For instance, high work morale and honesty as accepted values and group norms in a sport team could promote fair play behaviors and considerably minimize cheating in sport.

From social psychological and environmental perspectives, it is important to realize that competitive sport is a part of the society at large. Therefore, equitable or non-equitable conduct found in society is generally reproduced in sport settings. When sport traverses racial, ethnic, social, national, and gender boundaries, it has the power to bring diverse people together while de-emphasizing social or cultural differences. In other words, fair treatment in sport, in contrast to other settings, can provide conditions that significantly extend existing personal (or team's) resources. As an external resource, while sport offers the opportunity to traverse cultural identities and unite different peoples, it also can have the opposite effect. Thus, unfair treatment or even discrimination in sport can overtax the athletes' and teams' resources and considerably slow down their situational success and long-term development. In some cases, however, this negative treatment can be a strong motivator for athletic and personal excellence.

As Krane argues, fair treatment in sport occurs only when there are equitable resources and opportunities for all participants, regardless of social group membership (e.g., gender, race, ethnicity, nationality, sexual orientation, and social class). In reality, that is not always the case. In some cases, differential treatment occurs, in which minority social group members are treated unfairly. In other situations, sport is an avenue for educating people and increasing awareness about a wide range of social issues, fighting social injustice, and providing humanitarian assistance. Therefore, applied sport psychologists may employ a variety of strategies to promote fair and equitable sport.

### 3.3. Barriers to Athletic Excellence

What are the barriers to optimal athletic performance? How can athletes (and teams) minimize or cope with them? The threat of a net loss of resources, the net loss of resources, or a lack of resource gain following the investment of resources can be strong barriers to successful performance.

Four groups of internal and external resources (situational states, personality traits, psychological skills, and group dynamics factors) proposed earlier

can provide a framework for describing potential barriers to athletic excellence. Specifically, the notion of resources and their role in enhancing athletic performance is dialectic. A lack of resources or a failure to identify, recruit, and use them effectively could become a potential serious barrier to consistent athletic excellence. Examples of such barriers are dysfunctional emotional states, an overemphasis on apparent deficiencies, and a lack of performance-related skills. Finally, environmental barriers include inadequate motivational climate in the team, selfish behaviors of teammates, media pressures, and conflicts between a coach and an athlete. The typical consequences of the impact of barriers include performance slumps, overtraining, burnout, and injuries.

To minimize detrimental effects of internal and external barriers, it is recommended that an athlete's awareness of available resources and the strategies of their ongoing development is enhanced. Such awareness should be extended to a better understanding of causes of sub-standard performance and learning better risk management by maintaining self-efficacy, emotional control, and individually effective coping skills. Although the emphasis here is on situationally effective coping strategies, their role should also be understood from a wider (career development) perspective.

## **4. ATHLETIC EXCELLENCE FROM A DEVELOPMENTAL PERSPECTIVE**

### **4.1. Athletic Career Demands, Coping Resources, and Barriers**

What do athletes have to go through in order to achieve athletic excellence? Can any athlete reach it? What factors help athletes to reach excellence, and what might act as barriers along the way? These questions relate to the athletes' development during their athletic career and the demands they have to cope with by using specific resources.

A rapidly growing body of research in sport psychology focuses on "athletic careers" in an attempt to better understand how different athletes in different sports become expert performers and how they reach and maintain consistent excellence. Metaphorically, the athletic career (from initiation of sport participation to the retirement from sport) can be described as a miniature lifespan course involving a number of important transitions between the predicted stages. Understanding the mechanisms of these transitions

and stages is important for coaches, athletes, parents, and sport psychologists.

It is very common for an athlete taking the first steps in his sport to dream of reaching athletic excellence, turning professional, and winning the world championship or Olympic games. However, it usually takes a long time to make this dream true. A so-called "athletic pyramid" shows metaphorically that only a few athletes achieve athletic excellence and have successful (elite, recognized, professional) athletic careers. For instance, a pyramid with one professional soccer player at the top contains 6000 soccer players at bottom; a pyramid with one professional basketball player at the top has 14,000 players at the bottom. According to Csikszentmihalyi and Robinson (1986), "in highly competitive domains, such as music, math, or sports, the way down is always much broader than the way up. Year by year, it becomes more difficult to catch up, and dropping out becomes increasingly easy" (p. 275).

The athletic career of each individual athlete is unique, and there are still debates in sport psychology about factors contributing to individual differences in sport achievements. It is becoming increasingly clear that interplay of several groups of factors can help or hinder an athlete's development and achievement of athletic excellence. These factors include an athlete's innate talent/potential, environmental factors (competent coaches, family support, adequate conditions for practice, etc.), and an athlete's ability to develop, recruit, and use effectively all resources necessary to cope with the increasing demands of the athletic career.

An athletic career usually starts at the age of 7 to 10 years, sometimes even earlier depending on the sport event (e.g., in swimming, artistic gymnastics, figure skating, and ice-hockey). First, children perceive sport as merely "playing a game," however, later their attitudes change and sport becomes as "a sphere of education". Much later for those who reach the top it becomes a job or professional activity. It takes usually about 10 years of deliberate practice to reach an expert performance level in sports, and once there, the period before retirement usually lasts between 5 and 15 years. At the most general level, an athletic career typically consists of several stages: initiation, development, mastery/perfection/culmination, maintenance, and discontinuation.

Athletes striving for athletic excellence and staying at the top have to cope with increasingly complicated demands related to their practice, competitions, communication, and life outside of the sport. There are specific demands at the beginning (transitional) part of each athletic career stage. Research findings

summarized in the 1997 FEPSAC Position Stand # 3 indicate that the beginning of sport specialization requires adjustments to the demands of the sport event, coach, sport group, and a new schedule of everyday life. Young athletes must ensure the right choice of sport, show ability in learning sport skills, and test themselves in their first competitions. When the athlete and coach decide to work for results, they enter the transition to the development stage or intensive training in the chosen sport, characterized by more intense and specialized practice and participation in higher level competitions. This transition requires that athletes adjust to higher physical and psychological loads, improve their technical and tactical skills, achieve relatively stable results in competition, and balance the time and energy taken by their sport with other activities (studies, leisure, etc.).

The first significant success brings the athlete to top-level sport with its tough competitions, and that indicates a transition to high-achievement and “adult” sport, or to the mastery/perfection/culmination stage of athletic career. Further progress requires that athletes revise their lifestyle so that it works for their sport goals. They should also find their individual paths in sport, the ways to cope with pressures of selection to important competitions, and to gain respect of the team, opponents, and judges. In short, this stage is where an athlete earns her reputation, which later will work for her. Transition from amateur to professional sport is marked by adaptation to specific requirements and pressures of professional sports, competitions with very strong opponents, more independent training, and striving not only for the victory but also for fans’ sympathies. The transition from the culmination to the maintenance stage of an athletic career is characterized by the necessity to search for additional resources in order to maintain a high level of achievements and to plan athletic retirement. The termination of athletic career is marked by leaving sports and transitioning to some other professional career, with adjustments to a new status, lifestyle, and social network.

The career demands briefly described above characterize so-called normative transitions. However, each athlete also experiences a number of non-normative (idiosyncratic) transitions related to his or her particular situation or environment. Transition demands create developmental conflict between “what the athlete is” and “what he or she wants or ought to be,” which stimulates the athlete to find additional coping resources. The effectiveness of coping depends on the

dynamic balance between transition resources and barriers. Transition resources includes all internal and external factors that facilitate the coping process (e.g., the athlete’s self-knowledge, skills, personality traits, motivation, availability of social and/or financial support). Transition barriers include all internal and external factors that interfere with effective coping (e.g., a lack of necessary knowledge or skills, interpersonal conflicts, difficulties in combining sport and studies or work). Interestingly, the same experience may be either a resource or a barrier depending on the situation. For example, athletic identity, which according to Brewer *et al.* (1993) is “the degree to which the individual identifies herself with the athletic role” (p. 237), is usually an important resource for an athlete, especially when she is at the peak of her career. However, it can turn into a serious barrier in the process of adaptation to post-athletic career life.

Typically, at the beginning of their athletic career, athletes experience a lack of internal resources (sport-specific knowledge and skills), which are compensated for by social support from a coach, the family, and peers. At the culmination of their athletic career, athletes are usually at their most resourceful and their career demands are the highest. Elite athletes often rely very much on their relatively stable experience patterns and meta-experiences. At the maintenance stage of their athletic career, athletes often lack social support; their health deteriorates; they are bothered by the consequences of injuries, a lack of energy, and pressures in other spheres of life. However, all these concerns can be compensated for by the individualization of all aspects of the athletes’ preparation. For example, veteran athletes typically train less than their younger counterparts, but they use their individual strengths more effectively. This allows them to maintain results at a high level until the very end of their athletic career.

#### **4.2. Athletes’ Successful Transitions and Crisis Transitions**

What happens in a career transition when the athlete is either able or fails to cope with transition demands? Do athletes need any psychological help while in such a transition? If yes, what kind of help would be useful? These questions relate to the coping process, the outcomes, and consequences of career transitions.

The coping process is central in a transition and includes all strategies the athlete uses in order to adjust

to particular transition demands. An adequate match between the perceived demands and available resources creates a state of readiness to the career transition and a higher probability of successful transition. Successful transition is associated with effective coping when the athletes are able to recruit, use, or rapidly develop necessary resources and avoid (or overcome) potential transition barriers. One of the principles in effective coping is relying on athletes' strengths, which can compensate for potential and existing weaknesses or barriers.

An alternative outcome is a crisis transition, when an athlete is unable to cope effectively on his own with the demands of the transitional situation. Research identified a set of symptoms or markers describing typical reactions of athletes in crisis transition, including a decrease in self-esteem (as a first reaction to ineffective coping) and chronic emotional discomfort. Athletes also report new fears, increased sensitivity to failures, poor decision-making, and inadequate behaviors. Attempts to change the situation are usually ineffective, and instead of improvement new mistakes (and failures) are committed. Therefore, athletes in a crisis transition often describe feeling like they are in a blind alley or a dead end. For instance, as reported by Stambulova and Lindwall in 2002, one elite athlete who dropped out of sport after becoming caught up in doping described her feelings in the crisis as follows: "I totally panicked and did a terrible error . . . I took . . . forbidden substances as a final effort to get away from the feeling of being useless. My head was in chaos and there were no open roads left to take" (*Svensk Idrottspsykol.*, 2, 2–5).

Athletes in crises need psychological assistance to shift them from a "dead end" situation to a "cross-road" situation and to see several new coping alternatives. Moreover, psychological intervention influences the consequences of the transition. Effective intervention leads to successful but delayed transition. Alternatively, ineffective or no intervention situations are followed by negative consequences or so-called costs for failure to cope with the transition. Possible costs include decline in sport results, injuries, overtraining, neuroses, psychosomatic illnesses, prematurely quitting sports, and also different forms of rules violation and degradation of personality (e.g., alcohol and drug use, criminal behaviors). All these costs can be seen as negative effects of sport participation and also as barriers to coping with forthcoming career demands.

A developmental perspective provides a framework for a better understanding of career transitions. For instance, Vygotsky's constructs of the zone of actual development (ZAD) and the zone of proximal

development (ZPD) could be instrumental in prediction of transition consequences. The ZAD is a range of the tasks that a person can solve on her own; the ZPD is a range of the tasks that a person can solve only if assisted by others. If most of the athlete's coping resources are in her ZAD, a successful transition can be predicted; in contrast, a crisis transition is expected if most of the athlete's resources are in her ZPD. Therefore, a psychological intervention should focus on helping an athlete to develop new resources and overcome potential transition barriers, especially if transition demands exceed available resources.

The lifespan model of developmental challenge earlier applied to athletic performance (from a short-term perspective) can also be used for the interpretation of career transitions. Typically, transition demands require a long-term coping process and many resources. Successful coping means adding new resources and an outcome in the form of development. If no new demands are made and the athlete simply repeats everyday routines, development eventually turns into positive stagnation. Crisis transition can be seen as negative stagnation, which might turn into development (under condition of qualified psychological assistance to the athlete) or into decay (i.e., negative consequences of not coping with the transition).

Developmental psychology interpretations demonstrate the dialectic nature of career transitions and their role in achieving athletic excellence. Each career transition with its accompanying demands is a step to athletic excellence. There is a risk of not meeting the demands, resulting in negative stagnation or decay. But there is also a chance to develop further and to experience positive stagnation on a higher level.

Three types of psychological interventions can be useful for helping athletes in career transitions: (1) a crisis prevention, (2) psychological crisis coping, and (3) psychotherapeutic (clinical) interventions. Crisis prevention involves career planning and goal setting, mental skills training, and organization of a social support system. This intervention aims to prepare athletes for a transition in advance by developing their resources for effective coping. This approach actually enhances their readiness for the transition either on their own or by using an expert assistance. Psychological crisis coping is an intervention for athletes already in crisis transition; it includes mainly individual counseling and psycho-correction programs. The focus here is on helping the athlete to analyze her situation, to find the best option for coping, and to develop and realize the action plan. These interventions usually deal with negative

stagnation and help the athlete to turn it into a development situation. Psychotherapeutic or clinical interventions are applied when the athlete has already experienced one or several of the above-mentioned negative consequences of not coping with a crisis transition. In other words, these interventions deal with a decay situation, trying to stabilize and then to improve the athlete's situation.

### 4.3. From Athletic to Personal Excellence

What are the benefits and costs of many years of participation in sport? How can sport psychologists help athletes to maximize the benefits of an athletic career and to minimize its costs? How can a successful athletic career contribute to the athlete's life outside sport?

An athletic career can be evaluated not only as a stage-like developmental process, but also as a developmental event contributing to the lifespan development in and outside sport. From this perspective, several parameters characterize an athlete's development during his or her athletic career. These include duration of sport participation from start to peak and finish, the sport event(s) practiced, the degree of specialization, and achieved sport titles/records/results. Subjective indicators include perceived benefits of sport participation and its costs (in terms of time, energy, health, money, etc.) as well as career satisfaction (one's self-esteem in regard to the athletic career) and career successfulness (social recognition of one's athletic career).

Successful (or elite) careers are usually associated with athletic excellence, whereas satisfactory careers are associated with achieving individual peaks corresponding to the individual resources and environment. Satisfaction is based on a set of self-referenced criteria, but most often it consists of perceived potential in relation to level of achievements and athletic career costs. Interestingly, some athletes are often satisfied with non-elite careers, if they value the developmental effects (benefits) of sport participation (e.g., self-knowledge, physical fitness, good health, skills, qualities, social contacts that can be used in other spheres of life). In contrast, other athletes may be dissatisfied with their elite careers, especially if they perceive the costs as too high (e.g., deteriorated health, deficits in education, a lack of close personal contacts or any interests outside sports).

To achieve athletic excellence, athletes have to start and specialize in a particular sport event quite early.

This can facilitate young athletes' progress in a chosen sport, but it also can result in several negative consequences such as high pressures, fears, and one-sided development. To avoid this, coach effectiveness training encourages coaches working with children and youth to focus more on optimal development of young athletes than on "winning at all costs." Positive developmental effects related to athletes' self-esteem, skills level, and satisfaction with various aspects of sport participation should be provided for all young athletes, and then allowing the most talented of them to move further to the athletic excellence level.

In a broader sense, sport psychology aims to help all athletes, including top performers facing tough transitions and pressures of their careers, to achieve optimal development and their individual peaks in sport. Therefore, career/developmental perspective in applied work with athletes includes several aspects: (1) "whole career" approach, which spans the athletic career—from initiation to termination—as well as the post-athletic career; (2) "whole person" approach (taking into account not only athletic but also non-athletic developments of athletes); (3) developmental approach (links between past, present, and future); (4) activity-specific approach (taking into account general and sport event-specific factors); (5) individual approach (taking into account typical and individual patterns); and (6) transferable skills approach. For instance, the latter refers to a series of sport-based life skills programs that aim to teach physical and mental skills (e.g., emotion self-regulation, effective communication, goal setting, coping with success and failure), which can be generalized to various spheres of the participants' life outside sport. This approach can be useful at each stage of athletic career, especially for retired athletes, to help them adapt their skills and experiences acquired during sport participation to their post-athletic career life.

A challenge for sport psychologists helping athletes to reach athletic and personal excellence is to find the right balance between situational current problems and future career development issues. For example, what is more important for the athlete: to prepare well for a competition in the next week or to make sure she joins a national team in the next year? The other dilemma, for example, with a veteran athlete, is whether to focus on searching for additional resources to help him keep his sport results at a high level or to plan for retirement and post-athletic career life. The best answers to these and other similar questions can be provided by viewing applied sport psychology as both a science and

an art. The science viewpoint tells us that it is important to keep in mind both the situation and career perspectives; whereas the art viewpoint, based on past experiences, skills, and intuition, can help answer the question of how to do this.

## 5. CONCLUSION: FUTURE DIRECTIONS IN SPORT PSYCHOLOGY

What has been achieved in applied sport psychology? What are the main concerns of the field right now? What is on its future agenda?

In order to enhance the effectiveness of scientific support in elite sport, several new future directions from a research-oriented and a practical (organizational) perspective can be identified. These include a new emphasis on the role of elite coaches in psychological preparation of athletes and team, more focus on team-building, environmental, and organizational factors, and the development of closer international cooperation between scientists, practitioners, and sports organizers. Each of these aspects is briefly described in the sections that follow.

*More psychological support for elite coaches.* Initial focus of most sport psychology research and interventions on athletes and teams is well documented in the literature. However, the role of coaches in the psychological preparation of athletes and teams should be further emphasized. In practice, this means that the coach should be the central figure in preparation of the team, and sport psychologists should work more through the coach and with the coach-athlete team rather than only with the athlete. Enhancing the psychological competence of coaches can be a decisive factor in enhancing the quality of coaching.

In the past, sport psychology interventions and mental training programs usually focused on competing athletes who were coping with competition stressors. Less attention was paid to high-quality practices and prevention of overtraining, staleness, burnout, and injuries. Therefore, a most urgent and promising area of research and applications in sport psychology now and in the future should be the optimal performance of coaches and their coping skills for handling short-term and long-term chronic (e.g., burn-out) stresses. Qualitative research on careers of outstanding coaches to identify the factors of their consistent excellence would be a challenge for future researchers and

practitioners. On the applied side, it would be helpful to summarize experiences of how ongoing individualized consultancy (personal coaching) for coaches has helped them anticipate the critical transition periods in their careers.

*Team-building and effective management.* In the past, social psychological research in sport psychology comprised 8–10% of all efforts, and the role of environmental and organizational factors in elite sport is still underestimated. Therefore, sport psychology should focus more on a holistic approach to the interpersonal and group processes that are determining performance and the life of a team in a wider social and cross-cultural context. Optimization of communication in the team is a very promising and productive area of research and applications. Practically, very little is known about the psychology of effective management in elite teams, sports federations, and clubs. Considering the quick development of elite sports, such areas as organizational development, change, and change management are potentially very important as new directions for research and applications. Experiences and practices of organizational psychology and management already available in non-sports high achievement settings could be beneficial for sport. On the other hand, the findings obtained in elite sports might be of interest to top management, business, army, and police.

*Cross-cultural adaptation of athletes and coaches.* Recent developments in Europe and worldwide indicate that more and more elite athletes and coaches are working abroad. These professionals need new skills for successful adaptation to a new environment and its constant changes. Quick adaptation to a new team, teammates, and coach, effective contacts with the media, and negotiation skills, for example, are much-needed resources for elite athletes and coaches. Moreover, with more migration and higher mobility rates among elite coaches, a critical factor is the assessment of a candidate's potential for cross-cultural adaptation and individualized programs that could facilitate his or her entry to a host country. This is especially important in view of the fact that tradition and values vary by country, and, for instance, a well-meaning but authoritarian coach with a clear orientation on success can be less than effective when starting his work in an amateur-oriented environment of the host country. A follow up with the coach or athlete could be instrumental in helping them to quickly adapt and effectively function both professionally and personally in the new environment.

*International cooperation of sport psychologists.* There are indications that in the future, a better collaboration

between applied sports psychologists from different countries could be useful not only for research but also in consulting. With recent developments in world-wide communication, joint consulting and psychological support for coaches and athletes across different countries seems like a reality in the near future. Developing such a network of sport psychologists could be an interesting initiative, especially in places where there is a lack of experts who could provide high-quality services (in research and applications) for elite athletes and coaches. One possible solution would be to use the expertise of internationally recognized applied researchers and practitioners in sport psychology who could deliver the necessary services for elite athletes, teams, and coaches and provide hands-on experiences for the local young aspiring sport psychologists interested in working with elite performers.

To conclude, now as never before, the application of what is already available in sport psychology is extremely important. Practical experience and expertise available in sport psychology are important not only in competitive and elite sport settings but also in such high-achievement settings as the performing arts and business. There are promising indications that the gap between theoretical knowledge and experience-based knowledge in sport psychology is gradually being bridged. Moreover, there is a clear shift in applied sport psychology from a predominantly negative, problem-oriented, and deficit-repairing approach initially borrowed from clinical psychology to a more positive psychology focusing on optimal performance and on an athlete's and team's strengths rather than limitations. Another promising trend in sport psychology is more emphasis on idiographic (individual-oriented) and experience-based approaches rather than on traditional nomothetic (group-oriented) comparisons of successful and less than successful athletes. Finally, early attempts to use personality tests to predict situational performance proved to be unsuccessful. A new and more promising approach is to conceptualize the situationally oriented applied work focused on enhancement of athletic performance within the framework of developmental perspective. This may provide an opportunity for sport psychology to become the psychology of athletic and personal excellence.

### See Also the Following Articles

Anxiety and Optimal Athletic Performance ■ Arousal in Sport ■ Assessment in Sport Psychology ■ Attention and Concentration Training in Sport ■ Competition in Sport

■ Decision Making in Sport ■ Emotion in Sport: An Individualized Approach ■ Fair Treatment and Discrimination in Sport ■ Goal Setting and Achievement Motivation in Sport ■ Group Dynamics in Sport ■ Injury in Sport ■ Intrinsic and Extrinsic Motivation in Sport ■ Overtraining and Burnout in Sports ■ Performance Slumps in Sport: Prevention and Coping ■ Psychological Skills Training in Sport ■ Self-Confidence in Athletes ■ Successful Athletic Careers

### Further Reading

- Aspinwall, L. G., & Staudinger, U. M. (Eds.). (2003). *A psychology of human strengths. fundamental questions and future directions for a positive psychology*. Washington, DC: American Psychological Association.
- Brewer, B. W., Van Raalte, J., & Linder, D. E. (1993). Athletic identity: Hercules' muscles or Achilles heel? *International Journal of Sport Psychology*, 24, 237–254.
- Csikszentmihalyi, M., & Robinson, R. E. (1986). Culture, time and development of talent. In R. J. Sternberg, & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 264–284). New York: Cambridge University Press.
- Danish, S. J., Petitpas, A. J., & Hale, B. D. (1993). Live development interventions for athletes: Life skills through sports. *The Counseling Psychologist*, 21, 352–385.
- Definition of sport psychology. (1995). FEPSAC Position Stand # 1. *FEPSAC Bulletin*, 7, 3–5.
- Durand-Bush, N., & Salmela, J. (2001). The development of talent in sport. In R. Singer, H. Hausenblas, & C. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 269–289). New York: John Wiley & Sons, Inc.
- Ericsson, K. A. (1996). *The road to excellence: The acquisition of expert performance in the art and sciences, sports and games*. Mahwah, NJ: Erlbaum.
- Hanin, Y. L. (2003). Performance related emotional states in sport: A qualitative analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 4(1). Available at <http://www.qualitative-research.net/fqs-texte/1-03/1-03hanin-e.htm>.
- Hanin, Y. L., & Stambulova, N. B. (2002). Metaphoric description of performance states: An application of the IZOF model. *The Sport Psychologist*, 16, 396–415.
- Hanin, Y. L. (Ed.). (2000). *Emotions in sport*. Champaign, IL: Human Kinetics.
- Hanin, Y. L. (1999). Sport psychology research and consultancy with elite athletes and coaches. In E. Muller, F. Ludescher, & G. Zallinger (Eds.), *Science in elite sport*, (pp. 129–145). London, UK: E & FN Spon/Routledge.
- Hendry, L. B., & Kloep, M. (2002). *Lifespan development: Resources, challenges, and risks*. London, UK: Thomson Learning.
- Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, 44, 513–524.

- Orlick, T. (1990). *In pursuit of excellence. How to win in sport and life through mental training* (2nd ed.). Champaign, IL: Leisure Press.
- Smith, R., & Smoll, F. L. (1991). Behavioral research and interventions in youth sports. *Behavior Therapy, 22*, 329–344.
- Smith, R., Smoll, F. L., & Curtis, B. (1979). Coach effectiveness training: A cognitive-behavioral approach to enhancing relationship skills in youth sport coaches. *Journal of Sport Psychology, 1*, 59–75.
- Sports career transitions. (1997). FEPSAC Position Stand # 3. *FEPSAC Bulletin, 9*, 6–8.
- Stambulova, N. B. (1994). Developmental sports career investigations in Russia: A post-perestroika analysis. *The Sport Psychologist, 8*, 221–237.
- Stambulova, N. B. (2000). Athletes' crises: A developmental perspective. *The International Journal of Sport Psychology, 31, 4*, 584–601.
- Stambulova, N. (2003). Symptoms of a crisis-transition: A grounded theory study. In *SIPF yearbook* (pp. 97–109). Örebro, Sweden: Örebro University Press.
- Vygotsky, L. S. (1984). *Sobranie sochinenii, T.4* [The complete works, V.4]. Moscow: Pedagogika.



# Stereotypes

David L. Hamilton and Sara A. Crump

University of California, Santa Barbara, California, USA

1. Introduction
  2. Conceptual Approaches to Understanding Stereotypes
  3. Stereotype Activation
  4. Measuring Stereotypes
  5. When Do We Use Stereotypes?
  6. Consequences of Stereotyping
  7. Stereotype Change
  8. Conclusion
- Further Reading

## GLOSSARY

**authoritarian personality** Stereotypes and prejudice are thought to be the result of a personality syndrome that results from strict parenting style and is characterized by obedience to authority figures, strict adherence to rules, and intolerance of deviation from conventional norms.

**categorization** The mental process of organizing people, objects, and events into meaningful groups.

**contact hypothesis** The hypothesis that bringing people together in a supportive environment may help to reduce stereotyping and prejudice by fostering mutual understanding between members of different groups.

**out-group homogeneity effect** The perception that members of one's out-group are more similar to each other than are members of one's in-group.

**priming** An observed stimulus consciously or nonconsciously activates a related concept in the mind, which can then influence subsequent information processing.

**self-fulfilling prophecy** A person's prior expectancy (e.g., impression or stereotype) about another person influences

his or her behavior and leads the other person to behave in ways that confirm the original expectation.

**social desirability** The desire to act in ways that are socially acceptable, even if the behavior does not reflect one's true beliefs.

**subtyping** The mental act of placing individuals who violate group stereotypes into their own separate category, thereby maintaining the original stereotype.

Stereotypes are cognitive structures containing our knowledge, beliefs, and expectancies about social groups. These belief systems arise from personal experience with groups and their members, through the transmission of social norms, and from people's tendency to organize and understand the world through categorization processes. Once formed, stereotypes have a number of consequences for people's perceptions and judgments of others as well as for their behavior.

## I. INTRODUCTION

Throughout our daily lives, we encounter a vast array of people and situations that we perceive and interact with on a regular basis. Often, we perceive those people in terms of their memberships in some groups, and in many cases we have prior beliefs—stereotypes—about those groups. Stereotypes can then shape the way we perceive members of those groups, and they may influence our feelings and behavior toward them.

In 1922, journalist Walter Lippman coined the term stereotype to refer to people's penchant for consistently describing members of a group in the same ways. He conceived of stereotypes as "pictures in people's heads" that guide their perceptions of groups. More contemporary writers, building on Lippman's thinking, define stereotypes as people's beliefs about a particular group or social category. Specifically, stereotypes consist of a person's knowledge, beliefs, and expectancies about social groups. For example, we may believe that Italians are extraverted and passionate, whereas the Chinese are more polite and reserved. As these examples illustrate, stereotypes—even when based on some "kernel of truth"—are overgeneralizations in that they express the belief that all members of a social category share similar stereotypical characteristics. These beliefs may then influence our feelings about members of the group (prejudice) as well as our behavior toward them (discrimination). In short, whether accurate representations or not, we all use stereotypes to help guide our thoughts and expectancies toward others.

## 2. CONCEPTUAL APPROACHES TO UNDERSTANDING STEREOTYPES

Three major conceptual orientations have guided research on how and why people develop and use stereotypes: the psychodynamic/motivational, sociocultural, and cognitive approaches.

### 2.1. The Psychodynamic/Motivational Approach

The psychodynamic/motivational approach views stereotyping as serving the needs of the individual perceiver, protecting and enhancing the self by derogating others. There have been several variations on this theme.

In early work guided by Freudian theorizing, stereotypes were seen as a response to unacceptable impulses or tensions within the individual. For example, when an individual feels that he or she possesses certain negative qualities, rather than directly addressing those issues, the individual may do so indirectly by projecting his or her own negative characteristics onto others. When those attributes are projected onto an entire group of people, the result is stereotyping. This process serves the function of avoiding recognition of those qualities in oneself. Reflecting this approach, in 1950 Adorno *et al.* proposed that stereotypes and prejudice

arise from strict child-rearing practices that result in an "authoritarian personality." This personality syndrome is characterized by strict obedience to authority, rigid adherence to rules, and intolerance of deviation from conventional norms.

Recent research has focused on an individual's motivation to maintain a positive self-concept and its role in intergroup perceptions. For example, Tajfel and Turner's 1986 Social Identity Theory posits that people derive part of their self-esteem from their group memberships. Because people want to feel good about themselves, they therefore are motivated to feel good about the groups to which they belong. Consequently, people often derogate out-groups in order to make their own groups appear more favorable, thereby contributing to their self-esteem. In support of this view, some studies have shown that when a person's self-esteem is threatened, stereotyping of out-group members leads to an increase in self-esteem.

Finally, stereotypes may also help us to justify and perpetuate the status quo. For example, placing blame on those in our society who are disadvantaged (e.g., believing that African Americans are poor and of lower status because they are lazy) not only can help us to relieve guilt and maintain a positive self-image but also can perpetuate the disadvantaged status of less fortunate groups.

According to the psychodynamic/motivational perspective, then, stereotypes help us to deal with internal conflict and to maintain self-esteem.

### 2.2. The Sociocultural Approach

The second major approach to understanding stereotypes focuses on the influence of one's social and cultural environment on stereotype development and use. Specifically, according to the sociocultural approach, stereotypes are passed on to children through socialization processes and reflect the norms of society. At an early age, children become aware of the values held in their society through the teachings of their parents, in their interactions with their teachers and peers at school, and by what they see on television. For example, children may overhear their parents making derogatory comments about people of a different race or religion, their teachers may treat them differently depending on whether they are male or female, their peers may pick on certain children because they are developmentally delayed, or they may see members of some groups consistently presented in a stereotypical way on television programs.

In addition, certain groups of people are overrepresented in some roles in our society, potentially leading

children to believe that those positions reflect the inner dispositions of those group members. For example, children may observe that members of ethnic minorities are primarily employed in the lower paying jobs in the workforce, which may lead them to believe that those minorities are not hard workers or not intelligent enough to get better jobs. They may observe that men occupy most executive and leadership positions, fostering the belief that males are more aggressive and powerful than females. In each of these situations, children are learning the stereotypes held in their society by observing the distribution of group members in different social roles. Although that distribution may be arbitrary or may be constrained by external factors, observers may infer that it reflects internal, psychological characteristics, which then become part of the stereotype.

### 2.3. The Cognitive Approach

According to the cognitive approach, stereotypes develop in part as a result of people's natural tendency to categorize the people and events they encounter in the complex world in which they live. In fact, research guided by this approach has shown that the categorization of people into groups is a natural, fast, and effortless process that is used by everyone. For example, when meeting people for the first time we automatically categorize them in terms of their age, race, and gender—physical properties that are visually salient and are often believed to provide more information about a person than other physical features (e.g., brown hair). Social categorization can also be based on other features, including nationality, religion, occupation, sexual orientation, and many other characteristics. Through the categorization of individuals into groups, perceivers are able to more efficiently organize their knowledge about the world.

Categorization by itself does not mean that stereotypes are formed. However, once we differentiate people into different groups, we quickly begin to notice how those groups differ, and we form beliefs about the attributes that are characteristic of each group. Those beliefs, once formed, comprise the stereotype of the group. In this way, stereotyping derives from the foundation provided by the categorization process.

A related function of this categorization process is that the resulting stereotypes provide us with useful information. When encountering a member of a stereotyped group, the perceiver knows little about the individual. Yet by drawing on the stereotype, the perceiver can go beyond the information literally available and can infer a variety of attributes that are then

assumed to be true of the person, given his or her group membership. In this way, the perceiver has expanded his or her "knowledge" of the person. However, this process also entails the loss of information because in perceiving this person as being typical of members of the stereotyped group, the perceiver ignores information about this particular individual. Thus, through mental shortcuts, categorization and stereotyping allow us to preserve cognitive resources for other activities but they do so at potential cost.

In summary, each of the three conceptual orientations discussed previously approaches the study of stereotypes from a different perspective. Despite their different emphases and foci, elements of each approach undoubtedly contribute to stereotype formation and use.

### 3. STEREOTYPE ACTIVATION

In our everyday lives we often use stereotypes in making inferences and judgments about others. What factors influence whether stereotypes will be activated in the first place? In many cases, stereotypes may be activated simply when one encounters a member of a stereotyped group. In other words, the encounter may make certain beliefs more accessible in our minds, and these beliefs may then influence our perceptions of the person. For example, meeting an old person at a crosswalk might activate the stereotype "slow," whereas meeting an old person teaching on a college campus might activate the stereotype "wise." Also, there are situations in which one's membership in a group becomes salient. For example, the only woman on a committee of 12 people becomes salient, and that salience can make stereotypes about gender especially prevalent in that situation. Also, for some people stereotypes about certain groups can become chronically accessible so that the stereotype is routinely used in judging group members. Finally, research has shown that stereotypes may be activated automatically, even without the perceiver's awareness, and then can influence judgments and behaviors. Thus, stereotypes can be activated in a variety of ways, and in all cases they are likely to guide subsequent perceptions and interactions.

### 4. MEASURING STEREOTYPES

Research on stereotype measurement began in a classic study by Katz and Braly in 1933. Participants were given a lengthy list of trait adjectives and for each of

several groups (racial, national, and religious) were asked to select all of the words that were characteristic of the group. The adjectives that were most frequently selected were interpreted as representing the stereotype of that group. Follow-up studies using the same method have been conducted to assess the degree of stability or change of various stereotypes. This research has found, for example, that the characteristics selected as descriptive of African Americans have become less negative over the years. Does this mean that the stereotype has changed such that Americans now view African Americans more positively?

Perhaps this is so, but, unfortunately, it is difficult to draw such a conclusion with certainty. Although many researchers have used Katz and Braly's method over the years, it has some flaws that limit its validity. One of the most obvious limitations is that the checklist measures people's self-reports, and these self-reports are subject to bias. For example, when people report their beliefs and opinions, it is unclear whether they are indicating what they privately believe about African Americans, what other people think about African Americans, or what they believe is socially appropriate to say about African Americans. One explanation for the finding that negative stereotypes toward African Americans have decreased is that people are being influenced by social desirability concerns (i.e., they do not want to appear racist). To the extent that such concerns influence these measures, they do not provide accurate measures of people's stereotypes.

Recent studies measuring stereotype content have attempted to overcome some of these limitations by stressing that they are interested in the way society views members of various categories. Other research has assessed people's beliefs about various subtypes within broader social categories (e.g., within the category of women, one could separately assess beliefs about housewives, businesswomen, old maids, etc.). Despite these improvements, there still exists some ambiguity in the interpretation of the resulting data.

These concerns have prompted researchers to devise more subtle and indirect ways of measuring stereotypes. These methods often involve subliminally presenting a picture or words related to the stereotype (activating or "priming" the stereotype) and then assessing whether the participants respond to another task in stereotype-consistent ways. For example, studies have primed the stereotype of African Americans and have then shown that participants take less time to recognize negative (compared to positive) words. Other studies have shown that they make more stereotypic completions of

word fragments (e.g., *\_a\_y* = lazy). These and other indirect techniques have been used to assess people's thoughts about various groups without being subject to social desirability biases.

In summary, stereotypes may be measured in a number of ways, each of which has its strengths and limitations.

## 5. WHEN DO WE USE STEREOTYPES?

From categorization processes to automatic activation, stereotypes are used by everyone at some time or another. However, what determines whether or not we will use stereotypes when forming impressions of another person? Results of a number of studies on this topic have determined that stereotypes are most likely to influence impressions of another person when that person's behavior is ambiguous and therefore is open to stereotype-congruent interpretations. For example, seeing someone spend many hours studying for an exam, a perceiver's interpretation of that person's effort might be guided by racial stereotypes. If the student is Asian, then the extensive studying might be attributed to a desire to maintain a high grade point average. On the other hand, if the student is African American, then the perceiver might assume that the person is struggling to pass the class and therefore must study hard.

In contrast to this situation in which there are multiple possible interpretations of the behavior, there are other times when one's behavior is a direct indication of his or her underlying traits or dispositions. For example, a person who receives very high scores on the GRE exam would be seen as being intelligent, regardless of the person's race, gender, and so on.

The results of this research, then, have shown that stereotypes influence impressions of others when the behavior in question is ambiguous and open to stereotype-consistent interpretations. However, their influence on our construal of behavior is not the only consequence of using stereotypes.

## 6. CONSEQUENCES OF STEREOTYPING

### 6.1. Loss of Individual Information

Thinking of stereotypes as mental shortcuts seems not only efficient and useful but also fairly harmless.

However, holding stereotypic beliefs about group members can have a number of important, and often negative, consequences. On a cognitive level, stereotypes help us group similar others into meaningful categories. Although useful, the mere act of placing people into a category means that you overgeneralize and lose individuating information about them. In other words, the individual is perceived as having all of the characteristics stereotypic of the group and thereby loses anything that might differentiate the person as a unique individual.

At the group level, people tend to view members of the out-group as more similar to each other than members of the in-group (known as the out-group homogeneity effect). Perceiving homogeneity in a group is an important step in stereotyping because if all group members are seen as very similar, then it becomes easy to overgeneralize about them. Also, as already mentioned, overgeneralization is one of the hallmarks of stereotyping. In contrast, members of one's in-group retain their individuality and are seen as distinct from each other. A consequence of this in-group/out-group distinction is that people have a tendency to favor their own group (in-group bias) rather than the out-group. This in-group bias then becomes harmful if it results in prejudice toward out-group members or in favoritism toward in-group members based solely on their membership in the in-group.

### 6.2. Behavioral Implications

Once activated, stereotypes may influence both our behavior and our interpretation of another's actions. As mentioned previously, stereotypes act as expectancies that influence our construal of another's ambiguous behavior. In addition, they may unconsciously influence us to behave in ways that are consistent with the stereotype. For example, several studies have shown that priming the African American stereotype leads participants to behave in a more hostile way and to interpret another person's behavior as more aggressive. Subtly priming young people with the stereotype of the elderly can lead those young people to move more slowly. These effects become a very real problem when, even unconsciously, our beliefs about group members influence both our behavior toward them and, in turn, their subsequent behavior toward us.

### 6.3. Self-Fulfilling Prophecy

The process by which one's beliefs about a person lead that person to behave in ways that confirm one's

original expectations is called the self-fulfilling prophecy. The importance of the self-fulfilling prophecy was highlighted in a 1968 study by Rosenthal and Jacobsen. They led teachers to believe that a randomly chosen sample of children in their classes were very intelligent and were about to blossom. The research showed that the teachers then treated those students in ways that helped the students perform well, thereby confirming the teachers' expectations. Thus, the teachers' experimentally induced expectations guided their own behavior, which in turn produced the expected outcome in the students' performance.

These effects have been replicated in a number of other situations (including interactions between males and females and between blacks and whites). In each case, people's stereotypic beliefs about a target person influenced that person's behavior, thereby confirming the people's original expectations. These studies stress the important role that stereotypes play in several aspects of the interpersonal context. They not only can influence perceptions of behavior and guide the perceiver's own behavior but also can influence the stereotyped person to behave in a stereotype-confirming manner.

## 7. STEREOTYPE CHANGE

With the knowledge that stereotypes are often based on incomplete overgeneralizations, it seems that one obvious way to change people's stereotypes is to put them into contact with the very people they are stereotyping. Specifically, by interacting with out-group members people would have ample opportunity to notice behavior that was inconsistent with the group stereotype, thereby resulting in stereotype change. This premise is exactly what Allport's 1954 contact hypothesis is based on. Although the idea is simple, there are certain factors that influence whether or not contact leads to stereotype change. For example, if a certain white individual holds negative stereotypes about Latinos, then, according to the contact hypothesis, those stereotypes should be reduced by having the individual interact with Latinos in a supportive, friendly environment. In addition, stereotype change is facilitated when social institutions support the intergroup contact and when the two groups are believed to share equal status in some domain. Clearly, not every interaction with members of the stereotyped group will lead to change. In fact, situations in which the contact involves competition, members of one or both groups are frustrated, and members of the

two groups hold different morals or values will not be conducive for reducing stereotypes.

In the multicultural society in which we live, people of different races seem to be interacting with increasing frequency, yet stereotypes persist. One possible explanation for this persistence is that people rationalize inconsistent behaviors by stereotyped group members. For example, if someone meets a successful Latino businessman, then this might go against his or her stereotypic expectation that Latinos are lazy and unintelligent. However, rather than changing their stereotypes, people may instead group people that do not “fit” the stereotype into their own separate category or subtype. By placing those who display stereotype-inconsistent behavior into subtypes, people maintain the status quo and resist changing their stereotypes of the group as a whole.

Although the picture painted previously seems grim, researchers have developed other promising cognitive interventions that have been shown to decrease the use of stereotypes in certain situations. For example, encouraging people to think carefully about others before passing judgment forces them to focus on individual-level attributes rather than category-level information. This approach encourages people to look beyond group memberships and instead to see people as unique individuals. Moreover, if the motivation and cognitive resources are available, people are able to control their use of stereotypes in many situations.

## 8. CONCLUSION

Stereotypes are multifaceted, complex phenomena that influence us all in one way or another. At their most basic level, stereotypes enable us to effortlessly separate people into distinct groups, thereby allowing us to make inferences about them. In the process of categorizing people into groups we may gain information, but we also overlook unique information about individuals. In addition to categorization, stereotypes play a number of dual roles. They may be positive or negative, and they often lead to in-group favoritism or out-group derogation. Despite the wide array of consequences from the stereotyping process, stereotypes can be controlled and changed if circumstances permit. Stereotypes are a fundamental human resource but may become harmful if categorization and subsequent overgeneralizations lead to prejudice and discrimination.

## See Also the Following Articles

Groupthink ■ Intergroup Relations and Culture ■ Prosocial Behavior, Development of

## Further Reading

- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Allport, G. W. (1954). *The nature of prejudice*. New York: Addison-Wesley.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 20–33.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, 46, 735–754.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60, 509–517.
- Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 2, pp. 1–68). Hillsdale, NJ: Erlbaum.
- Katz, D., & Braly, K. W. (1933). Racial stereotypes of 100 college students. *Journal of Abnormal Social Psychology*, 28, 280–290.
- Lippman, W. (1922). *Public opinion*. New York: Harcourt Brace.
- Rosenthal, R., & Jacobsen, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart, & Winston.
- Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, 35, 656–666.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel, & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The non-verbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10, 109–120.



# Stress

Charles J. Holahan and Jennifer D. Ragan

University of Texas, Austin, Texas, USA

Rudolf H. Moos

Department of Veterans Affairs and Stanford University,  
Menlo Park, California, USA

1. Introduction
  2. Categories of Life Stressors
  3. Health Effects
  4. Stress Moderators
  5. Conclusions
- Further Reading

## GLOSSARY

**adaptive coping** Cognitive and behavioral efforts to manage stressful conditions or associated emotional distress; adaptive coping efforts operate as protective factors that decrease the adverse effects of life stressors when they occur and that can also reduce the likelihood of stressor occurrence.

**chronic stressors** Stressors that endure over time (e.g., social role strains, everyday hassles).

**hypothalamic–pituitary–adrenocortical axis** Mechanism mediated by the pituitary gland, which during stressful experiences stimulates the adrenal cortex to release glucocorticoids that suppress key components of the immune system when chronically expressed.

**negative life events** Acute life change events that occur at discrete points in time and that are objectively rated as negative (e.g., death of family members, financial setbacks).

**personal diatheses** Long-standing biological or psychological vulnerabilities that interact with life stressors and increase initial health risk and that can also impede the mobilization of social resources and adaptive coping efforts; they may reflect both hereditary and developmental influences.

**social resources** Emotional support and assistance from family members and friends; these resources operate as

protective factors that decrease the negative health effects of life stressors when they occur and that can also reduce the likelihood of stressor occurrence.

**stress moderators** Factors that increase or decrease the health effects of life stressors by functioning as either vulnerability factors or protective factors when stressors occur.

**sympathetic arousal** Mechanism mediated by the sympathetic nervous system, which during stressful experiences stimulates the release of catecholamines that mobilize the body to deal with immediate adaptive demands.

**traumatic stressors** Qualitatively more severe stressors than those indexed on standard life event or role strain scales (e.g., being exposed to war or a natural disaster, being the victim of a violent crime).

Life stressors are linked to both psychological and physical illnesses. Early stress studies focused on the direct negative effects of stressors on psychological and physical health, overlooking individual variability in response. However, later studies have emphasized factors that moderate the potential health effects of life stressors by functioning as either vulnerability factors or protective factors when stressors occur.

## 1. INTRODUCTION

The stress process is initiated by environmental factors (stressors) that place adaptive demands on individuals. Contemporary understanding of the physiology of stress

owes much to the groundbreaking research of Cannon on the role of the sympathetic nervous system in mobilizing the body for adaptive challenge and of Selye on the role of corticosteroids in producing the adverse biological effects of chronic stress. Stress research during the 1960s and 1970s extended early findings based on laboratory animals to humans, presuming a straightforward link between life change and pathology that overlooked individual variability in response. In fact, life stressors are linked to both psychological and physical illnesses. However, despite the consistency of stressor effects, early stress research was characterized by persistent anomalies. The amount of variance predicted in distress was typically less than 10%, and individuals showed highly variable reactions to stressors. Stressors made some individuals ill, others stayed healthy, and some grew stronger after managing challenges successfully.

Researchers recognized that these apparent anomalies reflected limitations in conceptualizing and measuring stressors and stress reactions; however, only slowly did they come to understand that these anomalies also represented important findings in their own right. The relatively poor empirical predictions of early stress-illness studies led researchers to focus increasingly on factors that moderate the health effects of life stressors. Contemporary research assumes that the human stress response is inherently complex, reflecting a dynamic interplay among stressors and vulnerability and protective factors in the individual and in the social environment.

## 2. CATEGORIES OF LIFE STRESSORS

Stress studies investigate (a) life change events, (b) major life crises and traumas, and (c) chronic stressors (Table I). It should be noted that although life change events have an acute onset, they often initiate chronic stressors such as persisting economic and family problems. In fact, these enduring secondary stressors may account for many of the adverse effects of life change events.

### 2.1. Life Change Events

Early stress studies focused on acute life change events that occur at discrete points in time such as deaths of family members and financial setbacks. A common measurement instrument was Holmes and Rahe's Social Readjustment Rating Scale (SRRS). The SRRS is a self-report questionnaire that uses objectively weighted life change units to index the relative level of adaptive demand associated with a particular life event, regardless of

TABLE I  
Examples of Life Stressors by Measurement Category

<i>Life change events</i>	<i>Traumatic stressors</i>	<i>Chronic stressors</i>
Death of spouse	War exposure	Family role strains
Death of friend	Victim of violent crime	Work role strains
Divorce	Natural disaster	Discrimination
Fired at work	Technological disaster	Poverty
Financial setback	Serious medical illness	Caregiving

whether the event is positive or negative. These weighted life change units are summed to reflect cumulative life change during a given period of time, typically the previous 12 months. Later research revealed that the adverse health effects of life change are linked to negative life events rather than positive events. Later studies also showed that weighted life change units contributed little beyond an unweighted sum of events experienced in predicting illness. Based on these findings, most contemporary research on life change events indexes the sum of negative events experienced during a given period of time.

### 2.2. Major Life Crises and Traumas

Stress researchers have also studied the adverse health effects of traumatic stressors that are qualitatively more severe than the stressors indexed on standard life event or role strain scales. Initially, traumatic stressors were defined as personal experience of life-threatening stressors such as being exposed to war and being the victim of a violent crime. Also included were being exposed to natural disasters (e.g., tornados, earthquakes) and being diagnosed with life-threatening medical illnesses (e.g., heart disease, cancer). Over time, the definition of trauma has been broadened to encompass witnessing events that place others at risk of death or serious injury or witnessing individuals being badly injured or killed by such events. Interest in traumatic stressors is evidenced in a growing body of research on posttraumatic stress disorder (PTSD).

### 2.3. Chronic Stressors

A major limitation of early stress research was its narrow preoccupation with laboratory studies of acute stressors.

Eventually, recognizing that many life stressors are ongoing, investigators began to examine chronic stressors that endure over time such as social role strains and everyday hassles. Researchers studying social role strains often measure chronic family or work difficulties. Some stress studies have focused on additional burdens that often are associated with family life such as caregiving for chronically ill family members. For example, increasing interest is being devoted to caregivers of aging family members suffering from Alzheimer's disease. Typical measurement instruments include Moos's Work Environment Scale and Family Environment Scale, which measure the social climate at work (e.g., work pressure) and in the family (e.g., family conflict), respectively. Moreover, there has been increasing research interest in role strains associated with the interface between work and family lives. Investigators have also studied minor life stressors that are experienced chronically such as everyday hassles with troublesome neighbors and traffic problems. Increasingly, stress studies are examining the health consequences of ongoing "macrostressors" such as discrimination, neighborhood violence, poverty, and terrorism.

### 3. HEALTH EFFECTS

Life stressors inevitably produce some emotional strain and physical tension. Stress researchers have posed the question of whether life stressors are also associated with more significant illnesses. In fact, life stressors are linked to a number of psychological and physical illnesses. Psychological and physical illnesses are often linked reciprocally, with each category of illness exacerbating the other. Moreover, psychological and physical illnesses often function as life stressors themselves, initiating a new cycle in the stress process.

#### 3.1. Psychological Illness

Life stressors are associated with psychological stress reactions that involve depression and anxiety. For example, life stressors are linked both to the onset of depressive disorders and to relapse among individuals recovering from depressive disorders. Interpersonal problems and losses are especially likely to be associated with depressive reactions. Life stressors can also precipitate both onset and relapse of anxiety disorders such as generalized anxiety disorder, panic disorder, agoraphobia, and obsessive-compulsive disorder, and they can play a role in the development and progression of

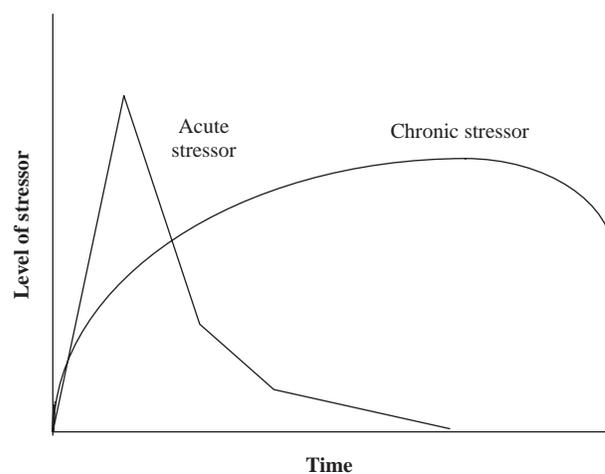
alcohol and drug abuse. Moreover, life stressors can trigger schizophrenic episodes among individuals who are vulnerable to this disorder. Trauma exposure produces a recognized pattern of PTSD symptoms, including reexperiencing the trauma psychologically through flashbacks and nightmares, emotional numbing, and experiencing heightened arousal and vigilance.

Exclusive of traumatic events, chronic stressors are more strongly linked to psychological distress than are acute events. Although chronic stressors generally are less severe than acute life events, their effects often last longer and are more pervasive (Fig. 1). Moreover, an event is more likely to have an adverse psychological outcome when it threatens or disrupts a domain in which a person has central commitments.

Psychological reactions themselves can exacerbate the stress process in two important ways. First, psychologically distressed individuals are more likely to perceive benign situations as threatening, and these perceptions of threat can trigger additional stress reactions. Second, persons who are psychologically distressed often create social conditions in their lives, such as conflictual family or work relationships, that are likely to produce new life stressors.

#### 3.2. Physical Illness

Stressor exposure also initiates a characteristic biological response that is associated with the onset or exacerbation of a wide spectrum of physical illnesses. The biological stress response involves interconnections among the nervous, endocrine, and immune systems.



**FIGURE 1** The time course of acute stressors versus chronic stressors. Adapted from Wheaton (1994).

The two most heavily studied stress-related biological mechanisms have been sympathetic arousal and activation of the hypothalamic–pituitary–adrenocortical (HPA) axis. Both mechanisms involve initial central nervous system input from the hypothalamus at the base of the brain, and both mechanisms operate through the adrenal glands located above the kidneys (Fig. 2). Short-term activation of these biological mechanisms is adaptive, mobilizing energy and enhancing alertness to respond to adaptive demands. However, chronic stimulation of these mechanisms can lead to (a) hyperarousal and subsequent wear and tear on body systems and (b) suppression of key components of the immune system.

Sympathetic arousal is mediated by the sympathetic nervous system, which during stressful experiences releases norepinephrine at multiple sites throughout the body and stimulates the adrenal medulla to release epinephrine (adrenaline). These catecholamines mobilize the body to deal with immediate adaptive demands. However, chronic sympathetic arousal is associated with tension-related complaints such as headache. In fact, stress is the triggering factor reported most often by migraine and tension-type headache patients. Sustained or frequent intense sympathetic arousal is also linked to wear and tear on arteries and coronary vessels and to shear stress associated with sharply increasing catecholamines. Chronic sympathetic arousal also increases blood clotting through coronary vasoconstriction, increased circulating lipids, and increased platelet aggregation.

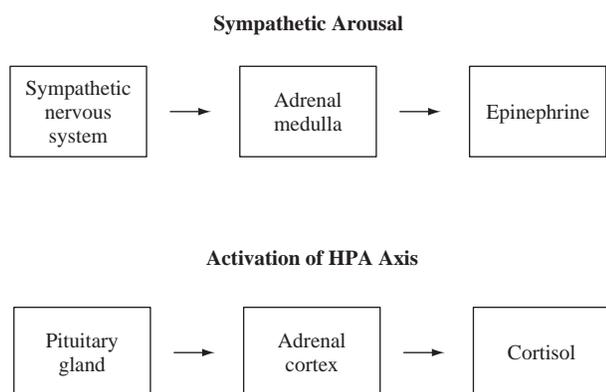
Hemodynamic and biochemical changes, in turn, may lead to health-related problems such as atherosclerosis,

hypertension, and coronary artery disease. For example, occupational stressors have been linked to elevated blood pressure in studies of activity at work. Moreover, stressful work conditions, particularly the combination of high work demands and low job control, have been associated with coronary heart disease in population studies of workers in Europe and the United States.

Activation of the HPA axis is mediated by the pituitary gland, which during stressful experiences stimulates the adrenal cortex to release glucocorticoids (primarily cortisol in humans). Chronic expression of glucocorticoids and excessively high levels of glucocorticoids have been associated with suppression of key components of the immune system in human *in vitro* studies (i.e., examining tissue or blood samples outside the body) and in experimental studies using animal models. For example, glucocorticoids are linked to reduced T cell proliferation and lower natural killer cell cytotoxicity. In addition, glucocorticoids can impede the maturation of developing lymphocytes and can destroy mature lymphocytes. These changes inhibit the ability of the cellular arm of the immune system to attack target pathogens directly. Glucocorticoids are also linked to decreased production of antibodies and certain components of complement. These changes inhibit the ability of the humoral arm of the immune system to target pathogens for destruction. It is unclear, however, whether the levels of immune system down-regulation observed in the context of stressful experiences are sufficient to increase vulnerability to disease.

In studies examining clinical illness, there has been evidence of a link between life stressors and the onset and progression of infectious diseases such as the common cold and influenza. Life stressors are also related to latent viral activity in the herpes viruses, resulting in complaints such as mononucleosis and cold sores. In addition, life stressors may be associated with the onset and exacerbation of some autoimmune diseases, such as rheumatoid arthritis, and with poor diabetic control and progression of diabetes. Experimental research with animals has demonstrated that stress can increase the development of artificially induced tumors; however, there is no clear evidence that stress can accelerate the growth of common human tumors.

Historically, stress research has emphasized the link between sympathetic arousal and cardiovascular response as well as that between the HPA axis and immune system response. However, there is increasing evidence that sympathetic arousal also affects the immune system and that the HPA axis also affects the cardiovascular system. In addition to these direct



**FIGURE 2** Two stress-related biological mechanisms mediating between central nervous system input and health outcomes. HPA, hypothalamic–pituitary–adrenocortical axis.

biological pathways between stressors and illness, life stressors can relate to physical illness indirectly through health risk behaviors. Life stressors are associated with common health risk behaviors such as increased cigarette smoking, alcohol abuse, poor dietary and exercise habits, disturbed sleep, and reduced adherence to medical regimens. In turn, each of these health risk behaviors plays a role in increasing the risk of physical illness.

#### 4. STRESS MODERATORS

Individuals have highly variable reactions to stressors, with some individuals showing little or no health effects from stressor exposure and others showing profound reactions. From an initial emphasis on a straightforward link between life stressors and illness, contemporary stress studies have evolved to placing increasing emphasis on stress moderators. Moderating factors increase or decrease the health effects of life stressors by functioning as either vulnerability factors or protective factors when stressors occur.

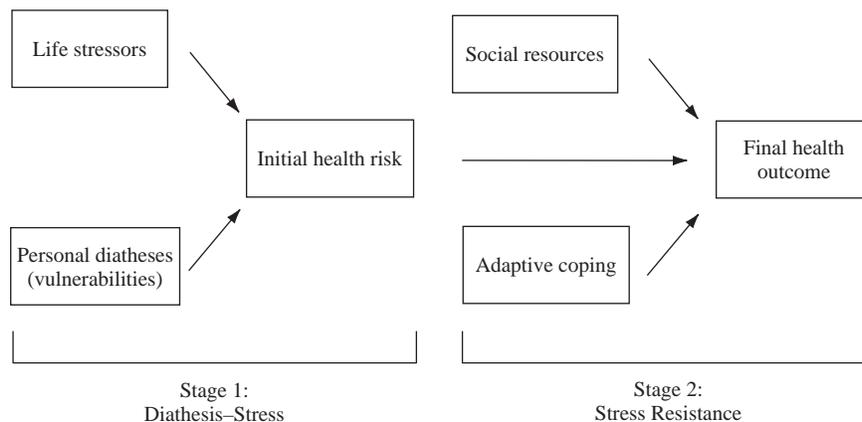
More recent stress research has been guided by two models that emphasize the key role of stress moderation: the diathesis–stress model and the stress resistance model. Figure 3 shows an integrative framework of key components of these two models. The diathesis–stress model focuses on long-standing vulnerability factors (i.e., personal diatheses) and emphasizes the first stage of the stress process where personal vulnerabilities interact with life stressors and increase initial health risk. The stress resistance model focuses on protective factors (i.e., social resources and adaptive coping

efforts) and emphasizes the second stage of the stress process where protective factors mobilized after stressor exposure modify initial health risk and determine the final health outcome. The figure is simplified in that arrows are shown only for primary relationships. Personal vulnerabilities can affect the second stage of the stress process by impeding the mobilization of social resources and adaptive coping efforts, whereas protective factors can affect the first stage of the stress process by reducing the likelihood of stressor occurrence.

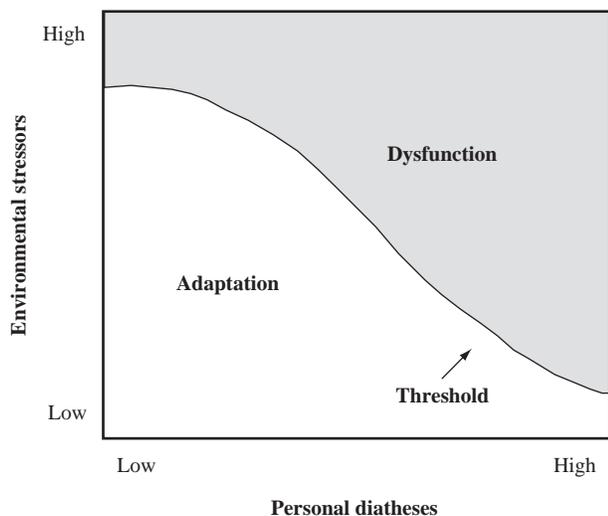
##### 4.1. Personal Diatheses

Personal diatheses are long-standing biological or psychological vulnerabilities that interact with life stressors and increase initial health risk (Fig. 4). Personal diatheses can also impede the mobilization of social resources and adaptive coping efforts. Thus, adaptation can be understood in terms of the match between the level of life stressors and individuals' personal vulnerabilities. For example, individuals with many personal vulnerabilities typically adapt poorly except in settings with exceptionally low life stressors. Conversely, individuals with few personal vulnerabilities typically adapt well except in settings with exceptionally high life stressors. Individuals with an intermediate level of personal vulnerabilities typically adapt poorly in circumstances that tax their particular areas of fragility.

Personal diatheses reflect both hereditary and developmental influences and can affect both psychological and physical health outcomes under stress. For example, excessive cardiovascular reactivity (a biological



**FIGURE 3** Integrative framework of key components of the diathesis–stress and stress resistance models. For simplicity, arrows are shown only for primary relationships.



**FIGURE 4** A matching model predicting psychological adjustment as a function of both environmental demands and personal vulnerability. Adapted from Zubin, J., & Spring, B. (1977). Vulnerability: A new view of schizophrenia. *Journal of Abnormal Psychology*, 86, 103–126. Copyright 1977 by the American Psychological Association. Reprinted by permission.

vulnerability) and hostility/cynicism (a psychological vulnerability) increase the risk of coronary artery disease under stress. Correspondingly, abnormalities in neurotransmitter systems, such as the serotonergic system in the brain (a biological vulnerability) and a proclivity to attribute failure to personal deficits (a psychological vulnerability), increase the risk of depressive disorders under stress. Although the diathesis–stress model emphasizes vulnerabilities, individuals also bring long-standing personal strengths to the stress process. Especially important are personality characteristics, such as dispositional optimism and self-efficacy, that enhance perceptions of personal control during recovery from stressors and that can also reduce the likelihood of stressor occurrence. The study of protective personality strengths is encompassed under the stress resistance model along with social resources and adaptive coping.

#### 4.2. Social Resources

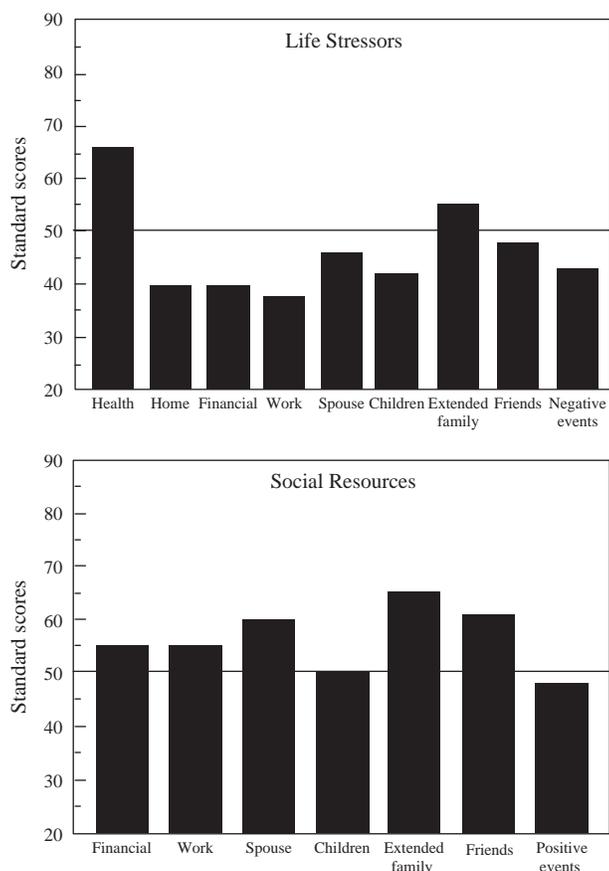
Social resources include emotional support as well as guidance and assistance from family members and friends. Social resources operate as protective factors that decrease the negative health effects of life stressors when they occur and that can also reduce the likelihood of stressor occurrence (Fig. 5). Social resources can

reduce both adverse psychological health outcomes, such as depression after bereavement, and adverse physical health outcomes, such as job-related hypertension. Some aspects of social support, such as social companionship, may be related to better health and well-being regardless of the level of stressors experienced. Other aspects of social support, such as informational support, may function as “stress buffers” that have stronger positive effects on health when individuals are experiencing a high level of stressors. Although the stress resistance model emphasizes resilience, social relationships can also present health risks. In fact, negative aspects of relationships, such as criticism or disapproval from family members or friends, are as strongly related to poorer health under stress as positive aspects of relationships are related to better health under stress.

Recently, stress studies have examined how life stressors can deplete social resources. In part, such depletion may be tangible in that negative life events, such as the deaths of family members or friends, entail reductions in social resources. However, resource depletion also may have a cognitive component in that perceived support is vulnerable to deterioration after life crises. Moreover, resource loss demonstrates another way in which life stressors are linked to adverse psychological outcomes. For example, individuals recovering from a natural disaster might feel that the support they received from friends was less than what they had expected. This perception of resource loss mediates part of the link between crisis exposure and emotional distress.

#### 4.3. Adaptive Coping

Adaptive coping includes cognitive and behavioral efforts to manage stressful conditions or associated emotional distress. Like social resources, adaptive coping operates as a protective factor that decreases the adverse effects of life stressors when they occur and that can also reduce the likelihood of stressor occurrence. People who rely more on active coping strategies, such as problem solving and information seeking, tend to adapt better to life stressors. Although the stress resistance model emphasizes resilience, coping styles may also involve vulnerabilities. For example, avoidance coping strategies, such as denial and wishful thinking, are generally associated with increased psychological distress. In addition, coping styles can affect physical health indirectly through these psychological reactions. Moreover, coping styles are linked to physical health indirectly through health risk behaviors



**FIGURE 5** A Life Stressors and Social Resources profile for a woman with rheumatoid arthritis. Moos and Moos's Life Stressors and Social Resources Inventory (LISRES) provides a picture of the overall balance between an individual's life stressors and social resources across representative life domains. The inventory consists of nine indexes of life stressors and seven indexes of social resources. The figure shows an illustrative LISRES profile for a 66-year-old woman with rheumatoid arthritis (scores are standardized, with a mean of 50 and a standard deviation of 10). Although she reported well above-average physical health stressors, she had below-average stressors in six of the other seven domains. Moreover, she also had above-average social resources in her work setting and in her relationships with her spouse, extended family, and friends. At an 18-month follow-up, this favorable balance of moderate stressors and high resources enabled her to manage quite well. She reported high self-confidence and below-average depression. Adapted from Moos, R. H., & Moos, B. S. (1994). *Life Stressors and Social Resources Inventory: Adult form manual*. Lutz, FL: Psychological Assessment Resources. Reproduced by special permission of the publisher, Psychological Assessment Resources Inc., 16204 North Florida Avenue, Lutz, FL 33549, from the LISRES-A by Rudolf Moos, Ph.D., and Bernice Moos, copyright 1994 by PAR Inc. Further reproduction is prohibited without permission of PAR Inc.

such as cigarette smoking, alcohol use, and lack of adherence to medical regimens.

Based on these observations, most conceptualizations of coping emphasize the distinction between strategies that are oriented toward approaching and confronting the problem and strategies that are oriented toward reducing tension by avoiding dealing directly with the problem (Table II). Lazarus and

Folkman proposed a related typology of coping processes that distinguishes between problem-focused strategies (i.e., efforts to modify the source of the stress) and emotion-focused strategies (i.e., attempts to regulate the emotional distress caused by a stressor). Emotion-focused strategies can be oriented toward dealing with a stressor; however, such coping is most often directed toward avoiding dealing with the source

**TABLE II**  
**Basic Coping Categories with Associated Coping Subtypes from Moos's Coping Responses Inventory**

<i>Basic coping category</i>	<i>Coping subtypes</i>
Cognitive approach coping	Logical analysis (e.g., "Did you think of different ways in which to deal with the problem?")
	Positive reappraisal (e.g., "Did you think about how you were much better off than other people with similar problems?")
Behavioral approach coping	Seeking guidance and support (e.g., "Did you talk with a friend about the problem?")
	Taking problem-solving action (e.g., "Did you make a plan of action and follow it?")
Cognitive avoidance coping	Cognitive avoidance (e.g., "Did you try to forget the whole thing?")
	Resigned acceptance (e.g., "Did you lose hope that things would ever be the same?")
Behavioral avoidance coping	Seeking alternative rewards (e.g., "Did you get involved in new activities?")
	Emotional discharge (e.g., "Did you yell or shout to let off steam?")

*Source.* From Moos, R. H. (1993). *Coping Responses Inventory: Adult form manual* (3rd ed.). Lutz, FL: Psychological Assessment Resources. Adapted and reproduced by special permission of the publisher, Psychological Assessment Resources Inc., 16204 North Florida Avenue, Lutz, FL 33549, from the Coping Responses Inventory by Rudolf Moos, Ph.D., copyright 1993 by PAR Inc. Further reproduction is prohibited without permission of PAR Inc.

*Note.* Sample coping responses are in parentheses.

of stress. Although one can draw overall conclusions about the relative efficacy of approach and avoidance coping strategies, such generalizations necessarily oversimplify the process of adaptation. Individuals adapt best when their coping efforts match situational demands. Individuals who are flexible in their choice of coping should show better adaptation than do persons who have more restricted or rigid coping repertoires. For example, avoidance coping can be adaptive in the short term with time-limited stressors and with stressors that are not controllable.

Stress studies have begun to develop an integrated understanding of how social resources and adaptive

coping are related to one another. Social resources are linked to health under high stressors in part because they encourage more adaptive coping strategies. For example, social resources can enhance coping efforts by bolstering feelings of self-confidence as well as by providing informational guidance that aids in appraising threat accurately and in planning effective responses.

## 5. CONCLUSIONS

Considerable evidence documents that life stressors are associated with both psychological and physical illnesses. However, the overall association between life stressors and illness is only small to moderate. Moreover, health effects are strongest with chronic stressors and severe stressors, which often precipitate associated ongoing demands. Even more important, the profound individual differences in stressor effects highlight the complexity of the stress process. At initial stressor exposure, potential health risk is moderated by long-standing biological and psychological vulnerabilities. Further moderation occurs after stressor exposure as social resources and individual coping strengths are brought into play to manage adaptive demands.

Finally, despite the pain of stressful experiences, some individuals are able to grow stronger under stress. For example, physiological resilience can be strengthened under stress. Physiological strengthening is most likely when stressors are acute rather than chronic and when periods of relaxation provide restorative intervals between stressor exposures. Similarly, life crises can be "constructive confrontations" that provide an opportunity for psychological growth. Individuals can emerge from a crises with greater self-confidence, new coping skills, closer relationships with family members and friends, and a richer appreciation of life.

## See Also the Following Articles

Coping ■ Posttraumatic Disorders

## Further Reading

- Baum, A., & Posluszny, D. M. (1999). Health psychology: Mapping biobehavioral contributions to health and illness. *Annual Review of Psychology*, 50, 137-163.
- Cohen, S., & Herbert, T. B. (1996). Health psychology: Psychological factors and physical disease from the

- perspective of human psychoneuroimmunology. *Annual Review of Psychology*, 47, 113–142.
- Coyne, J. C., & Downey, G. (1991). Social factors and psychopathology: Stress, social support, and coping processes. *Annual Review of Psychology*, 42, 401–425.
- Kessler, R. C. (1997). The effects of stressful life events on depression. *Annual Review of Psychology*, 48, 191–214.
- Krantz, D. S., & McCeney, M. K. (2002). Effects of psychological and social factors on organic disease: A critical assessment of research on coronary heart disease. *Annual Review of Psychology*, 53, 341–369.
- Lazarus, R. S. (1993). Coping theory and research: Past, present, and future. *Psychosomatic Medicine*, 55, 234–247.
- Moos, R. (1994). *The social climate scales: An overview*. Palo Alto, CA: Consulting Psychologists Press.
- Sapolsky, R. M. (2002). Endocrinology of the stress response. In J. B. Becker, S. M. Breedlove, D. Crews, & M. M. McCarthy (Eds.), *Behavioral endocrinology* (2nd ed., pp. 409–450). Cambridge, MA: MIT Press.
- Selye, H. (1993). History of the stress concept. In L. Goldberger, & S. Breznitz (Eds.), *Handbook of stress: Theoretical and clinical aspects* (2nd ed., pp. 7–17). New York: Free Press.
- Wheaton, B. (1994). Sampling the stress universe. In W. R. Avison, & I. H. Gotlib (Eds.), *Stress and mental health: Contemporary issues and prospects for the future* (pp. 77–114). New York: Plenum.



# Structure of Interests

Christy Hofsess and Terence J. G. Tracey

Arizona State University, Tempe, Arizona, USA

1. Evolution of the Structure of Interests
2. Extensions of the Circular Model: Spherical Structure
3. Relation of the RIASEC Types to Personality and Ability Measures
4. Current Research  
Further Reading

## GLOSSARY

**circumplex** Model in which the focus is on a pattern of relations in a matrix that forms a circular structure. Variables within a specific domain (e.g., vocational interest domain) can be understood and described by their placement on the circle; those more proximate are more highly related than those more distal.

**factor list** Model in which the focus is on the number and nature of independent factors that exist in a particular domain (e.g., vocational interest domain).

**hierarchical model** Model in which the focus is on the discrete sets of clusters within a domain (e.g., vocational interest domain).

**interest inventory** An assessment of one's interests for use in counseling and guidance.

**interests** Activities and occupations that an individual finds attractive.

**structure** How something is constructed, organized, or interrelated.

The structure of interests is a well-researched area within vocational psychology that focuses on how

individuals think about and conceptualize their interests. All theories are based on how interests are interrelated and aim to create a systematic framework for understanding the organization of interests, thus guiding interest inventory development and career counseling practice. Since the early 1950s, researchers have primarily developed theories categorizing and classifying the many types of interests and occupations individuals pursue into multiple factors or themes, and these are useful as organizing principles in inventory development and presentation. Interest inventories help individuals understand how their curiosities and involvement in different activities influence and shape career decisions. The focus on the structure of interests thus centers on gaining a better understanding of how individuals think about their interests and presenting more helpful information to the test taker.

## 1. EVOLUTION OF THE STRUCTURE OF INTERESTS

Much of the research on interests has been of the factor list variety as exemplified by the early work of Guilford in 1954 and others. Interest items were factor analyzed with the goal of yielding a parsimonious set of factors that characterize interests. This approach is useful but often results in numerous factors, making easy application somewhat difficult. As an alternative to the factor list approach, Roe introduced the first circular theory of

the structure of interests in 1956 by proposing a circular model to explain how interest types and occupations are related to each other. Her theory classified occupations into eight groups: service, business contact, organization, technology, outdoor, science, general culture, and arts and entertainment. These groups, referred to as fields, represent a common shared activity. Her theory arranged the fields in a circular manner so that similar fields are in close proximity. In addition to the circular arrangement of interests, she also introduced the concept of level of responsibility (analogous to complexity of tasks). Subsequent work by Tracey and Rounds demonstrated that the circular arrangement proposed by Roe does not have empirical support. However, Roe's model served as a stimulus to the circular model proposed by John Holland.

### 1.1. Holland's Model of Vocational Types

John Holland's theory of vocational interest types evolved into the dominant, standard organizing framework for interest assessment used by career counseling professionals. His theory classifies individuals and work environments according to six types: realistic (R), investigative (I), artistic (A), social (S), enterprising (E), and conventional (C), commonly referred to collectively as RIASEC. Although RIASEC types are primarily personality types, they typically are conceptualized and measured in terms of vocational interests. Holland posited these six types as existing in a hexagonal (or circumplex) structure with equal distance between the types, and proximity of the types representing the degree of similarity. His model is represented in Fig. 1. Individuals are categorized into the six types according to their personality dispositions. Similarly, work environments are characterized according to settings that evoke, develop, and reinforce the six personality interest types. The congruence of interest types and work environments is hypothesized to be related to satisfaction and productivity. The simplicity, elegance, and applicability of Holland's typology make it the most widespread theory of the structure of interests.

The prominence of Holland's theory has led to the incorporation of RIASEC scales into every major interest assessment instrument, such as the Strong Interest Inventory (SII), Self-Directed Search (SDS), American College Testing Program Interest Inventory (UNIACT), Kuder Occupational Interest Survey, Vocational

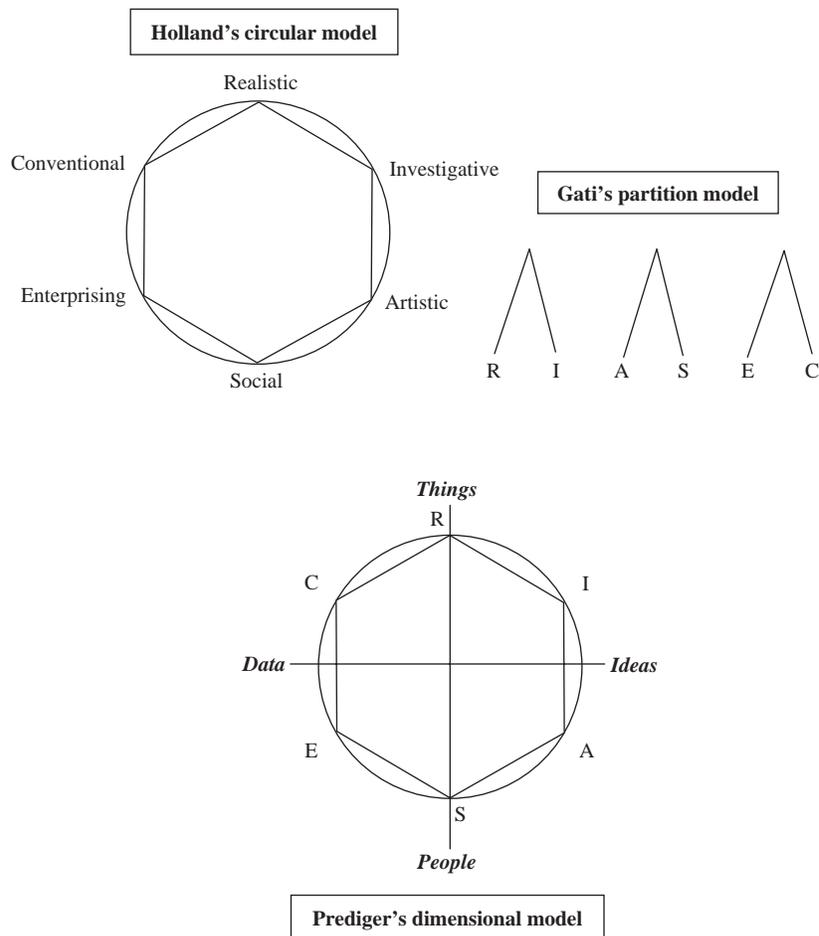
Preference Inventory (VPI), and the Career Assessment Inventory.

### 1.2. Alternative Approaches

Although Holland's model remains the most influential model within vocational psychology, two other notable structural models diverge from his hexagonal structure: the hierarchical model of Gati and the dimensional model of Prediger. Using the RIASEC types, Gati proposed a hierarchical model of the structure of vocational interests as a more accurate representation of the structure of the six types than the circular/hexagonal representation. His hierarchical model is also depicted in Fig. 1. Tracey and Rounds compared Gati's hierarchical model to Holland's circular model in a meta-structural analysis using 104 different RIASEC correlation matrices with a total sample size of 47,268 individuals. They found that Holland's circular model was better supported than was Gati's model, especially for U.S. samples. Subsequent work by these researchers on cross-cultural and ethnic minority U.S. samples provides some support for the viability of Gati's model in these contexts.

Prediger sought to explicitly label the underlying dimensions of the RIASEC circular model. His model, although an alternative approach, complements Holland's theory by demonstrating how the RIASEC circular structure contains two key underlying work task dimensions, people–things and data–ideas. The people–things dimension differentiates the social type from the realistic type, and the data–ideas dimension separates enterprising and conventional types from artistic and investigative types. His model is summarized in Fig. 1. Prediger posited that a focus on the two underlying dimensions rather than the six types provides a more parsimonious representation of an individual's interests and the work environment.

In a different meta-structural analysis, Rounds and Tracey compared the circular model to Prediger's dimensional approach. They confirmed that underlying Holland's circular structure are two dimensions, supporting Prediger's people–things/data–ideas model as complementing, not competing with, Holland's theory. The focus of Holland's structure is less on the underlying dimensions and more on the circular structure, yielding information about the relations among the interest types not revealed in the dimensional approach. Therefore, although the two dimensions are valid, they are viewed as supplementary because they are not as important as the circular structure.



**FIGURE 1** Various structural representations of Holland's RIASEC types (Holland circumplex, Gati partition, and Prediger dimensional). From Tracey, T. J. G. (1993). Evaluating Holland's and Gati's vocational interest models: A structural meta-analysis. *Psychological Bulletin*, 113, 229–246. Copyright © 1993 by the American Psychological Association. Adapted with permission.

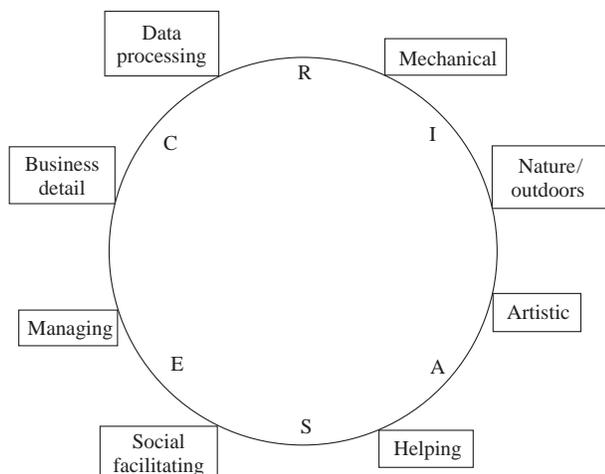
### 1.3. Generalizability of the Circular Structure

Most RIASEC interest score comparisons across gender, culture, and age focus on mean differences in scores and do not take into consideration variations that may exist in underlying interest structure. In an investigation of the structural invariance of the circular order model, Tracey and Rounds found no differences in gender or age (at least with respect to samples aged 14 years or older). The circular structure was also invariant across the four major interest inventories (VPI, SDS, UNIACT, and SII). As noted previously, the circular model was the best representation of U.S.

samples, whereas Gati's model appeared to be a better representation of non-U.S. samples.

## 2. EXTENSIONS OF THE CIRCULAR MODEL: SPHERICAL STRUCTURE

In two studies on the structure of interests, Tracey and Rounds demonstrated (i) the arbitrary nature of the RIASEC types and (ii) the presence of prestige in interest data. Interest items were found to be uniformly distributed around the interest circle, not centered on the six nodes of the RIASEC types. This implied that the



**FIGURE 2** Interest circle represented by Holland's six RIASEC types and Tracey's eight types.

RIASEC hexagon can be better represented by a circle composed of any number of types and the adoption of six is arbitrary. To examine this, Tracey and Rounds created an eight type model and found that it fit the data as well as or better than the six type RIASEC model. A version of their eight type model is presented in Fig. 2.

Prestige—also referred to as status, occupational level, level of training, and level of difficulty or responsibility—is a salient variable in vocational assessment, but it is commonly unaccounted for in interest inventories. Adding prestige to the RIASEC interest circle, Tracey and Rounds proposed a spherical structure wherein a prestige dimension exists orthogonal to Prediger's people–things and data–ideas dimensions. Tracey's Personal Globe Inventory is an inventory developed to measure the spherical model and a representation of the structure is depicted in Fig. 3.

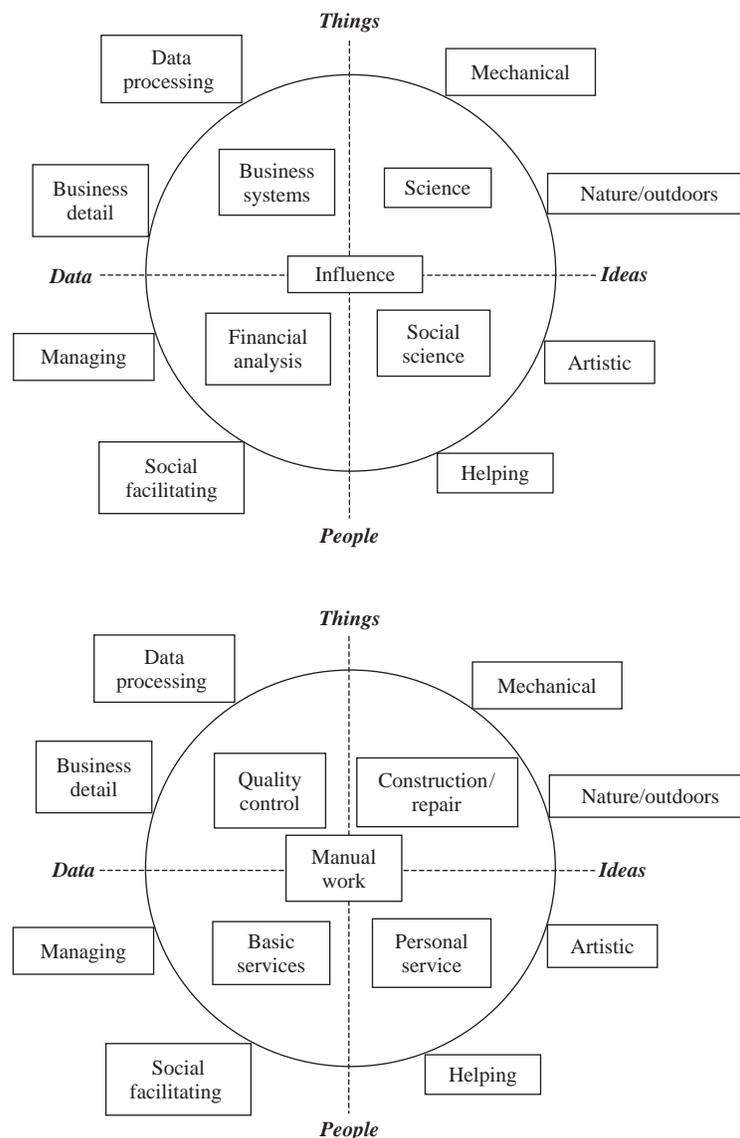
The model is best conceptualized as an occupational interest globe, with prestige represented along the north pole–south pole axis and serving to differentiate interest areas. For example, as one moves up the globe toward high prestige interest areas (e.g., social science, science, and financial analysis) or down the globe toward low prestige interest areas (e.g., personal service, basic services, and construction/repair) the salience of the people–things and data–ideas dimensions diminishes. Individuals with moderate levels of prestige preference would be characterized by standard RIASEC measures and appear near the equator on the occupational interest globe.

### 3. RELATION OF THE RIASEC TYPES TO PERSONALITY AND ABILITY MEASURES

Many theories suggest that vocational interests are an extension of personality and that interests are related to abilities. Holland used trait descriptive adjectives to describe the RIASEC interest types and claimed that they were essentially personality types. The most well-known theory of personality trait structure, the Big Five Model, is a factor list taxonomy for organizing personality variables. Correlations between the RIASEC and the Big Five are small to moderate and also mixed; enterprising and social types are correlated with extroversion, and artistic types are correlated with openness to experience. Research has generally ignored the circular structure of the RIASEC, with the exception of one study. One of the Big Five traits, extroversion, fit the RIASEC circular structure, separating enterprising and social types from realistic and investigative. Integrative work by Ackerman and Heggestad has focused on representing the overlap among the RIASEC interests, personality traits, and ability. They present initial models of the relation among each of these domains, and this research should serve as stimuli for subsequent research on the intersection of these three important domains.

### 4. CURRENT RESEARCH

Most research on the structure of interests focuses on adolescents and adults. Very little is known about the structure of interests in children or the development of the structure of interests. Initial studies demonstrate that the structure of interests in elementary school children does not mirror the circular structure found in those 14 years old or older. Work in the area of the development of interests and their structure is needed as an aid to design educational programs appropriate to the elementary school age group. Since the development of interests in children is inherently changeable and unique, research focusing on this process should examine differences in how children think about their likes and dislikes and how the development of interests changes over time.



**FIGURE 3** Spherical model of interests (as exemplified by the Personal Globe Inventory). (Top) Upper half (higher prestige); (bottom) lower half (lower prestige).

### See Also the Following Articles

Career Counseling ■ Holland's Theory (Vocational Personality Types) ■ Indecision, Vocational ■ Job Search ■ Traits ■ Vocational Interests ■ Vocational Psychology, Overview

### Further Reading

Gati, I. (1991). The structure of vocational interests. *Psychological Bulletin*, *109*, 309–324.  
 Holland, J. L. (1997). *Making vocational choices* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Prediger, D. J., & Vansickle, T. R. (1992). Locating occupations on Holland's hexagon: Beyond RIASEC. *Journal of Vocational Behavior*, *40*, 111–128.

Tracey, T. J. G. (2002). Personal Globe Inventory: Measurement of the spherical model of interests and competence beliefs [Monograph]. *Journal of Vocational Behavior*, *60*, 113–172.

Tracey, T. J. G., & Rounds, J. (1997). Circular structure of vocational interests. In R. Plutchik, & H. R. Conte (Eds.), *Circumplex models of personality and emotions* (pp. 183–201). Washington, DC: American Psychological Association.



# Study Skills Instruction

Maribeth Gettinger

University of Wisconsin, Madison, Wisconsin, USA

1. Introduction
2. Importance of Study Skills
3. Three Dimensions of Studying
4. Definition of Study Skills
5. Study Skills in High- Versus Low-Achieving Students
6. Rationale and Theoretical Framework for Study Skills Instruction
7. Key Components of Study Skills Instruction
8. Four Clusters of Study Skills
9. Conclusions and Future Directions  
Further Reading

**study skills** Cognitive tools associated with acquiring, recording, organizing, synthesizing, remembering, and using information.

**study strategy** Configuration of different tactics, deliberately selected by the learner for a particular purpose or learning situation.

**study tactic** Specific study technique, such as underlining, note-taking, outlining, summarizing, or self-questioning.

Study skills encompass a range of coordinated cognitive processes that enhance the effectiveness and efficiency of students' learning. Study skills contribute to success in both nonacademic and academic settings. Within an information-processing framework, study skills are grouped as repetition-based, procedural, cognitive, and metacognitive skills. Key elements of study skills instruction include cognitive modeling, guided practice and feedback, and self-regulation.

## GLOSSARY

**cognitive modeling** Teacher demonstrates the use of a study strategy while "thinking aloud" to verbalize the rationale and reasoning behind its use.

**cognitive study strategies** Strategies that guide students to engage in appropriate thinking about the information they are studying (e.g., self-questioning).

**imitation** Learner applies a study strategy in a way that approximates the teacher's demonstration.

**metacognitive strategies** Skills that guide students in how to select, use, and monitor study tactics.

**procedural study skills** Behaviors or habits that allow students to maximize the benefits of available study time (e.g., time management and material organization).

**repetition-based strategies** Study strategies that rely on repetition or rehearsal of information to be learned.

**self-regulation** Students initiate the use of a study strategy on their own and make adjustments or adapt the strategy to meet their needs.

## 1. INTRODUCTION

Study skills are fundamental for learning and academic competence. For many learners, lack of success in school is associated with deficiencies in their knowledge and application of effective study skills. Capable students at all grade levels may experience failure and frustration in school, not because they lack ability but because they lack study skills. Despite a well-documented link between academic success and

effective studying, students rarely receive explicit instruction in study skills. Teachers often assume that students acquire basic study skills on their own. Most learners, however, will not become proficient at studying without systematic instruction and repeated practice.

## 2. IMPORTANCE OF STUDY SKILLS

There are several reasons for schools to focus on developing effective study skills among students. First, study skills are necessary for learning. According to Devine's 1987 definition, study skills are the tools associated with all components of the learning process, specifically "acquiring, recording, organizing, synthesizing, remembering, and using information and ideas found in school" (p. 5). Second, effective study skills enhance students' sense of personal control and self-efficacy. Students with good study skills feel competent and confident as they learn; they are more likely to approach schoolwork with positive expectations and attitudes. Finally, study skills are basic "life skills" that apply to many activities outside of school. Study skills go beyond helping students achieve success in a classroom; they help individuals function effectively in multiple and diverse contexts (e.g., academic, athletic, social, or vocational). The development of study skills contributes to an individual's growth as an independent, lifelong learner. In his seminal 1984 paper, Rohwer noted that "studying is the principal means of self-education throughout life" (p. 1).

## 3. THREE DIMENSIONS OF STUDYING

A comprehensive definition of study skills derives from an understanding of what is meant by studying. Studying is defined on the basis of three key dimensions. First, studying is skillful; it requires training and practice with specific techniques that help a learner acquire, organize, retain, and use information. Second, studying is intentional. Studying requires not only the knowledge and application of skills but also volition. The act of studying is a deliberate, conscious effort by students. It depends as much on personal efficacy and motivational factors as it does on organizational and procedural skills. Finally, studying involves a self-regulatory dimension. When individuals study, they engage in self-regulatory behaviors, such

as self-monitoring. Good studiers have been shown to evaluate whether understanding is occurring and whether information is being remembered.

## 4. DEFINITION OF STUDY SKILLS

Given these characteristics of the process of studying, what does it mean to have good study skills? Study skills is a broad term that refers to a range of competencies that enhance the effectiveness and efficiency of learning. In 1991, Ellis *et al.* made an important distinction between a study tactic and a study strategy. A study tactic is a specific technique or skill, such as underlining, note-taking, outlining, summarizing, or self-questioning. A study strategy is a configuration of different tactics, deliberately selected by the learner for a particular purpose or learning situation. Having good study skills requires both knowledge of effective tactics and strategies to execute them in a planned sequence and for a specific purpose.

## 5. STUDY SKILLS IN HIGH- VERSUS LOW-ACHIEVING STUDENTS

A "think-aloud" assessment method developed by Kucan and Beck in 1997, whereby students express their thoughts and verbalize their strategies while engaged in studying, has allowed researchers to explore the cognitive processes involved in effective versus ineffective studying. Rich descriptive information gained from think-aloud assessment, as well as other methods of inquiry (e.g., interviews, observation, and self-report measures), has contributed to a profile of study behaviors that differentiate successful from unsuccessful students. Characteristic study skills of high- versus low-achieving students are summarized as follows:

High-achieving students

- Do not use one study tactic; instead, they choose from a repertoire of study tactics and select study tactics in a purposeful manner.
- Take an active role in studying and justify the use of study tactics for a given learning task.
- Allocate their attention appropriately to focus on important information, jumping forward and backward to process information effectively.
- Monitor their own studying and take action to "fix up" inaccuracies in their comprehension or change study strategies.

Low-achieving students

- Have a restricted range of tactics, and use the same, often ineffective, study approach for all learning, regardless of the task content, structure, or expectations.
- Assume a passive role in learning and rely on others to tell them how to study.
- Are unaware that they must extend effort beyond simply reading content to fully comprehend and retain information.
- Do not monitor their own studying and show little evidence of looking back, or employing “fix-up” strategies to remedy comprehension problems.

## 6. RATIONALE AND THEORETICAL FRAMEWORK FOR STUDY SKILLS INSTRUCTION

During the past 20 years, research has provided evidence supporting the effectiveness of study skills instruction. Through explicit training in study skills, students become more efficient, thoughtful, and independent as learners, and they perform better in school. Even students who develop study skills on their own learn to study more effectively and efficiently through explicit instruction.

Although different theoretical perspectives support the benefit of equipping students with study skills, the most applicable theory is information processing. In brief, an information-processing perspective assumes that information to be learned is manipulated by the student to enhance acquisition and retention. The level of processing, or manipulation, is affected by the type of study strategy the learner uses. The more elaborate the strategy, the deeper the level of processing. An information-processing perspective provides the theoretical framework for two broad areas of research related to study skills instruction, as described next.

## 7. KEY COMPONENTS OF STUDY SKILLS INSTRUCTION

In recent years, there has been a trend toward developing and evaluating teaching models that emphasize the flexible use of study tactics. Evaluations of two approaches provide evidence that study strategy training improves academic performance, especially for low-achieving students. These approaches are the Strategies Intervention Model, developed by Deschler,

Schumaker, and other researchers at the Kansas Institute for Research in Learning Disabilities and Center for Research on Learning, and the Good Information Processing Approach, based on the work of Pressley and colleagues.

The most effective study skills instruction incorporates a sequence of instructional phases that proceed from social modeling to a gradual increase in self-directed functioning. First, instruction begins with breaking the study strategy into basic steps, followed by explicit instruction and frequent modeling of the strategy by the teacher. Research highlights the importance of showing students, not just telling them, how to use a study skill. This type of modeling is termed cognitive modeling because the teacher demonstrates the use of a study strategy while “thinking aloud” to verbalize the rationale and reasoning behind its use. The second phase is imitation, whereby the learner applies the strategy in a way that approximates the teacher’s performance. The primary role of the teacher is to provide multiple opportunities for students to practice study skills and receive feedback and guidance. Teacher guidance is gradually reduced over time as students assume increasing responsibility for using the strategy on their own. In the third phase, referred to as self-regulation, learners use the study skill independently. Opportunities are provided for students to practice using study skills with their daily assignments. Although teachers may provide minimal guidance, the expectation is that students initiate the use of a study strategy on their own and make adjustments or adapt the strategy to meet their needs. In summary, effective study skills instruction incorporates three critical components: (i) simplification of the strategy so that the basic steps are explicit, justified, and amenable to individualization; (ii) frequent modeling or demonstration of the learning strategy; and (iii) guided instruction, practice, and feedback to help students move from teacher dependence to self-dependence in judging a strategy’s utility and potential effectiveness.

## 8. FOUR CLUSTERS OF STUDY SKILLS

In 1985, Weinstein and Mayer offered a useful framework for classifying study skills within an information-processing perspective. Specifically, study strategies are grouped on the basis of the degree of manipulation or level of processing of the information to be learned.

Using this framework, four clusters of study skills are identified: repetition-based, procedural, cognitive, and metacognitive.

The most basic study skills are repetition-based. Rehearsal strategies are most useful when storing small bits of information for the short term (e.g., studying 15 weekly spelling words) or when the content being studied is used frequently (e.g., multiplication facts). From an information-processing perspective, repetition-based study skills, although easy to learn and apply, afford minimal processing of content. Repetition strategies may be augmented to promote deeper processing of information. In 1993, Levin reviewed 20 years of research supporting the use of mnemonic devices, such as mental imagery, to enhance the effectiveness of repetition-based study skills, particularly for long-term retention and acquisition of complex information.

Procedural study skills encompass the behaviors or habits that allow students to maximize the benefits of their available study time. Several competencies underlie the development of procedural study skills, including time management, material organization, and development of schedules or consistent study routine. Organizational routines and study procedures are most effective when they are personalized by having students construct their own study plans. Although research has failed to document significant benefits for any single study routine over another, such as SQ3R (Survey, Question, Read, Respond, Review), what does contribute to positive outcomes is the consistency with which a study routine is implemented and the extent to which it is personalized or adapted by individual learners.

The goal of cognitive study strategies is to guide students to engage in appropriate thinking about the information they are studying. From an information-processing orientation, the more knowledge students have about content, the more likely they are to think about, understand, and remember it. Thus, studying is enhanced when new material is meaningful to learners and integrated with their existing knowledge. In addition, information that is stored in a network of connected facts and concepts, called schemata, is more easily learned and retained. Cognitive-based study strategies are designed to help students activate background knowledge prior to studying a topic; connect new information or concepts to what they already know; and develop new schemata, when necessary, to integrate content to be learned. Specific study skills that allow students to achieve these cognitive goals include the use of cognitive organizers or semantic maps, which are graphic representations of the interrelatedness of key

ideas; question generation and summarization to activate and make connections with prior knowledge; and peer-assisted learning approaches, including tutoring, mentoring, and cooperative learning.

Whereas cognitive-based study skills relate to how learners process information, metacognitive strategies relate to how students select, monitor, and use strategies. Students with well-developed metacognitive skills are able to adjust their studying according to varying task demands; they know how to study effectively; and they understand which study strategies to deploy, monitor their studying, and allocate study time wisely. In contrast, students with weak metacognitive skills are often disorganized in their studying and have limited understanding about what to do or how to proceed with studying. Research reported by Deshler *et al.* in 1996 demonstrated that explicit training in self-questioning (e.g., “Do I understand the material I am studying?” “Should I reread or revise my study strategy?”) can significantly improve students’ metacognitive abilities and task performance.

## 9. CONCLUSIONS AND FUTURE DIRECTIONS

Several key principles are important to keep in mind when designing study skills instruction. First, students must recognize the need to implement a specific study strategy. Not all strategies are appropriate for all study tasks. For example, the most effective approach for studying spelling words is likely to be different from an effective approach for studying for a history test. Similarly, any single study tactic will require some personalization on the part of the students. In developing an awareness of different strategies, students should be encouraged to explain the appropriateness of a particular study approach for an assigned task. Routine or rote application of a study technique does not ensure effective studying. Note-taking is a good example. Students use a note-taking tactic in different ways. Some students routinely copy parts of a text verbatim and process information only superficially; others use their own words when they take notes or write notes that elaborate on the text and promote a deeper processing of information. Some students may take notes randomly; others may take notes strategically, selecting information that is important, relevant, or worthy of extra processing. How students take notes and what they do with a note-taking tactic affect how

successful their studying will be, not the activity of note-taking itself. Therefore, students need training or direction in how to use a technique effectively and to employ one that best fits the task they are studying.

In summary, acquiring effective study skills is important for academic success. Problems with studying can contribute to academic failure and inability to profit from even the most optimal instruction. To be effective students, student must, first, have a wide arrange of evidence-based strategies at their disposal and, second, know where, when, and how to apply these strategies. Research has demonstrated that study skills can be taught at all grade levels, thus highlighting the need for an emphasis on the development and maintenance of study skills across the curriculum for all grade levels.

### See Also the Following Articles

Learning ■ Learning Styles and Approaches to Studying  
 ■ Teaching Effectiveness ■ Transfer of Learning

### Further Reading

- Deshler, D. D., Ellis, E. S., & Lenz, B. K. (1996). *Teaching adolescents with learning disabilities: Strategies and methods*. Denver, CO: Love Publishing.
- Devine, T. G. (1987). *Teaching study skills: A guide for teachers*. Boston: Allyn & Bacon.
- Ellis, E. S., Deschler, D. D., Lenz, B. K., Schumaker, J. B., & Clark, F. L. (1991). An instructional model for teaching learning strategies. *Focus on Exceptional Children*, 23, 1–24.
- Gettinger, M., & Seibert, J. K. (2002). Contributions of study skills to academic competence. *School Psychology Review*, 31, 350–365.
- Hoover, J. J., & Patton, P. R. (1995). *Teaching students with learning problems to use study skills: A teacher's guide*. Austin, TX: Pro-Ed.
- Kucan, L., & Beck, I. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67, 271–299.
- Levin, J. R. (1993). Mnemonic strategies and classroom learning: A twenty-year report card. *Elementary School Journal*, 94, 235–244.
- Pressley, M., Woloshyn, V., & Associates. (1995). *Cognitive strategy instruction that really improves children's academic performance* (2nd ed.). Cambridge, MA: Brookline.
- Robinson, F. P. (1970). *Effective study* (4th ed.). New York: Harper & Row.
- Rohwer, W. D. (1984). An invitation to an educational psychology of studying. *Educational Psychologist*, 9, 1–14.
- Scheid, K. (1993). *Helping students become strategic learners: Guidelines for teaching*. Cambridge, MA: Brookline.
- Schunk, D. H., & Zimmerman, B. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. New York: Guilford.
- Strichart, S. S., Mangrum, C. T., & Iannuzzi, P. (1998). *Teaching study skills to students with learning disabilities, attention deficit disorders, or special needs*. Boston: Allyn & Bacon.
- Weinstein, C. E., & Mayer, R. F. (1985). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 315–329). New York: Macmillan.
- Winne, P. H., & Hadwon, A. F. (1997). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 234–256). Hillsdale, NJ: Erlbaum.
- Wood, E., Woloshyn, V. E., & Willoughby, T. (Eds.). (1995). *Cognitive strategy instruction for middle and high school*. Cambridge, MA: Brookline.



# Subjective Culture

Harry C. Triandis

University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

1. Introduction
  2. Elements of Subject Culture
  3. Conclusion
- Further Reading

## GLOSSARY

*emic* Refers to culture-specific associations.

*etic* Refers to a quality that is universal.

*loose culture* A quality of a group in which deviation from the group norms is tolerated.

*tight culture* A quality of a group in which conformity to the group norms is strictly enforced.

A broad definition of culture is that it is the human-made part of the environment. According to this definition, culture can be categorized as material and subjective culture. Material culture consists of such elements as dress, food, houses, highways, tools, and machines. Subjective culture is a society's characteristic way of perceiving its social environment. It consists of ideas about what has worked in the past and thus is worth transmitting to future generations. Language and economic, educational, political, legal, philosophical, and religious systems are important elements of culture. Ideas about aesthetics and how people should live with others are also important elements. Most important are unstated assumptions, standard operating procedures, and habits of sampling information from the environment.

## 1. INTRODUCTION

Subjective culture is a shared pattern of beliefs, attitudes, norms, role perceptions, and values. Thus, the first step when studying culture is to ask whether the element of culture is shared. The next step is to determine whether the shared responses correspond to a specific language, a time period, and a geographic region. Usually, people who share a language dialect at a particular time and place are members of the same culture.

Subcultures emerge because people share other elements, such as gender, physical type, neighborhood, occupation, standard of living, climate, and resources. For example, lawyers throughout world share some elements of subjective culture. Japanese lawyers have a subculture that differs from that of other lawyers as well as from general Japanese culture. A nation consists of thousands of cultures, but many of these cultures have common elements.

Since there are many elements of subjective culture, a strategy is required to study them most economically. Focusing on some of the elements of subjective culture is one way to proceed economically.

## 2. ELEMENTS OF SUBJECT CULTURE

### 2.1. Categories

Much can be learned about a culture by analyzing the categories that people use. Indigenous psychologists

study the meaning of specific words, such as the Japanese meaning of *amae* or the Greek meaning of *philotimos*. Such words do not have a corresponding meaning in other languages, so they tell much about a particular culture. *Amae* means something like expecting another person to indulge you. It is the kind of feeling one can find between a mother and a child. *Philotimos* literally means friend of honor and is a common adjective used by traditional Greeks to describe themselves. It can be translated into “a person who does very frequently what family and friends expect done.”

A category can be identified by noting that people give the same response to discriminably different stimuli. Our eyes, for instance, are capable of discriminating 7.5 million different color stimuli. However, we do not have millions of words that refer to color. In fact, most of us get along well with a few dozen color names.

When studying categories, we discover that some cultures have many words for a particular domain, and others have few words. For instance, the Eskimo have many words for snow; also, we have many words for cars (e.g., Ford, VW, Dodge, Toyota, truck, and vehicle). This tells us at once that the Eskimo deal with snow a lot, and we deal with cars a lot. Categories can also tell us about the way people behave. For instance, the Pawnee of Oklahoma use the same word for “mother’s brother’s wife,” “ego’s wife,” and “sisters of ego’s wife.” We note that unacculturated Pawnees have sexual relations with all these women.

## 2.2. Categories Have Associations

Categories are associated with other categories. Extensive work by Osgood *et al.* has shown that throughout the world people associate categories with evaluation (good, beautiful, and moral), potency (strong, heavy, and large), and activity (fast, alive, and noisy). Since these associations occur universally, they are called etic. The term etic refers to a quality that is universal. However, there are also emic (i.e., culture-specific) associations. When we compare cultures we need to use etic constructs, but when we describe cultures we need to use emic constructs. A metaphor may help: If we compare apples and oranges we can use etic elements such as weight, size, thickness of skin, and price. However, obviously one does not learn much about such fruit with this kind of information. One needs to learn about apple flavor and orange flavor, apple texture and orange texture, and the like. These are emic qualities. Therefore, when comparing

cultures etic qualities are used, but when describing cultures emic qualities are needed.

We can learn a good deal about a culture by examining its emic associations. For instance, some research has indicated that in South Korea “democracy” and “socialism” are strongly associated, but this is not the case in the United States. This tells us something about the political culture of South Korea.

## 2.3. Beliefs

Categories are linked to each other in other ways as well. For instance, “If a relative asks for help, you must give it” is a strong belief in some cultures, but in other cultures it is not. People in the latter cultures consider more complicated ideas, such as “Is the asking legitimate?” or “Do I like this relative?”

## 2.4. Attitudes

Attitudes are ideas charged with affect (emotion) predisposing action. Any category can be the core of an attitude.

## 2.5. Norms

Norms are ideas about behavior expected of members of a group or culture. In some cultures, called tight, people are expected to behave exactly as specified by norms. In other cultures, called loose, one can deviate from norms. In tight cultures one is punished if one does not behave according to the norms. In loose cultures a person is less likely to be punished. Punishment in some cases is very severe, whereas in other cases it is quite lenient.

When a culture is homogeneous and people are very interdependent and can be supervised closely, the culture is usually tight. When a culture is under the influence of many other cultures, or when people are not too interdependent or supervision is difficult (e.g., people live far from each other), it is more likely that the culture will be loose. Cultures are tight or loose in different domains. For instance, American culture in 2000 was very tight about writing bad checks but rather loose about who one has as a roommate. Nevertheless, across domains cultures tend toward tightness or looseness. Thai culture is loose, American culture is in-between, Japanese culture is rather tight, and theocracies such as the Taliban culture of Afghanistan are very tight.

### 2.6. Roles

Roles are a special category of norms. Roles are ideas about the correct behavior of people who hold a position in a social group. For example, foreman, father, aunt, or sister are roles. Roles include both prescriptive elements (e.g., fathers should advise and protect their daughters) and proscriptive elements (e.g., fathers should not hit their daughters).

### 2.7. Tasks

A sequence of behaviors can be defined as a task. For example, passing a law can include a series of actions. In different cultures, different sequences of actions will correspond to such a task.

### 2.8. Values

Values are conceptions of the desirable state of affairs. Schwartz, who has studied them extensively in many countries, defined them as beliefs that pertain to desirable states or behaviors, that transcend specific situations, that guide the selection or evaluation of behavior and events, and that are ordered by relative importance. To collect data, he uses a concept, such as “freedom,” and asks people to rate its importance as “a guiding principle in my life.”

## 3. CONCLUSION

Triandis discusses subjective culture more extensively. He also provides a discussion of methodological problems and solutions when studying subjective culture. Various chapters give examples of intensive studies of several of the elements of subjective culture.

### See Also the Following Articles

Attitudes ■ Motivation and Culture

### Further Reading

- Kluckhohn, C. (1954). Culture and behavior. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 921–976). Cambridge, MA: Addison-Wesley.
- Osgood, C. E., May, W., & Miron, M. (1957). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25). New York: Academic Press.
- Triandis, H. C. (1972). *The analysis of subjective culture*. New York: Wiley.



# Successful Athletic Careers

Paul Wylleman and Marc Theeboom

Vrije Universiteit, Brussels, Belgium

David Lavalley

Loughborough University, Loughborough, United Kingdom

1. Nature and Conceptualization of the Athletic Career
  2. Theories and Models Related to Career Development of Athletes
  3. Career Intervention Programs for Athletes
  4. Conclusion
- Further Reading

**successful athletic career** The athletic career of an athlete who, in comparison with her or his age group peers, achieves and sustains a consistent level of athletic excellence at the national, continental, and/or world levels.

**transition** Generally results from one event, or a combination of events, perceived by an athlete to bring about personal and social disequilibria that are beyond the ongoing changes of everyday life and that cause a change in the athlete's assumptions about herself or himself.

## GLOSSARY

**athletic career** A succession of stages and transitions that includes an athlete's initiation into and continued participation in organized competitive sport and that is terminated with the athlete's (in)voluntary but definitive discontinuation of participation in organized competitive sport.

**career program** An integrated and comprehensive combination of workshops, seminars, educational modules, and individual counseling and/or a referral network providing individualized or group-oriented multidisciplinary support services to elite athletes with regard to athletic participation, developmental and lifestyle issues, and educational and vocational development.

**non-normative transition** An idiosyncratic transition that is generally unpredicted, unanticipated, and involuntary and that does not occur in a set plan or schedule but rather is the result of important events that take place in an athlete's life and to which she or he responds.

**normative transition** A transition that is part of a definite sequence of age-related biological, social, and emotional events or changes and that can be said to be generally related to the socialization process and to the organizational nature of the setting in which an athlete is involved.

A successful athletic career refers to the athletic career of an athlete who achieves and sustains in a consistent manner a level of athletic excellence in international competitions (e.g., continental or world level). Although research on the development of the athletic career during the late 1960s to early 1970s originally focused on anecdotal evidence of the process of retirement from high-level competitive and professional sports, it has become a well-delineated and growing topic of study among the sport psychology community. Research reveals that the athletic careers of elite athletes consist of succeeding normative stages, including the initiation, development, mastery, and discontinuation stages. Athletes progress in their athletic development when they cope successfully not only with the characteristics of each stage but also with the demands of the transition from one stage to another. Furthermore, athletes must cope with the demands of the stages and transitions that occur concurrently with those at the athletic level in their psychological, psychosocial, academic, and/or vocational development. Career intervention programs

are implemented to support athletes in optimizing and furthering their holistic development.

## **1. NATURE AND CONCEPTUALIZATION OF THE ATHLETIC CAREER**

The academic interest in the athletic career has clearly known an increase during the past two decades. Whereas only 20 works pertaining to this area were published before 1980, more than 270 works related to the career of the athlete could be identified at the time of this writing. In addition, an inspection of the proceedings of the recent world and European sport psychology congresses reveals that nearly 10% of the program was dedicated to topics related to the athletic career. Moreover, for more than a decade, an international special interest group has focused on the exchange of information on applied and investigative work in the area. Finally, governing bodies and sport institutes around the world have developed a number of intervention programs to assist athletes in their career development.

Empirical research on the career development of elite athletes emanates from studies conducted during the 1960s and 1970s on the sport career termination among elite and professional athletes. The earliest studies revealed that athletes retiring from elite sports experienced a wide range of psychological, interpersonal, social, and financial problems initiated by a range of negative or even traumatic experiences. These studies tended to center almost exclusively on the dysfunctional issues (e.g., depression, alcoholism) experienced by retiring athletes and on male professional athletes in specific (i.e., professional) sports. However, research data gathered during the 1980s not only challenged the widespread assumption that sports retirement is inherently stressful but also led researchers to suggest that the athletic career termination should be seen as a transitional process rather than as a singular event. The increased understanding of the career termination process also motivated sport psychologists during the 1990s to broaden their attention to the occurrence of other career transitions faced by athletes. This shift toward a career development perspective was also enhanced by the integration of conceptualizations emanating from research into youth sport participation and the development of talented and expert performers. Recently, research into the

development of the athletic career has taken a life span perspective by focusing not only on athletes' development from "beginning to end" at an athletic level but also on other domains of their lives (e.g., academic, psychosocial, professional).

In sum, research on the career development of elite athletes has been seen to evolve during the past four decades or so from studies on the sport career termination to studies taking a holistic, life span, multilevel approach to the sports and postsports careers of elite athletes.

## **2. THEORIES AND MODELS RELATED TO CAREER DEVELOPMENT OF ATHLETES**

Several theories and models have been outlined in the sport psychology literature to conceptualize the athletic career. These have been employed predominantly from the mainstream psychological literature and have been developed and revised following research conducted in the area. Comparing the "end" of the athletic career to that of retirement from the workforce or even to the process of dying, sport psychologists drew from the fields of social gerontology (i.e., the study of the aging process) and of thanatology (i.e., the study of the process of dying and death). From social gerontology, "subculture" theory asserted that prolonged social interactions among individuals lead to the development of a group consciousness and that people can be less active and well adjusted during retirement even if their situations are different from overall social norms. Although athletes have fairly distinguishable (sub)cultural characteristics, and although this theory assists in revealing potential adjustment problems experienced by athletes in ending their athletic careers, the use of this theory was questioned because retiring athletes are moving out of, rather than into, the proposed subculture.

From thanatology, the "social death" approach was used to explain how, after their sports career termination, retired athletes were treated as if they were dead even though they were still biologically alive, resulting in a loss of social functioning, isolation, and/or even ostracism. This approach failed due to, among other reasons, the obvious fact that athletic retirees continued functioning in society, albeit in a different social role. A more popular use of thanatology involved describing athletic retirement in the series of stages experienced when facing death: denial and isolation

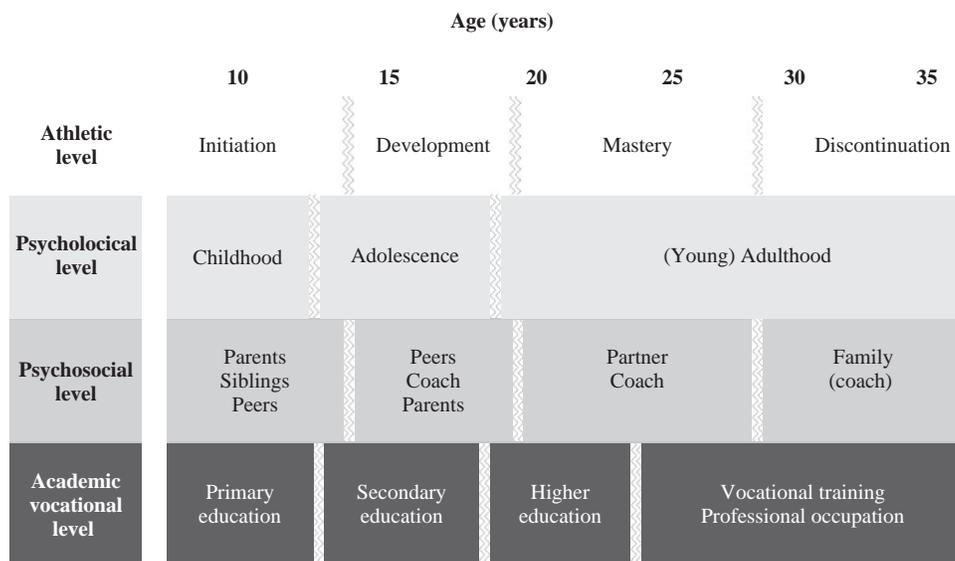
(where athletes initially refuse to acknowledge the inevitability of their career termination), anger (where retiring athletes become disturbed at the overall changing situation), bargaining (where they try to negotiate for a lengthened career in sport), depression (where they experience a distress reaction to retirement), and acceptance (where they eventually come to accept their career transition). Notwithstanding their intuitive appeal, thanatological models were criticized for the lack of an analogy between terminal illness and career termination.

More recently, models of transition have also been adopted as alternative frameworks for working with athletes in transition. These models characterize retirement from sport as a process and propose that the attributes of these athletes, cognitive appraisals of the career termination, and characteristics of the postretirement environment all interact throughout a sports career transition. Transition frameworks used in research with athletes include Sussman's "analytical model" and especially Schlossberg and colleagues' proposed "model of human adaptation to transition." In this latter model, three major sets of factors interact during a transition: the characteristics of the individual experiencing the transition, the perception of the particular transition, and the characteristics of the pre- and posttransition environments. A number of researchers also developed a sport-specific career termination model, including Taylor and Ogilvie's "model of athletic career termination." Although these models have been instrumental in stimulating research on the quality of the process of athletic retirement, their focus has remained limited to this one particular transition ending the athletic career.

Therefore, sport psychologists have adopted a developmental approach by outlining a series of predictable or "normative" transitions throughout the athletic career. This approach relates to the work by Bloom, who identified three stages of talent development (within the fields of science, art, and sport), with each stage being delineated by specific transitions. These normative transitions are part of a definite sequence of age-related biological, social, and emotional events or changes and can be said to be generally related to the socialization process as well as to the organizational nature of the setting in which the individual is involved (e.g., school, family). In the sporting domain, normative transitions include, for example, transitions from junior to senior level, from regional- to national-level competitions, from amateur to professional status, and from active participation to discontinuation from

competitive sport. During these types of transition, the athlete exits one stage and enters another stage, making these transitions generally predictable and anticipated. Non-normative transitions, on the other hand, do not occur according to a set plan or schedule and are the result of important events that take place in an individual's life and to which she or he responds. For the athlete, such transitions may include a season-ending injury, the loss of a personal coach, or an unanticipated de-selection from the team. As a result, these transitions are generally unpredicted, unanticipated, and involuntary. Non-normative transitions also include those that were expected or hoped for but that did not happen (labeled "nonevents"). Not making the senior team after having made the final preselection and not being able to participate in a major championship after years of preparation are examples of nonevents.

Based on research findings showing the strong concurrent, interactive, and reciprocal nature of normative stages and transitions occurring in the athletic career and those occurring in other domains of athletes' lives, a developmental model on the athletic career that includes the normative stage transitions faced by athletes at the athletic, individual, psychosocial, and academic/vocational levels has been developed (Fig. 1). The top layer of this model reflects four stages and transitions that athletes face in their athletic development: the initiation stage (where young athletes are introduced to organized sports and where they are identified as talented athletes), the development stage (where athletes become more dedicated to their sport and where the amount of training and level of specialization is increased), the mastery (or perfection) stage (where athletes reach their highest levels of athletic proficiency), and the discontinuation stage (where elite athletes discontinue further participation in high-level competitive sport). Although this layer reflects athletes' development in elite sports, these stages can also be considered for athletes at the non elite level. The second layer of the model reflects the developmental stages and transitions that occur at the psychological level: childhood, adolescence, and (young) adulthood. The third layer of the model reflects the changes that occur in athletes' psychosocial development relative to their athletic involvement: the athletic family, peer relationships, coach-athlete relationships, marital relationships, and other interpersonal relationships significant to athletes. The final layer of the model reflects the stages and transitions at the academic and vocational levels: the transition to primary education/elementary school, the



**FIGURE 1** A developmental perspective on transitions faced by athletes at the athletic, psychological, psychosocial, and academic/vocational levels. Dotted lines indicate that the ages at which the transitions occur are approximations. Reprinted from Wylleman, P., & Lavallee, D. (2004). A developmental perspective on transitions faced by athletes. In M. R. Weiss (Ed.), *Developmental sport and exercise psychology: A lifespan perspective* (p. 516). Morgantown, WV: Fitness Information Technology. Reprinted with permission.

stage of secondary education/high school, the transition to higher education (college/university), and the transition to vocational training and/or a professional occupation (which may also occur at an earlier age). This model illustrates, among other things, the concurrence of transitions occurring at different levels of development (e.g., the occurrence of the transition from secondary education to higher education close to the athletic transition from the development stage to the mastery stage). This developmental model is currently used not only to determine the type of support required by talented athletes to progress in their athletic development but also to assist elite athletes in preparing for and transitioning to their postathletic careers.

In a similar vein, other researchers have identified other normative transitions. For example, Côté described the stages of sampling, specializing, investment, and mastery or performance in the development of deliberate play and practice, whereas Stambulova considered the beginning of the sport specialization, the transition to intensive training in the chosen sport, the transition to high-achievement and adult sport, the transition from amateur sport to professional sport, the transition from culmination to the end of the sport career, and the end of the sport career as predictable stages and transitions.

From an applied perspective, this developmental model enables counselors to situate and reflect on the developmental, interactive, and interdependent nature of transitions and stages faced by an individual athlete. For example, when using this model, a sport psychology consultant not only would situate a talented 16-year-old athlete in the developmental phase at the athletic level (e.g., en route to a possible selection for the national junior team) but also would consider the changes that the adolescent might be experiencing at psychological level (e.g., separation–individuation process) and psychosocial level (e.g., distancing from parents, growing influence of peers, intensifying relationship with coach) and would highlight the importance of the athlete to complete secondary education (and perhaps prepare for higher education).

However, the general focus on the identification of normative transitions should not obscure the occurrence of non-normative transitions. The unpredictable and involuntary nature of these transitions requires the development of conceptual models that include the mechanisms required by athletes to cope successfully, that is, being able to negotiate and reduce the possible negative cognitions, affects, and behaviors that may precede, accompany, or follow from these transitions.

### 3. CAREER INTERVENTION PROGRAMS FOR ATHLETES

During recent years, numerous career intervention programs for athletes have been developed in countries around the world (Table I). The existing programs vary in format and often include workshops, seminars, educational modules, and individual counseling. The majority of programs focus on lifestyle management and the development of transferable skills that can assist individuals in making the transition from lives in sport to postsport careers. As such, these programs provide athletes with an introduction to career planning and development by focusing on values and interest exploration, career awareness and decision making, résumé preparation, interview techniques, and job search strategies. Career development programs for athletes are primarily managed by national sports governing bodies, national Olympic committees, specific sport federations, universities, and independent organizations linked to sport settings. Although some programs address the needs of professional athletes, the majority have been developed for a much wider population (e.g., amateur players, nonelite athletes).

The Olympic Job Opportunities Program (OJOP) is an international program that has been initiated in Australia, South Africa, and the United States. The principal goal of OJOP, which is sponsored by a private company, is to develop and source career opportunities for Olympians and potential Olympians. Eligible athletes are either current or Olympic-caliber athletes who must be certified as such by their respective national federations. In addition to providing

direct employer contacts and identifying job positions, OJOP provides career analysis services, personality aptitude testing, and interview skills training.

A grant from the U.S. Olympic Foundation was awarded to OJOP in 1988 to create a program to assist elite athletes in coping with the transition out of active sport competition. Following a survey of approximately 1800 Olympic and other elite-level athletes, the U.S. Career Assistance Program for Athletes (CAPA) was established to introduce individuals to the career development process while they were still competing. This program focused on increasing athletes' sense of personal competence through understanding and identifying transferable skills. A number of one-day workshops were organized around three main themes: managing the emotional and social impact of transitions, increasing understanding and awareness of personal qualities relevant to coping with transitions and career development, and introducing information about the world of work. Although these workshops were well received by the athletes, funding for the CAPA program was terminated in 1993.

One of the first career development programs to be developed for athletes was the Canadian Olympic Athlete Career Center (OACC). The OACC was launched in 1985 as part of the Canadian Olympic Association (COA) following a series of needs-based surveys conducted in 1983 and 1984. The original center, which was based in Toronto, had a mandate to assist athletes through the transition process to second careers, primarily through career and education planning. Athletes who had achieved approved rankings by way of their performances at Olympic, Commonwealth, and Pan-Pacific competitions had access to the OACC. The following career development services were initially offered through the OACC:

1. Clarification of career planning needs, self-assessment, aptitude/interest assessment to assist in identification of specific occupations of interest, decision making, and action planning skills were included.

2. Booklets were written and made available to all eligible athletes on the topics of curriculum preparation, interview preparation, job search techniques, and information interviewing. Retirement planning focused on what athletes should expect during the adjustment period. Transition workshops and peer support groups were established to help athletes deal effectively with the career transition process.

3. The COA provided reference letters of support and personalized business cards bearing the COA logo for networking purposes.

TABLE I  
Selected Overview of Career Transition Programs

Program	Institution	Country
Athlete Career and Education (ACE) program	Australian Institute of Sport	Australia
Career Assistance Program for Athletes	U.S. Olympic Committee	United States
Olympic Athlete Career Center—National Sports Center	Olympic Athlete Career Center	Canada
Olympic Job Opportunities Program	Australian Olympic Committee	Australia

4. A “shadow” program was developed in 1990 to provide athletes with the opportunity to explore career options by shadowing professionals in the field of their choice.

During recent years, the COA has initiated a reorganization process that has resulted in an increase in the number of centers operating in Canada and in additions to the career development services provided. Currently, OACCs are operating in Calgary, Montreal, Ottawa, Toronto, and Vancouver, with each center employing consultants to work with athletes living in or around these locations. The standardization of service provision across centers is ongoing and is influenced by the need and utility of the services for the athlete population in each region.

The Athlete Career and Education (ACE) program was developed in 1990 in Australia by the Victorian Institute of Sport and was later amalgamated with the Lifeskills for Elite Athletes Program (SportsLEAP). SportsLEAP was originally established in 1989 by the Australian Institute of Sport based on the results of a needs-based survey following the 1988 Seoul Olympics. This program was highly successful, but the scope and content of each program varied according to athlete demand in each state institute/academy of sport. Thus, a decision was made to integrate SportsLEAP and the ACE program (under the ACE name) to form a national program in Australia.

The overall objective of the ACE program is to assist athletes in balancing the demands of their sporting careers while enhancing their opportunities to also develop their educational and vocational skills. A major component of the program is to assist individuals in developing career plans that integrate both sporting and nonsporting components. The philosophy is to create an environment where athletes can be encouraged to be independent and self-reliant and to have a capacity to meet the demands associated with elite sport. To be eligible for assistance in Australia, athletes must be scholarship holders with the Australian Institute of Sport, or with state institutes/academies of sport, or Olympic athletic program participants. ACE managers and advisers are employed in each state institute/academy of sport, and a national manager coordinates the program. The following services are provided through the ACE program:

1. Individual athlete assessments are used to provide a structured process in which to assess individual athletes’ educational, vocational, financial, and personal development needs.

2. More than 30 personal development training courses are structured to assist athletes in meeting their sporting, educational, and career aspirations. Many of these courses provide individuals with nationally accredited competency-based education programs.

3. A nationally consistent career and education planning process is employed to enable athletes to manage their own individual vocational requirements. Career development is provided in the form of direct assistance in finding employment through career advice, training paths, and vocational training. Secondary and tertiary education support is provided through networking with individuals in secondary schools, universities that can offer unit or course selection advice, and assistance in negotiating appropriate academic and residential arrangements (e.g., quiet halls of residence, appropriate and sufficient dietary provisions at appropriate times, distance and online learning opportunities).

4. A transition program provides career and education guidance for elite athletes who are transitioning to postsport careers.

5. Training opportunities and supervised practice for ACE staff are provided through a graduate certificate in athlete career and education management.

6. Direct athlete needs-based assessments provide a structured process to assess athletes’ eligibility for support.

ACE services are available to more than 3000 elite-level athletes throughout Australia. In addition, the U.K. Sports Institute initiated an ACE U.K. Program across England, Scotland, Wales, and Northern Ireland.

The worldwide need for services and programs that optimize athletes’ development is recognized in the establishment of (inter)continental forums (e.g., International Athletes’ Services Forum, the European Forum of Lifestyle Management) where relevant career and life balance issues, interventions, and programs are shared and discussed by athlete service providers and career advisers to elite athletes and teams.

#### **4. CONCLUSION**

To better appreciate how an athlete’s sporting career develops, it is essential to understand not only the nature and conceptualization of the athletic career but also the theories and models related to career development. The athlete can develop a successful athletic career if she or he is able to cope with the different normative athletic and nonathletic stages

and transitions. Therefore, it is important to consider each athlete in terms of her or his athletic, psychological, psychosocial, academic, and/or vocational development and to be aware of the reciprocal influence that occurs among these various developmental contexts. As such, the existing career intervention programs for athletes around the world focus on supporting individuals through the stages and transitions that they experience both during and after their sporting careers. Applying the knowledge and conceptualization of the non-normative athletic and nonathletic transitions that affect the development of athletes' sport careers will further the development of these programs.

### See Also the Following Articles

Goal Setting and Achievement Motivation in Sport  
 ■ Intrinsic and Extrinsic Motivation in Sport  
 ■ Psychological Skills Training in Sport ■ Self-Confidence in Athletes

### Further Reading

- Lavallee, D., Gorely, T., Lavallee, R. M., & Wylleman, P. (2002). Career development programs for athletes. In W. Patton, & M. McMahon (Eds.), *Career development programs: Preparation for life long career decision making* (pp. 125–133). Camberwell: Australian Council for Educational Research Press.
- Lavallee, D., & Wylleman, P. (Eds.). (2000). *Career transitions in sport: International perspectives*. Morgantown, WV: Fitness Information Technology.
- Taylor, J., & Ogilvie, B. (2001). Career termination among athletes. In R. N. Singer, H. E. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (pp. 672–691). New York: John Wiley.
- Wylleman, P., & Lavallee, D. (2004). A developmental perspective on transitions faced by athletes. In M. Weiss (Ed.), *Developmental sport and exercise psychology: A life-span perspective* (pp. 507–527). Morgantown, WV: Fitness Information Technology.
- Wylleman, P., Lavallee, D., & Alfermann, D. (Eds.). (1999). *Career transitions in competitive sports*. Biel, Switzerland: European Federation of Sport Psychology.



# Suicide Intervention in Schools

Scott Poland

Cypress–Fairbanks Independent School District, Houston, Texas, USA

1. Prevalence of Youth Suicide
  2. Risk Factors for Youth Suicide
  3. Most Common Misperceptions about Youth Suicide
  4. Most Common Methods of Suicide
  5. Role of Schools in Suicide Prevention
  6. What if the Death Was a Suicide?
  7. Status of State and Individual Initiatives
  8. Suicide Prevention National Resources
  9. Summary
- Further Reading

## GLOSSARY

**emergency notification form** A form to be signed by a suicidal student's parents that acknowledges their being notified of the suicidal emergency and that provides referral information for local mental health services.

**no harm contract** An agreement to be signed by a suicidal individual that promises that the individual will not harm himself or herself and that provides alternatives for the individual in addition to suicide and crisis hotline numbers.

**postvention** A series of carefully planned steps to manage emotionality after a suicide, with special care being taken to be truthful but not to glorify the suicide victim.

**precipitating event** An emotionally troubling event that occurs just prior to a suicide attempt; this event causes the suicidal individual to act on a previously thought-out plan to attempt suicide.

**suicidologists** Professionals who study the problem of youth suicide and develop prevention strategies.

Youth suicide has been a constant problem for schools over the past three decades or so, with many researchers stating that it remains at or near an all-time high. Suicidologists have made a number of recommendations for school personnel, but suicide prevention has not been a primary focus in many districts and there are few or dwindling resources directed toward prevention. In 1989, Poland outlined the role of suicide prevention in the schools and identified many practical strategies and resources to assist schools in prevention efforts. In 2001, Portner found that few school districts have formal suicide prevention/intervention policies and that few counselors and psychologists have been prepared to deal with suicidal students. Portner concluded that progress in youth suicide prevention has been slow, with most prevention efforts taking place only after one or more youth suicides occurred in the same community or school. There is debate over what information should be presented to students about suicide prevention and whether the warning signs of suicide should be taught in key classes such as health. The American Association of Suicidology (AAS) has excellent resources available for schools on both prevention and postvention; however, very few schools take advantage of these resources. Surviving parents, such as those who founded Yellow Ribbon and those in Texas who have been working to convince the state legislature to pass laws so that schools have policies, procedures, and training, are making a difference; however, progress has been slow. A great

deal of national attention has been focused on school shootings. A U.S. Secret Service and U.S. Department of Education report on its study of 37 school shootings over several decades emphasized that the majority of school shooters were suicidal. An Associated Press article described a recent example in which a fifth-grade boy in Wellsboro, Pennsylvania, who went into a school restroom and committed suicide after his friend decided not to participate in their planned school shooting.

## **1. PREVALENCE OF YOUTH SUICIDE**

The suicide rate for high school students ages 15 to 19 years has remained relatively stable during the past decade or so, whereas the suicide rate for young people ages 10 to 14 years has increased more than 100% during the same time period. Citing a "silent epidemic," Portner in 2001 noted that 1 in 13 teenagers had already made some form of suicide attempt. Statements that the suicide rate of young people has at least tripled since the 1950s and that it remains at or near an all-time high, are, unfortunately, very accurate. Suicide rates by age and state are available from the Centers for Disease Control and Prevention (CDC) and are typically given per 100,000 individuals in a particular age group. In the United States, suicide rates are highest in western states that have a high percentage of Native Americans, and Alaska currently has the highest rate of all states. Professionals in suicidology acknowledge that suicide rates are undoubtedly underestimated due to underreporting and the unknown causes of single-car fatal accidents. The United States has a relatively high suicide rate for young people, with suicide consistently ranking second or third as the leading cause of death among adolescents. In 2000, Poland and McCormick cited figures indicating the United States had the highest suicide rate of 26 industrialized nations that were studied. However, suicides among children under 10 years of age is rare in the United States (typically fewer than 50 suicide deaths per year), and the primary reasons given for this are as follows:

- Children are more involved with their parents.
- Children are not in the advanced intellectual level of development and, therefore, are less likely to feel alienated and lonely.
- Children are involved in less substance abuse and have less access to guns.

Native American and Caucasian male youth die by suicide most often, whereas Caucasian girls attempt suicide most often. The suicide rate for African American boys has increased the most dramatically during the past decade or so. There is also growing awareness that gay and lesbian youth are particularly at risk for suicide, although more studies are needed to clarify conflicting research. The prevalence of suicidal ideation and actions by young people necessitates that schools have policies and procedures in place to address this problem, and all school personnel must be aware of the risk factors.

## **2. RISK FACTORS FOR YOUTH SUICIDE**

It is very important to examine the risk factors that are most involved with youth suicide. Awareness of these factors will assist school personnel in their important prevention efforts. Risk factors identified by research fall into five key areas: psychopathology, family factors, biological factors, environmental factors, and situational crises.

### **2.1. Psychopathology**

The majority of children and adolescents who die by suicide had experienced an emotional or mental disorder, with the most common ones being affective disorder and substance abuse. Major depression and a previous suicide attempt are significant risk factors for girls, whereas a previous suicide attempt, depression, disruptive behavior, and substance abuse are significant risk factors for boys. Intoxication is involved in approximately 50% of all suicides, and alcoholics are six times more likely to die by suicide than are nonalcoholics. Researchers further emphasize the comorbidity factors involving depression, conduct problems, and alcohol abuse, especially among high school students.

### **2.2. Family Factors**

The most significant characteristics are a family history of suicide and medical and/or psychiatric illness. In 2002, Davis and Brock found that other significant factors were economic stress, family strife, and family loss. Suicidal adolescents experience more parental divorces and report more conflict and less cohesion in their families.

### 2.3. Biological Factors

Evidence of serotonergic dysfunction in adult suicide attempters and completers has been found, and research continues in this area. The identification of neurochemical factors might be helpful in youth suicide prevention at some point in the future.

### 2.4. Environmental Factors

The strongest situational factor in the United States is the presence of a firearm. Approximately 60% of all suicides across all ages involve firearms. Many professionals believe that the most effective prevention strategy is to remove guns from ready access of troubled teens. It is noteworthy that states that have enacted strict gun laws have lower suicide rates than, for instance, western states with liberal gun laws. School personnel must not hesitate to inquire directly, during discussions with suicidal students and their parents, about gun availability and to make strong recommendations about locking up or removing guns from the homes of suicidal students.

### 2.5. Situational Crises

Interpersonal crises are the most common precipitating events to adolescent suicide attempts, with the breakup of romances and arguments with parents being the most common ones. Other significant precipitating events include loss, humiliation, and discipline or grade problems. School personnel and parents must be alert to precipitating events that may cause young persons to act on their previously thought out suicidal plans.

## 3. MOST COMMON MISPERCEPTIONS ABOUT YOUTH SUICIDE

School personnel often have many misperceptions about youth suicide, and it is important that they have accurate information. In 1999, Poland and McCormick debunked the following common myths about suicide:

*Myth:* Young people who talk or write about suicide do not commit suicide.

*Fact:* These young people are at risk and crying out for help.

TABLE I  
Situational Crises

- 
- Loss (Today's youth seem to possess more avenues to loss in their lives. Among these are loss of a loved one, losses suffered when families break up, losses caused by transience and relocation, and losses of self-esteem or romance. Traumatic grief heightens vulnerability to suicidal ideation.)
  - Disciplinary crisis (e.g., fear of incarceration or trouble at school). (The rising rate of litigation against schools might indicate a need for suicidal assessment after student suspension, expulsion, or exclusion.)
  - Stressful life events (e.g., poverty, academic and peer pressures)
  - Family crises (e.g., family violence, parental arguments, physical or sexual abuse)
  - History of running away
  - Suicide completion in the community or exposure to suicidality in others
- 

*Source.* Poland and Lieberman (2002). Copyright (2002 and 1995) by the national Association of School Psychologists, Bethesda, MA. Reprinted by permission of the publisher.

*Myth:* Suicidal youth really want to die.

*Fact:* Suicidal youth are ambivalent about dying but want the pain to end and wish that someone would change the situation in a positive way so that their lives can continue.

*Myth:* Discussing suicide will give a suicidal young person the idea to commit suicide.

*Fact:* All young people are aware that suicide is an option, and discussing it will provide assurances that others care about them. It will empower them that they can save their own lives or the lives of friends by getting adult help.

*Myth:* Suicide is inherited or destined.

*Fact:* Although parental depression is a predictor of adolescent suicide, the most important variables are situational crises, as outlined in Table I.

*Myth:* There are usually no warning signs of suicide.

*Fact:* Nearly all adolescents who die by suicide gave some warning to friends or family members. Many professionals believe that these adolescents were seeking help and options besides suicide.

## 4. MOST COMMON METHODS OF SUICIDE

Prior to 1980, the use of pills and overdoses was the number one method of suicide among girls. The use of

guns, followed by hanging and explosions, has long been the number one method among boys. More recent female suicide methods, in order of frequency, are gunshots, pill overdoses, and hanging.

Many professionals believe that removing access to guns from a suicidal individual is essential to reducing suicides given that approximately 60% of all suicides across all ages involve a gun. Professionals recommend that school personnel inquire very directly and specifically about the presence of any guns in the homes of suicidal students. Parents who are notified by school personnel that their children are suicidal must be convinced to securely lock up their guns or remove them from their homes.

It is very important that school personnel be very aware of the most common methods of suicide so that appropriate actions can be taken to remove any available lethal means of suicide. Parents and school personnel must be alert to the precipitating events that often precede youth suicide so that appropriate interventions can take place as soon as possible to prevent the suicide from occurring. The most common precipitating events are the following:

- Severe arguments with parents
- Breakup of a romance
- Loss of a loved one
- Discipline or grade problems

## 5. ROLE OF SCHOOLS IN SUICIDE PREVENTION

There has been a growing awareness of the need for schools to be more knowledgeable about suicide prevention and to regularly in-service school personnel on the warning signs of suicide as well as on the importance of all school personnel working as a team to detect suicidal students and secure assistance for them. Many state lawmakers have passed legislation addressing prevention in the schools. Unfortunately, most of the legislation that has already been passed has involved recommendations rather than mandatory actions. The nationwide emphasis on high-stakes testing and academic accountability has also minimized the focus on suicide prevention as well as on other important mental health issues. For example, a school superintendent in Austin, Texas, commented that, given a choice of providing teachers with an in-service on preventive mental health topics or with one that

focuses on improving test scores, he would opt for the one on test scores.

It is also important to recognize that there have been attempts to hold schools accountable for the suicides of students, and school districts have been subjected to lawsuits over the suicides of students. The question is not whether the school somehow caused the suicide (regardless of whether the suicide occurred on or off school grounds) but rather whether the school took reasonable action to prevent it from occurring. The issues, as outlined by Lieberman and Davis in 2002, are as follows:

- *Foreseeability*: Did the school have reason to foresee that the young person was at risk for suicide? Were warning signs obvious?
- *Negligence*: It is negligent to fail to secure counseling assistance for a suicidal student. It is negligent to fail to notify parents when a child is suicidal.

In 1999, Poland and McCormick reviewed a number of cases where school personnel did not follow commonsense practices of notifying parents, increasing supervision, and providing referral and follow-up services. School personnel and school districts were found by courts to be liable for monetary damages. One such case in 1997 was *Wyke v. Polk County School Board* following the death by suicide of a student who school administrators knew was suicidal. The student's parents were awarded a monetary judgment. The 1991 case of *Eidseil v. Board of Education of Montgomery County* illustrated the importance of parents being notified when school personnel suspect that a student is suicidal, even though the student himself or herself denies suicidal intent.

Administrators are in a particularly difficult situation when a student is facing suspension and expulsion consequences and states, "If you expel me, then I am going home to kill myself." There have been situations where discipline incidents precipitated suicides. Administrators must be alert to suicidal ideation for students who are involved in discipline situations and must follow procedures that include parent notification. The role of the schools in suicide intervention is outlined in [Table II](#), and various suicide warning signs are provided in [Table III](#). [Figures 1 and 2](#) illustrate a sample no suicide or no harm contract and a sample parent form for notification of child's suicidal ideation, respectively.

**TABLE II**  
**The School's Role in Suicide Prevention**

- 
- *Awareness and detection:* All school personnel who supervise students (i.e., who have any form of contact with students), including administrators, teachers, counselors, psychologists, speech therapists, classroom aides, and bus drivers, should be in-serviced annually on the warning signs and what to do when they suspect a student might be suicidal. In addition, it is recommended that awareness training include personnel such as office secretaries, custodians, and cafeteria workers.
  - *Assessment:* Support personnel such as counselors and psychologists often experience anxiety when working with suicidal students and need specialized training to increase their confidence level to assess the severity level of the student's suicidal behavior. It is important that support personnel placed in this role collaborate with other professionals and maintain frequent communication with the school administrators. In 1995, Poland outlined Texas legislation stating that parental permission is not necessary prior to seeing a minor when the minor is believed to be suicidal. Poland further stressed the importance of a direct inquiry with questions such as the following:
    - Have you thought about harming yourself or trying to end your life?
    - Have you previously attempted suicide?
    - Have you thought about suicide lately?
    - How would you end your life?
 The answers to these questions are essential to know whether parents must be contacted immediately and whether the student must be kept under close supervision until parents can be notified.
  - A "no harm contract" is another essential step in suicide prevention. Copies of the signed contract should be given to the suicidal student, his or her parents, and the school administrator.
  - *Parent notification:* Parents must be made aware that it is suspected that their child is suicidal. The question is not whether to call the parents but rather what to say. The conversation might need to include only statements such as that the child is depressed and has lost the joy for life or that the child has an imminent plan to commit suicide and that parents need to come to the school immediately. Documentation of parent notification is essential. The role of the school should be outlined, and referral for additional community services should be made.
  - *Referral and follow-up services:* Students known to be suicidal should be referred to community agencies and private practitioners for more in-depth counseling than school employees can provide. School personnel, such as administrators, counselors, and psychologists, should monitor suicidal students closely, including frequent contact with the students, their parents, and any outside professionals providing treatment. Protective service agencies should be notified when parents do not seek counseling assistance for suicidal students.
- 

## 6. WHAT IF THE DEATH WAS A SUICIDE?

In the case of a student's death, it is very important for school personnel to talk with the family to verify the cause of death and to ask permission to inform the staff and students at the deceased child's school of the circumstances. The AAS's Guidelines for Postvention are an excellent source outlining the role of the schools as well as media guidelines. These guidelines, summarized in Table IV, emphasize that being truthful about the circumstances is important and that postvention is an ideal opportunity in which to raise awareness to prevent the future suicides. Fearing that attention to a recent suicide will be perceived as glamorizing the deceased and the circumstances, schools often prefer to do very little; however, it is important to deal with survivors' loss, confusion, shock, and grief by empowering them with a resolve to prevent further suicides.

A particularly troubling issue for school administrators is the question of whether to use a suicide victim's name in a memorandum to teachers and in a note to parents. The AAS guidelines specifically give examples of stating the name of the suicide victim; however, schools commonly do not provide that information, even though the cause of death is a matter of public record.

Few events in a school are more troubling than the suicide of a student. The school crisis team would greatly benefit from reviewing the AAS's guidelines for postvention that also outline the need to be very cautious about memorializing the suicide victim. School superintendents and the principal must carefully consider the requests of family and friends to memorialize the suicide victim. Permanent markers and memorial activities at school are not recommended. It is hoped that the victim's family can be convinced to donate money to suicide prevention crisis centers or to scholarship funds.

**TABLE III**  
**Warning Signs**

- *Threat*: Threats may be direct (e.g., “I want to die,” “I am going to kill myself”) or indirect (e.g., “The world would be better without me,” “Nobody will miss me anyway”). During adolescence, indirect clues could be offered through joking or through references in school assignments (e.g., creative writing, art pieces). In concrete and preoperational children, indirect clues may come in the form of acting out or violent behavior that is often accompanied by suicidal/homicidal threats.
- *Plan/Method/Access*: Suicidal thoughts must be distinguished from actual planning. The greater the planning, the greater the potential for suicide. In evaluating the suicidal potential of a student, the lethal potential and availability of the means and the level of sophistication of the plan (including the developmental level of the interviewee) must be taken into account. Familiar themes for children include running into traffic, jumping from high places, and cutting/scratching/markings their bodies.
- *Previous attempts*: Adolescent attempters are at increased risk for a variety of negative outcomes, including repeat attempts, psychiatric symptoms, and academic, social, and behavioral problems. Approximately 15% of individuals with a history of one or more suicide attempts will go on to kill themselves.
- *Final arrangements*: This behavior may take many forms. In adolescents, it might be giving away prized possessions (e.g., jewelry, clothing, journal). It seems likely that preoperational elementary children lack the cognitive skills necessary to plan for making final arrangements.
- *Depression*: When symptoms of depression include pervasive thoughts of helplessness and hopelessness, children or adolescents are conceivably at greater risk for suicide.
- *Sudden changes*: Changes in behavior, friends, or personality are warning signs.

Source. Poland and Lieberman (2002). Copyright (2002 and 1995) by the national Association of School Psychologists, Bethesda, MA. Reprinted by permission of the publisher.

I/We, \_\_\_\_\_, the parents of \_\_\_\_\_, were involved in a conference with school personnel on \_\_\_\_\_. We have been notified that our child is suicidal. We have been further advised that we should seek some psychological/psychiatric consultation immediately from the community. School personnel have clarified the district’s role and will provide follow-up assistance to our child to support the treatment services from the community.

\_\_\_\_\_  
(parent or legal guardian)

\_\_\_\_\_  
(school personnel/title)

**FIGURE 2** Sample of parent form for notification of child’s suicidal ideation. Reprinted from Poland (1995). Copyright (2002 and 1995) by the national Association of School Psychologists, Bethesda, MA. Reprinted by permission of the publisher.

In 2002, Poland and Lieberman cited literature predicting that with the occurrence of one youth suicide in a community, the chances of a second suicide increase dramatically. Teenagers in particular are more susceptible to imitating classmates who have died by suicide. The question of suicide contagion or clusters continues to be studied by the CDC.

### 7. STATUS OF STATE AND INDIVIDUAL INITIATIVES

The majority of states have some form of legislation addressing the problem of youth suicide. California has long been a leader in youth suicide prevention, but prevention initiatives have lost momentum in many states and are largely only recommendations. Many well-meaning school administrators have difficulty in finding the time for teacher training as well as in developing and implementing curriculum units to teach students about suicide prevention. Many of the youth suicide prevention initiatives currently under way at the state level are being driven by surviving parents.

I, \_\_\_\_\_, agree not to harm myself.

If I am having thoughts of harming myself or committing suicide, then I will do the following until I receive help:  
Get assistance from an adult.  
Call the Crisis Hotline at \_\_\_\_\_.  
Call the school psychologist/counselor at \_\_\_\_\_.

I understand the contract that I am signing and agree to abide by it.

\_\_\_\_\_  
(student signature)

\_\_\_\_\_ (school psychologist)

**FIGURE 1** Sample of no suicide or no harm contract. Reprinted from Poland (1995). Copyright (2002 and 1995) by the national Association of School Psychologists, Bethesda, MA. Reprinted by permission of the publisher.

**TABLE IV**  
**General Suicide Postvention Guidelines for Schools**

- 
- Plan in advance of any crisis.
  - Select and train a crisis team.
  - Verify report of suicide from collaboration with the medical examiner, police, and family of the deceased.
  - Do not dismiss school, and do not encourage funeral attendance, during school hours.
  - Do not dedicate a memorial (e.g., yearbooks, tree, bench).
  - Do contribute to a suicide prevention effort on behalf of the school or community.
  - Do contact the family, apprise them of the school's intervention efforts, and assist with funeral arrangements.
  - Do not release information in a large assembly or over intercom systems. Instead, disseminate information to faculty, students, and parents in small groups or in a classroom setting. Always be truthful.
  - Follow the victim's classes throughout the day with discussion and counseling.
  - Arrange for counseling rooms in the school building, and provide individual and group counseling.
  - Collaborate with media, law enforcement, and community agencies.
  - Emphasize the following points with media and parents:
    - Prevention is crucial.
    - No one thing or person is to blame.
    - Help is available.
    - Provide counseling or discussion opportunities for the faculty.
    - Discourage media from including front-page coverage of the suicide or including a picture of the deceased or a description of the method.
    - Discourage media from using simplistic or romantic explanation of the suicide.
    - Encourage media to include a list of suicide warning signs and specific examples of what to do to secure professional help for a suicidal individual.
- 

*Source.* American Association of Suicidology (1998) and Poland and Lieberman (2002). Copyright (2002 and 1995) by the national Association of School Psychologists, Bethesda, MA. Reprinted by permission of the publisher.

School-based suicide prevention programs fall into three categories as outlined by Kalafat and Lazarus in 2002: universal interventions, selective interventions, and indicated interventions. Universal interventions are directed to an entire population and involve increasing likelihood that administrators, faculty, and peers who come into contact with suicidal youth can more readily identify them and secure help for them,

an approach that is often referred to as "gatekeeper training." Teaching school personnel the warning signs and developing guidelines for schools to follow when students are identified as suicidal has been the most widely used approach to suicide prevention. An additional universal intervention is having students participate in a national suicide screening day where large numbers of students answer questionnaires that address depression and suicide. This program targets secondary student and is sponsored by the National Association of School Psychologists. The program includes as a first step an educational component where students are taught the warning signs and the importance of getting adult help for anyone believed to be suicidal. Students then fill out a questionnaire that asks questions about depression and suicide. Students who are identified as depressed or suicidal, based on their responses, are evaluated further and are provided with interventions. Typically, the screening is done 1 day each year at participating schools; thus, a limitation is that a student might become suicidal at a later point in the year without having reported suicidal behavior/thoughts on the screening date. Approximately 1500 schools have implemented this program and have reported increases in adolescent help-seeking behaviors.

Selective interventions target subpopulations that are characterized by shared exposure to risk factors such as making the transition to a new school and attending a school where a number of students have died by suicide.

Indicated interventions target specific individuals who have been identified as suicidal by screening procedures, self-referral, and/or referral by a concerned peer or faculty member.

Kalafat and Lazarus pointed out that it is important to view these as complementary approaches rather than competing approaches. They summarized the research on suicide prevention programs and emphasized that although early research results during the 1980s about the effectiveness of suicide prevention programs were mixed, more recent controlled analysis of suicide prevention programs in the schools has resulted in reduced suicidal behavior and in lives being saved. The authors stressed the need for additional research and linkage between school-based programs and community resources.

A very promising development at the national level was initiated in the U.S. surgeon general's "Call to Action to Prevent Suicide" in 1999 and in the U.S. Public Health Service's "National Strategy for Suicide

**TABLE V**  
National Suicide Prevention Efforts

---

Surgeon general's "Call to Action to Prevent Suicide" (1999)

- *Awareness*: Broaden the public's awareness of suicide and its risk factors.
- *Intervention*: Enhance services and programs, both population based and clinical care.
- *Methodology*: Advance the science of suicide.

National Strategy for Suicide Prevention's Goals and Objectives for Action (2001):

- *Goal 1*: Promote awareness that suicide is a public health problem that is preventable.
- *Goal 2*: Develop broad-based support for suicide prevention.
- *Goal 3*: Develop and implement strategies to reduce the stigma association.
- *Goal 4*: Develop and implement suicide prevention programs.
- *Goal 5*: Promote efforts to reduce access to lethal means and methods of self-harm.
- *Goal 6*: Implement training for recognition of at-risk behavior and delivery of effective treatment.
- *Goal 7*: Develop and promote effective clinical and professional practices.
- *Goal 8*: Improve access to and community linkages with mental health and substance abuse services.
- *Goal 9*: Improve reporting and portrayals of suicidal behavior in the entertainment and news media.
- *Goal 10*: Promote and support research on suicide and suicide prevention.
- *Goal 11*: Improve and expand surveillance systems.

---

*Note.* Goals outlined in National Strategy for Suicide Prevention are very important. However, little funding has been attached to them, and there has been little clarification of where schools fit into these national goals.

Prevention" in 2001. These actions outlined the goals listed in Table V.

## 8. SUICIDE PREVENTION NATIONAL RESOURCES

School personnel are encouraged to familiarize themselves with the initiatives in their own states as well as at the national level. They should also contact local crisis hotline centers and use them as a resource in working on youth suicide prevention in the school and community. Table VI lists national resources.

**TABLE VI**  
Suicide Prevention Resources

---

- American Association of Suicidology. [www.suicidology.org](http://www.suicidology.org). Many resources are available from AAS. Its members are a combination of professionals working in suicide prevention and laypersons who are suicide survivors. An annual convention during the spring.
- Yellow Ribbon Suicide Prevention Foundation. [www.yellowribbon.org](http://www.yellowribbon.org). Founded by parents who lost a child to suicide, this organization offers support for survivors and prevention information for schools. Local chapters are located in nearly every state as well as in several countries. The annual international convention is held in Denver, Colorado.
- Americans for Suicide Prevention. [www.afsp.org](http://www.afsp.org).
- Suicide Prevention Awareness Network. [www.spanusa.org](http://www.spanusa.org).
- International Association for Suicide Prévention. [www.med.uio.no/iasp](http://www.med.uio.no/iasp).
- National Suicide Prévention Hotline. 1-800-SUICIDE.

---

## 9. SUMMARY

The World Health Organization's report on suicide in 2000 emphasized the following protective factors to prevent youth suicide:

- Good family relations and support
- Good relationships with teachers and peers
- Social integration and participation in school and community
- Social skills and self-confidence
- Willingness to seek help when difficulties arise

These factors are important for the success and satisfaction of all children. However, they involve a partnership/collaboration among schools, parents, and community leaders. School psychologists and other mental health professionals, such as school counselors and social workers, are in a unique position to encourage and help schools to develop and implement policies and procedures to prevent youth suicide through outlining each school employee's role in the prevention effort. It is also essential to educate the students themselves about how they can help to prevent suicides. It is hoped that the National Strategy for Suicide Prevention will result in many highly funded initiatives that will clarify the role of schools, provide more mental health services to suicidal students, and encourage schools, agencies, and organizations to work together. It is also

hoped that no more schools and communities will suffer the shock, devastation, and confusion that result from the suicides of students or staff members.

### See Also the Following Articles

Behavioral Observation in Schools ■ Bullying and Abuse on School Campuses ■ Stress

### Further Reading

- American Association of Suicidology. (1998). *Suicide post-vention guidelines: Suggestions for dealing with the aftermath of suicide in the schools*. Washington, DC: Author.
- Associated Press. (2003, June 12). School shooting in Pennsylvania ends with student suicide. *Houston Chronicle*, P. A6.
- Brent, D. A., & Perper, J. A. (1995). Research in adolescent suicide: Implications for training, service delivery, and public policy. *Suicide and Life-Threatening Behavior*, 25, 222–230.
- Davis, J. M., & Brock, S. E. (2002). Suicide and the schools: Incidence, theory, research, and response. In J. Sandoval (Ed.), *Crisis counseling, intervention, and prevention in the schools* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Eisel v. Board of Education of Montgomery County*, 597 A.2nd 447 Maryland (1991).
- Kalafat, J., & Lazarus, P. (2002). Suicide prevention in the schools. In S. E. Brock, P. H. Lazarus, & S. R. Jimerson (Eds.), *Best practices in school crisis prevention and intervention*. Bethesda, MD: National Association of School Psychologists.
- Lieberman, R., & Davis, J. (2002). Suicide intervention. In S. E. Brock, P. J. Lazarus, & S. R. Jimerson (Eds.), *Best practices in school crisis prevention and intervention*. Bethesda, MD: National Association of School Psychologists.
- Moscicki, E. K. (1995). Epidemiology of suicidal behavior. *Suicide and Life-Threatening Behavior*, 25, 22–35.
- Poland, S. (1989). *Suicide intervention in the schools*. New York: Guilford.
- Poland, S. (1995). Best practices in suicide intervention. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 3, pp. 155–166). Bethesda, MD: National Association of School Psychologists.
- Poland, S., & Lieberman, R. (2002). Best practices in suicide intervention. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 4, pp. 1151–1166). Bethesda, MD: National Association of School Psychologists.
- Poland, S., & McCormick, J. S. (1999). *Coping with crisis: Lessons learned*. Longmont, CO: Sopris West.
- Poland, S., & McCormick, J. (2000). *Coping with crisis: A quick reference guide*. Longmont, CO: Sopris West.
- Portner, J. (2001). *One in thirteen: The silent epidemic of teen suicide*. Beltsville, MD: Robins Lane Press.
- Ramsay, R. F., Tanney, B. L., Tierney, R. J., & Lang, W. A. (1996). *Suicide intervention workshop* (6th ed.). Calgary, Alberta: Living Works Education.
- Shaffer, D., & Piacentini, J. (1994). Suicide and suicide attempts. In M. Rutter, L. Hersov, & E. Taylor (Eds.), *Child and adolescent psychiatry* (3rd ed., pp. 407–424). London: Blackwell Scientific.
- U.S. Public Health Service. (1999). *The surgeon general's call to action to prevent suicide*. Washington, DC: Author.
- U.S. Public Health Service. (2001). *National strategy for suicide prevention: Goals and objectives for action*. Washington, DC: Author. Available: [www.mentalhealth.org](http://www.mentalhealth.org).
- Vossekuil, B., Reddy, M., Fein, R., Borum, R., & Modzeleski, W. (2000). *U.S. Safe School Initiative: An interim report on the prevention of targeted violence in schools*. Washington, DC: U.S. Secret Service, National Threat Assessment Center.
- World Health Organization. (2000). *Preventing suicide: A resource for teachers and other school staff*. Geneva, Switzerland: Author.
- Wyke v. Polk County School Board*, 11th Cir. (1997).



# System Safety

Bernhard Wilpert

Berlin University of Technology, Berlin, Germany

1. Large Complex Sociotechnical Systems
  2. Safety Management Strategies
  3. The Role of Psychology in System Safety
- Further Reading

the safety and reliability of highly hazardous systems and began to focus on the avoidance of total system breakdowns. Applied psychology has an important role in the development and implementation of system safety strategies.

## GLOSSARY

*event analysis* The systematic approach to identifying factors that contributed to an event (incident, accident, or near accident) in organizations.

*large complex sociotechnical systems* Very large industrial installations of high complexity with a concomitant increased hazard level for the environment and people.

*safety management systems* The portfolio of concrete structural and organizational measures initiated by management in order to maintain safety and reliability of a given sociotechnical system.

*system safety* The quality of a system that allows the system to function without major breakdowns under predetermined conditions with an acceptable minimum of accidental loss and unintended harm to the organization and its environment.

Applied psychology traditionally studied safety as a problem of individuals in their respective workplaces and was concerned with the avoidance of workplace accidents. The growth of large and highly complex sociotechnical installations brought with it a drastic increase in concomitant hazards. Thus, applied psychology enlarged its scope of reference to include

## 1. LARGE COMPLEX SOCIOTECHNICAL SYSTEMS

During the second half of the 20th century, the number and complexity of very large technical installations increased greatly. Along with their growth was a concomitant increase in hazardousness. Examples of these installations include those of the chemical and petrol industry, civil and military aviation, railway systems and space industry, and nuclear energy production—that is, installations in unforgiving environments in which the accumulation of high energy or toxic substances must be controlled to avoid catastrophic consequences for people and environment. Their characteristic features include the involvement of multiple actors, high levels of complexity, tightly coupled subsystems and production processes with their concomitant opaqueness and control problems, and comparatively high hazard levels. The thorough analysis of any major industrial accident (e.g., Three Mile Island, Chernobyl, Challenger, the collision of two jumbo planes in Tenerife, and the train accident in Eschede, Germany) invariably shows that it is never a single component failure or a single human error that causes the accident. Rather, it is the

often little understood interaction of factors from various system levels that “causes” accidents: failures on the technical (component) level, erroneous individual (operator) action, inadequate work team activities, and dysfunctionalities in the organizational and extraorganizational environment.

Münsterberg, in his seminal publication *Psychology and Industrial Efficiency*, showed in 1913 that safety in work systems has from its very beginning been a preoccupation of applied psychology. However, its focus tended to be occupational safety of individuals in their specific workplace settings. “Human factors” in the tradition of ergonomics and occupational safety were considered to refer to all characteristics at the immediate man–machine interface that enable proper and safe operations. Although recently there has been a shift in theory and practice of safety, health, and environment to more holistic approaches encompassing the whole organizational system, large-scale, highly hazardous systems pose novel challenges for psychology in terms of theory, methodological analysis, and safety-oriented intervention practices. Safety and reliability of complex systems depend on failure avoidance and effective process and production control. From the analysis of major accidents, we know that human action always plays a major role in triggering accidents as well as mitigating their consequences.

Work and organizational psychology has for a long time neglected to focus on the role of humans in safeguarding highly complex sociotechnical systems. In 1974, Graumann noted that one reason may have been the preoccupation with individual organismic perspectives at the expense of studying the interaction of humans with “the world of things.” The focus of organizational psychology was occupational hazards and individual workplace accidents as influenced by individual constructs such as motivation, perception, attitude, and decision making. The world of things (i.e., complex technologies and large-scale installations) was seen to fall into the domain of engineers, who traditionally tended to focus on component failure. However, neither the exclusive individual or workplace orientation nor the limited preoccupation with technical aspects help to derive a proper understanding of the genesis of system breakdowns.

The comparative neglect of the multitude of human and physical/environmental factors in producing intended or unintended outcomes is surprising because there are at least two important theoretical strands in applied psychology that shed crucial light on these relations: Kurt Lewin’s field theory and the

sociotechnical systems approach of the Tavistock Institute of Human Relations. They have been developed for more than half a century, and both of these perspectives call for a comprehensive consideration of all factors that impact a given outcome under study, including the behavior of people (e.g., safety relevant actions, commitment, and questioning procedures) or systems (e.g., efficiency of goal getting, safety, reliability, and profitability). These factors may be located on various systems levels—the individual, the work group, the organization, the interorganizational, or the technological. For traditional psychology, however, such a holistic systemic thinking is rare. The focus is traditionally restricted to the individual subsystem.

Regarding system safety, using Amalberti’s classification we can distinguish three safety levels in different systems:

Level 1 refers to nonprofessional domains in which the safety level is lower than  $10E-3$ ; for example, in situations such as mountaineering in the Himalayas, with a survival chance of only 30% after three expeditions.

Level 2 corresponds to systems that over time reach a safety level of  $10E-5$ , such as street traffic, anesthesiology, and helicopter flights.

Level 3 covers systems with a safety level near or higher than  $10E-6$ —so-called ultracomplex and ultrasafe systems such as nuclear power plants, civil aviation, and space travel.

The nature of ultrasafe, large-scale technical systems internationally brought about a paradigmatic shift in safety thinking. Traditional workplace safety concerns tried to avoid individual accidents in work settings by creating safe working conditions. Ergonomics, work medicine, work and organizational psychology were the disciplines in demand. The emergence of large-scale technical systems required that specialists in psychology, engineering sciences, and political and legal sciences cooperate in order to prevent “systems disasters”—large industrial catastrophes due to total systems failures (e.g., the Chernobyl accident). In agreement with the formulation of Roland and Moriarty, we can define system safety as a quality of a system that allows the system to function without major breakdowns under predetermined conditions with an acceptable minimum of accidental loss and unintended harm to the organization and its environment. The critical elements of this definition refer to the systemic quality of safety in the sense that all subsystems of a given organization are considered

contributors to the intended outcome: (i) safety as avoidance of system breakdown; (ii) the system's functioning is linked to predetermined performance criteria; and (iii) the system performance is measured against a socially acceptable limit of uncontrollable, if negative, side effects. One consequence of such a holistic view of system safety is that "human factors" not only relate to the immediate man-machine interface but also human action on all system levels in as much as they contribute to the critical outcome variable: safety.

## 2. SAFETY MANAGEMENT STRATEGIES

In 1991, Rasmussen, in a somewhat abstract manner, distinguished three different approaches to the control of safety. First, the feedback strategy reactively attempts to control relatively low-risk and unassociated hazards by empirically monitoring acceptable safety goals as measured by accident and injury rates. Such a strategy may be appropriate in traffic and construction systems, which have a slow pace of change. It is principally oriented toward learning from past events and data suggesting specific behavior adjustments.

Second, the feedforward strategy concentrates on control through enlightened design and planning of operations in the sense of analyzing probable system hazards in advance and implementing requisite safeguards into the system. The approach seems appropriate for highly hazardous systems with slow change such as railway systems, in which methods of probabilistic analyses (e.g., Probabilistic Safety Analysis) are fruitfully employed.

Finally, the feedback and feedforward strategy combined tries to learn from past events. It integrates such learning into probabilistic methods and is oriented toward evolutionary choice, which makes it particularly appropriate for ultrasafe systems such as nuclear power plants.

In a different approach, Reason identified three safety management models: a person model, which focuses on individual work-related competencies and behavior; an engineering model, with emphasis on improving technology and reducing man-machine mismatches to enhance safety; and an organizational model, in which avoidance of latent weaknesses (insufficient staff training, inadequate managerial decisions regarding division of work, and inappropriate incentive systems) in the organization are to be minimized. Applied psychology

has significant contributions to make for all three of Reason's models: in the person model, action regulation theory; in the engineering model, physiological psychology, with its insights into the characteristics (possibilities and limits) of the sensory nervous system of operators that need to be matched with workplace demands in order to maximize safety; and in the organization model, leadership theories, with their implications for requisite training measures.

Different managerial safety strategies have operationally been made more precise under the term "safety management systems." The term covers a large number of concrete activities and measures, such as the outlining of a precisely defined safety policy by management, the clear distribution of responsibilities among staff and departments, the adequate allocation of resources to organizational units, application of procedures to assess operational risks, learning from operational experience through incident reporting and incident analyses, competence-enhancing training measures, monitoring of safety-related data, and internal and external audits and peer reviews.

## 3. THE ROLE OF PSYCHOLOGY IN SYSTEM SAFETY

System safety is not the exclusive interest of psychologists, as demonstrated by the founding of the System Safety Society in 1964, which is international in scope and has members from various professional backgrounds. However, there are five important reasons why the field of psychology should be involved in issues of system safety.

First, an important, if not the only, precondition of individual growth is the personal, bodily integrity of working people and their environment. This is the ethical mandate for psychologists to contribute to this precondition.

Second, since people are almost always involved in triggering or mitigating accidents, it seems to be self-evident that the scientific discipline that deals with human perception, competence, motivation, and action is challenged here as well. This is the mandate that follows from the self-image of the psychological discipline.

Third, no other scientific discipline is as prepared and geared to deal with many of the problems raised by highly hazardous organizations as is psychology. Psychology deals with the physiological bases of human action (e.g., information-processing capacities),

cognitive dynamics (perceptual distortions), organizational conditions of work, strain and stress, group dynamics, and risk communication.

The fourth reason concerns the enormous hazard potential implied in highly hazardous organizations. Hence, from an ecological standpoint, one must not leave the area to engineers alone.

Finally, technological development provides an important reason: The increasing complexity of new technological systems calls for an understanding of technology that transcends material objects (machines, software, etc.) traditionally understood as technology. Today, even engineers understand technology in broader terms: Technology comprises the material artifacts such as machines but also the human action facilitated by them as well as the set of human actions in which these artifacts are used. In other words, the conditions of emerging material technological systems as well as their use must be considered as technology. This is the interdisciplinary contribution of psychology that is called for.

Three examples illustrate the potential contributions of applied psychology to system safety beyond the general reasons given previously.

### 3.1. Event Analysis and Event Reporting

The feedback strategy of safety control suggests that learning needs to take place in highly hazardous organizations through a systematic analysis of the organization's operational experience. Individuals learn from mistakes, whereas organizations learn from incidents. Hence, it has become a generally accepted tenet in highly hazardous industries that it is necessary to thoroughly analyze, document, and report events (incidents and accidents). Given that the analysis of events is a socially accepted reconstruction of the processes leading up to the event, the analysis implies the need to conjecture about causal relations that usually goes beyond the immediately observable facts. Such attributional processes are fraught with potential biases that are well analyzed and understood by various subdisciplines of psychology, such as cognitive psychology and social psychology. Biases that result from general human attributional tendencies in the search for causal relations include the following:

- Premature hypothesis formation
- Overweighting of immediately present factors and neglect of distant ones ("out of sight—out of mind")

- Monocausal thinking
- The fundamental attribution error (causes foremost attributed to acting persons rather than situational factors)

The Research Center Systems Safety of the Institute of Psychology and Ergonomics, Berlin University of Technology developed a methodology, Safety through Organizational Learning (SOL), that is available as a computer-assisted tool. SOL is used by German and Swiss nuclear power plants as the method of choice for analyses of their events.

### 3.2. Training and Personnel Development

Maintaining and improving the competence of managers and staff in highly hazardous industries is a constant and increasingly difficult challenge for applied psychology due to the rapid social and technological changes that affect these industries. Deregulation, mergers, increased competition, changes in regulatory practices, and changing public responses are a few of the catchwords that mark the new environmental trends. Applied psychological know-how, especially from organizational psychology and organizational development, is required in at least two areas: analyses of changing work demands and the design of requisite training measures. Applied psychology has in many countries made important contributions to the improvement of system safety via management training and safety-oriented organizational change interventions.

### 3.3. Safety Culture

Safety culture is a relatively new concept that originated in response to the Chernobyl catastrophe when international analysts and experts were puzzled about the factors that caused the dramatic events of 1986 in the Ukrainian nuclear power plant. The origin of the concept lies in anthropology and organization sciences. Its fundamental assumption is that safety-oriented behavior can only be fully understood by the complex interaction of basic assumptions, values, implicit and explicit norms, and artifacts in the organization. Although a growing number of organizational psychologists have made efforts to obtain a better understanding of the role of organizational cultures, the term safety culture requires further theoretical and methodological refinement. However, the term has gained remarkable currency in many highly hazardous industries far beyond the nuclear industry: aviation, railway safety, chemical

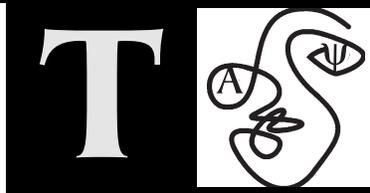
process industry, and medicine. Limited knowledge surrounding the concept of safety culture calls for the decisive involvement of applied psychology.

### See Also the Following Articles

Safety and Risks: Errors and Accidents in Different Occupations

### Further Reading

- Amalberti, A. (2002). Revisiting safety paradigm to meet the safety challenges of ultra complex and safety systems. In B. Wilpert, & B. Fahlbruch (Eds.), *System safety—Challenges and pitfalls of intervention* (pp. 265–276). Amsterdam: Pergamon.
- Fahlbruch, B., & Wilpert, B. (1997). Event analysis as problem solving process. In A. Hale, M. Freitag, & B. Wilpert (Eds.), *After the event—From accident analysis to organisational learning* (pp. 113–130). Amsterdam: Elsevier.
- Fahlbruch, B., & Wilpert, B. (1999). System safety—An emerging field of I/O psychology. In C. L. Cooper, & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 55–93). Chichester, UK: Wiley.
- Graumann, C. G. (1974). Psychology and the world of things. *Journal of Phenomenological Psychology*, 4, 389–404.
- Reason, J. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate.
- Roland, H. E., & Moriarty, B. (1990). *System safety engineering and management*. New York: Wiley.
- Wilpert, B., & Itoigawa, N. (Eds.). (2001). *Safety culture in nuclear power operations*. London: Taylor & Francis.



# Teaching Effectiveness

Jac J. W. Andrews

University of Calgary, Calgary, Alberta, Canada

1. Introduction
  2. A General View of Teaching
  3. Indicators of Effective Teaching
  4. Assumptions about Teaching Effectiveness
  5. Moving from Effective Teaching to Excellent Teaching
  6. Conclusion
- Further Reading

## GLOSSARY

**direct instruction** Teacher-guided instruction aimed at the mastery of basic skills and concepts.

**extrinsic motivation** The desire to engage in some behaviors to gain outside rewards such as praise, high grades, and money.

**intrinsic motivation** The desire to engage in some behaviors to gain inside rewards such as satisfaction and insight.

**personality trait** A stable characteristic of a person that is inferred from a tendency to behave in particular ways in certain situations (e.g., assertive, outgoing, withdrawn).

**self-concept** A person's awareness of his or her characteristics and the ways in which the person is like and unlike others.

**self-efficacy** A person's judgment of his or her competence to execute courses of actions that are required to deal with certain situations.

**self-esteem** The value that a person puts on himself or herself and the value that the person puts on his or her behavior.

**strategy** Organized skills and processes by which a person plans and regulates his or her behavior and accomplishes particular goals.

Effective teaching is indicated by certain behaviors exhibited in particular situations that lead to positive student outcomes. Essentially, it is when approaches to teaching increase students' knowledge and facilitate their learning. In addition to their content expertise and methodological technique, excellent teachers approach their teaching in ways that are congruent with their students' approach to learning and result in increased student understanding.

## 1. INTRODUCTION

Although there has been a substantial amount of research over the years in the area of teacher effectiveness, there is still little consensus about how to define and identify effective teaching. One of the reasons for the different opinions about teaching effectiveness is that teaching is such a complex, multidimensional, and idiosyncratic process. However, there are some generally agreed-on notions of teaching and indicators of effective teaching. This article focuses on some of the basic conceptions of teaching and some general understandings of what constitutes effective teaching. From a discussion about teaching and teacher effectiveness, the article concludes with some ideas about moving from effective teaching to excellent teaching.

## 2. A GENERAL VIEW OF TEACHING

Teaching is a process aimed at imparting knowledge, facilitating learning, and increasing understanding. Two factors that influence teaching are teacher characteristics and situational-contextual factors (Fig. 1). Teacher characteristics include personality traits, knowledge, abilities, experience, values, and beliefs. Teachers have different types of personality traits, abilities, knowledge, and experiences as well as particular values and beliefs that affect their approaches to teaching. Hence, teachers' approaches to teaching are somewhat dependent on their personal backgrounds, competencies, and viewpoints. Coincidentally, it is generally understood that teachers' competencies vary in relation to their acquired skills and self-efficacy. Moreover, it is generally well known that the values and beliefs of teachers influence their perceptions and judgments and also affect their behavior in the classroom. In 1992, Pratt noted that teachers' values and beliefs are anchored in cultural, social, historical, religious, and personal realms of meaning, and although there are different conceptions about teaching, they are not mutually exclusive. For example, one teacher might have the view that teachers have the primary

responsibility for developing students' knowledge. Furthermore, this teacher might believe that students should be rewarded for good work and penalized for lack of effort and poor work. In addition, this teacher might believe that teachers need to direct their students with respect to what they need to learn. Another teacher might believe that the responsibility for learning lies primarily within the learners. This teacher also might believe that students need to know how to learn and understand the importance of their learning and behavior. In contrast to the second teacher, the first teacher would appear to value the product of learning over the process of learning and to consider students as passive learners rather than active learners. In this regard, it is likely that the first teacher would use direct instruction as an approach to teaching and use extrinsic reinforcements to motivate students to learn. The second teacher would likely use an interactive and collaborative approach to teaching and foster an intrinsic motivation for learning within students.

Teachers' approaches to teaching are also mediated by situational and contextual factors. For example, how one teaches will be influenced by things such as the various characteristics of the community

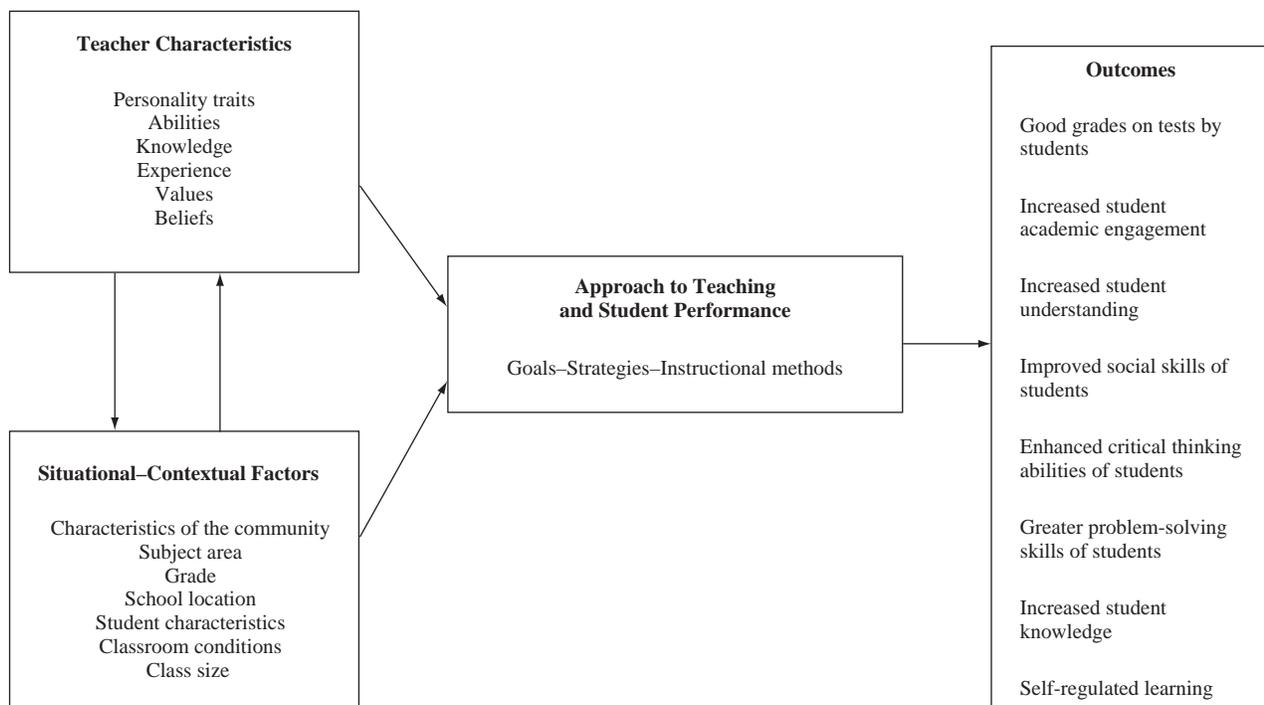


FIGURE 1 General model of teaching.

(e.g., social, political, cultural), the subject area being taught, the grade of the students, the location of the school (e.g., rural or urban), the school district policies, the knowledge and skills of the students, the physical conditions of the classroom, and the class size. For example, teachers will have classrooms that include students with different learning and behavioral styles, abilities, and difficulties that have specific needs and require diverse learning goals, instructional strategies, and methods. Teachers who have students with advanced cognitive abilities must provide opportunities to enrich students' thinking, extend their academic achievement, facilitate their independent learning, and maximize their leadership capabilities. In contrast, teachers who have students with learning disabilities must focus on ways in which to improve their learning strategies, develop their organizational skills, increase their academic self-concepts, and develop their ability to recognize, discriminate, and interpret information. Hence, the various characteristics of students in these classrooms will influence the approach to teaching as well as the outcomes of teaching. For example, in the first case, the general aim is to build on the students' strengths, whereas in the second case, the aim is primarily to remediate the students' weaknesses.

The interaction of teacher characteristics and situational-contextual factors affects the goals, strategies, and instructional methods adopted by teachers in their approaches to teaching. Although there are a number of teaching approaches, the literature indicates that teachers have two primary orientations to teaching: an orientation toward transmitting knowledge and an orientation toward facilitating learning and increasing understanding. Accordingly, each orientation affects the teaching strategies and instructional methods used by a teacher and influences how students approach their learning. For example, if the teacher's goal of teaching is to transmit and increase knowledge, he or she will likely focus on the content (i.e., basic concepts and facts), provide structured learning experiences, and present information in ways that could be reproduced through rote learning. On the other hand, if the teacher wants to maximize students' understanding, he or she will likely focus on the integration and elaboration of information in ways that could facilitate meaningful learning and critical thinking.

Typically, the outcomes of teaching are primarily related to student performance and are the result of teacher characteristics, situational-contextual factors, and the teaching approach. Some of these outcomes

could include good grades on tests, increased academic engagement, improved social skills, enhanced critical thinking abilities, greater problem-solving skills, and increased knowledge.

### 3. INDICATORS OF EFFECTIVE TEACHING

#### 3.1. Teacher Characteristics

In light of the preceding framework of teaching, effective teachers tend to be typified as enthusiastic, charismatic, caring, motivational, fair-minded, supportive, flexible, outgoing, accessible, organized, and adaptable. In terms of abilities and skills, effective teachers tend to prepare for teaching by establishing goals and objectives for their students, determining students' learning strengths and weaknesses, deciding on appropriate instructional strategies and methods, and planning for assessment and evaluation of student performance. During teaching, effective teachers match appropriate learning tasks with students' abilities, execute systematic instructional procedures routinely (e.g., review previously learned concepts and skills, show students the skills they need to be able to perform successfully, explain to students how they can apply these skills to their learning tasks, monitor students' progress, provide feedback to students regarding their learning and development), provide sufficient time for students to engage themselves in their learning, help students to generalize their learning to other situations, communicate with students in ways that promote positive self-concepts, and emphasize student effort and self-control. After teaching, good teachers assess the effectiveness of their teaching strategies and methods, determine the extent to which their goals and objectives have been met, and reevaluate the learning needs of their students.

A notable characteristic of good teachers is their ability to use effective classroom management techniques. For example, effective teachers introduce content and procedures gradually, provide enjoyable and meaningful activities in which students can engage, establish warm and secure classroom environments, use reasonable work standards, ensure high rates of success in assigned activities, anticipate and deal well with problems (e.g., late arrival of students to the classroom), collectively develop rules and routines with the students, promote the social competence of students, stop disruptive behavior appropriately and

immediately, and organize and arrange a classroom that is conducive to student cooperation, active learning, and communication.

Effective teachers tend to have diverse and experience-based knowledge. There is some evidence to suggest that effective teachers have command of the subject area being taught and have a good understanding of students' learning capabilities. Concomitantly, they know how to build on students' background experiences and facilitate their academic and social learning.

### 3.2. Situational–Contextual Factors

Teaching effectiveness is determined in part by teachers' ability to deal with contextual demands. For example, a good teacher is able to modify his or her teaching approach to adapt successfully to the needs of the students and the requirements of different situations (e.g., teaching in elementary school vs teaching in high school, teaching undergraduate courses versus teaching graduate courses at a university). To some degree, a requisite of effective teaching is being able to determine the situations in which one teaches most effectively. There is some evidence in the literature that teaching effectiveness is situation specific; that is, effective teachers are not necessarily good at teaching all subject areas. Moreover, how effective a teacher is viewed by students may be somewhat dependent on how personable the teacher is and on how meaningful and valuable the students consider the subject area that is being taught. Hence, it is important to keep in mind that because the aim of teaching is to facilitate learning and understanding, the learners' perspective must also be included in any attempt to fully appreciate the nature and scope of an effective teaching–learning transaction.

### 3.3. Approach to Teaching and Student Performance

There is support in the literature for the view that the quality of students' learning is significantly affected by teachers' approach to teaching. In general, effective teachers tend to take an approach to teaching that encourages students' active engagement in learning, fosters self-regulated learning, results in both increased competency and understanding, and makes learning meaningful, interesting, and enjoyable. Teaching effectiveness is about spending time interacting with students, engaging them in meaningful tasks, and developing their ability to make their own learning

choices. It is also about knowing and addressing the individual needs of students and having high expectations for their learning.

## 4. ASSUMPTIONS ABOUT TEACHING EFFECTIVENESS

In consideration of the preceding discussion and the expansive literature on teaching effectiveness, the following general assumptions are made:

- Although the literature points out a number of characteristics of effective teaching, it is likely that not all effective teachers perform at the highest level in all of these areas.
- Effective teachers are likely able to excel in their teaching without using all of the highly regarded effective instructional strategies and methods.
- Each teacher can be effective in different ways, using different teaching approaches or a combination of teaching approaches with different instructional strategies and methods.
- The outcomes achieved by effective teachers (e.g., students' increased knowledge and grades on tests) may vary from time to time due to situational–contextual factors.
- Knowledge of factors associated with effective teaching is not enough. Teachers must be well trained and mentored.

## 5. MOVING FROM EFFECTIVE TEACHING TO EXCELLENT TEACHING

Excellent teaching involves processes and practices beyond the descriptors typically associated with effective teaching (e.g., enthusiasm, availability, organization, knowledge of subject matter). It involves more than just content expertise and effective technical performance. Excellent teaching is about focusing on high-quality outcomes (e.g., increased student understanding, self-regulated learning, independent thinking) and ensuring that teaching approaches and students' learning approaches are congruent in achieving these outcomes. In this regard, how students approach their learning is as much a concern for excellent teachers as are the situational–contextual factors and approaches to teaching.

In 1996, Andrews and colleagues discovered that excellent teachers appear to have a conceptual basis for their teaching and value their teaching as well as their students and subject matter. They tend to adopt a meaningful approach to teaching so as to achieve high-quality learning outcomes. In this regard, excellent teachers seem to measure their success in relation to how well their students engage in critical thinking and take responsibility for their own learning. Excellent teachers appear to select teaching approaches that fit with their personalities, abilities, values, and beliefs as well as situational–contextual factors such as subject area and student characteristics.

Excellence in teaching is multifaceted and difficult to achieve. The literature suggests that excellent teachers monitor not only students' misunderstandings but also the incongruencies between teacher and student perceptions of the teaching–learning transaction and desired outcomes. Excellent teaching takes a holistic perspective of the teaching–learning transaction and ensures an integration of process and outcome. It results when there is congruence between teachers' approach to teaching and students' approach to learning.

## 6. CONCLUSION

Teaching is an instructional process that involves the implementation of strategies and methods relative to particular situations and contexts that are designed to get learners to achieve particular outcomes. Effective teaching is about engaging in particular behaviors that have been shown to lead to desirable student learning outcomes and adapting these behaviors to accommodate diverse learners and educational contexts. Excellent teaching is more than just having effective

teaching characteristics and technical performance and being responsive to individual and contextual factors. It is linking students' ways of learning with teachers' ways teaching so that learning becomes authentic, meaningful, and enriching.

### See Also the Following Articles

Educational Achievement and Culture ■ Intrinsic Motivation and Self-Determination ■ Learning ■ Literacy, Improvement of ■ Organizational Justice ■ Reading, Teaching of ■ Traits ■ Writing, Teaching of

### Further Reading

- Andrews, J., Garrison, D. R., & Magnusson, K. (1996). The teaching and learning transaction in higher education: A study of excellent professors and their students. *Teaching in Higher Education, 1*(1), 81–103.
- Beishuizen, J. J., Hof, E., van Putten, C. M., Bouwmeester, S., & Asscher, J. J. (2001). Students and teachers' cognitions about good teachers. *British Journal of Educational Psychology, 71*, 185–201.
- Good, T. L., & Brophy, J. E. (1994). *Looking in classrooms*. New York: HarperCollins.
- Lortie, D. C. (1975). *Schoolteacher*. Chicago: University of Chicago Press.
- Minor, L. C., Onwuegbuzie, A. J., Witcher, A. E., & James, T. L. (2002). Preservice teachers' educational beliefs and their perceptions of characteristics of effective teachers. *Journal of Educational Research, 96*, 116–127.
- Pratt, D. (1992). Conception of teaching. *Adult Education Quarterly, 42*, 203–220.
- Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.
- Young, S., & Shaw, D. G. (1999). Profiles of effective college and university teachers. *Journal of Higher Education, 70*, 670–686.



# Territoriality

Robert Sommer

University of California, Davis, California, USA

1. The Psychological Concept of Territoriality
  2. Methods of Measurement
  3. Research Findings
  4. Applications
- Further Reading

applications have been in design, urban planning, and crime prevention.

## 1. THE PSYCHOLOGICAL CONCEPT OF TERRITORIALITY

### GLOSSARY

**defensible space** An application of territoriality to public housing that creates areas under the control of residents.

**home range** An area habitually traversed by an individual or group but not personalized or defended. Home range may be a network of familiar paths and destinations.

**jurisdiction** A space temporarily controlled by an individual or group.

**territory** A fixed geographical space marked, defended, and used for life-sustaining activities. At the psychological level, human territory is linked to concepts of personalization and privacy. Turf is a colloquial synonym for territory. The concept has a broader meaning in political science, sociology, and geography, in which it involves national boundary and migration issues.

Territories are areas marked and defended by their owners and used for important life-sustaining activities. Although the concept originated in ethology, there are many differences between human and animal territories. Research on human territory is conducted in real-world settings rather than in the laboratory. The most fruitful

Most animals in the wild have fixed geographic areas called territories that they mark, defend against intruders, and use for vital life-sustaining activities. Marking by animals is done in various ways. Some deer species use a glandular secretion rubbed against trees, bears often make claw marks on trees, and birds sing. Humans also mark or personalize and defend spaces important to them—the two major aspects of territory (Fig. 1). However, because humans rely on symbolic means of marking and employ outside agencies to defend their spaces, many differences exist between human and animal territories. Human territories may be less rooted in survival needs than in a desire for status, privacy, and solitude. There may be a complex system of ownership and control in human territories that includes a distant owner, a mortgage holder, a local manager, and tenants. People hire others to patrol and defend their territories (Fig. 2). If the possession is transient, such as when someone places a coat over a chair at a restaurant or a teacher controls a classroom for an hour each day, it seems more useful to describe this temporary occupation as a jurisdiction than as a territory (Fig. 3).



**FIGURE 1** Graffiti announces territorial control of a beach.

In Altman's classification system, primary territories are central to the day-to-day lives of the occupants, clearly marked as theirs, and controlled by them on a regular basis. Examples are a person's home and a private office. Secondary territories are less central and exclusive in a person's life. A neighborhood coffee shop may be a secondary territory because the space, although associated with a specific neighborhood and clientele, is not marked as belonging to a single group and is open to others. Public territories such as sidewalks, streets, and parks, although owned, marked, and policed by a municipality or agency, are open to all those who obey rules of occupancy that may limit usage to certain times and activities.

Territories contribute to a stable social order, sharpen survival skills as occupants patrol and defend them, increase security because people on their turf know its resources and dangers, and define individual and group identity. Territories increase predictability



**FIGURE 2** Physical barriers used to keep out intruders.



**FIGURE 3** Transient occupations are better described as jurisdictions than territories.

because people know where to find others, although this may be becoming less true in an era of aspatial and asynchronous digital communication. Because the concept of territory originated in ethology, some of the most promising theories to explain why humans mark and defend territories employ a Darwinian (functional) framework, in terms of the role of territories in natural selection. There are also social learning theories that emphasize its role in the social order and place attachment.

## 2. METHODS OF MEASUREMENT

Although there have been laboratory experiments with animal territories, it is difficult to conduct them with humans. Most studies of human territories are done in the field using natural observation, surveys, and field

experiments. Invading someone's territory as part of an experiment can be stressful and raises ethical issues.

### 3. RESEARCH FINDINGS

Places with strong territorial definition, such as well-maintained dwellings, tended yards, and many signs of occupancy are less likely to be burglarized and vandalized. Clear territorial definition and many signs of personalization deter criminals, who prefer indistinct border areas where they can remain anonymous.

The more strongly a territory is valued, the more actively it will be defended. Neighbors play an important role in territorial defense.

Dominant individuals in a group or organization are likely to have larger and more suitable territories. Assigned spaces typically reflect the existing dominance order, with higher status individuals receiving the most desirable locations.

People identify themselves with territories, and any threat to the territory becomes a threat to the self. Urban renewal projects are often viewed by local residents as the destruction of a community and a diminution of self.

People on their turf are more opinionated, assertive, and in control. They are more likely to get their way in an argument on their own turf. When children are off their territories, they tend to remain close to a parent. As they become more familiar with a setting, distance between them and the parent increases. Also, males tend to claim larger territories than do women.

Microterritories can be created in small spaces, such as the division of a shared office or dormitory room into two separate territories with each person's belongings on separate sides of the room. The belongings become the territorial markers reinforcing each individual's ownership of half the space. Such microterritories are found in institutional settings in which individuals or groups take possession of certain spaces and keep out intruders. Staff may reinforce such microterritories in order to maintain a stable social order.

## 4. APPLICATIONS

### 4.1. Defensible Space

This concept was developed by city planner Oscar Newman to define how public and semipublic areas in public housing can be brought under the control of

residents. The two major components of defensible space are surveillance and territoriality. Good surveillance is achieved by proper positioning of windows and video cameras. Residents' feelings of territoriality are enhanced by real and symbolic barriers that mark areas as belonging to particular buildings. Widely used in public housing, defensible space is credited with reducing crime, increasing neighborhood cohesion, and raising property values. This concept is the basis of Crime Prevention through Environmental Design.

### 4.2. Home Field Advantages

Sports teams play better on their home fields than away. This has been found in football, baseball, hockey, basketball, and soccer. Team members are more familiar with the home field and its idiosyncrasies; are adapted to local weather, lighting, and wind conditions; have local fans cheering for them; can live at home; and are less debilitated by travel. A team will work and play hard an entire season to gain home field advantage for playoff games. Sports owners encourage place identification on the part of both team members and fans, defining the issue in terms of team spirit and fan loyalty.

### 4.3. Tracking Gang Activities

Urban gangs explicitly mark their turf boundaries using graffiti showing gang symbols. Police departments have officers trained to read wall writing to know what is happening in neighborhoods and head off potential conflict. When symbols from one gang appear on the turf of another gang, authorities may intervene quickly to head off gang warfare. "187" is graffiti code for homicide, a particularly ominous display. Prison authorities are cognizant of informal territories, and new inmates may be cautioned by fellow inmates or staff to avoid sections of the yard or dining hall controlled by gangs. Turf disputes among drug sellers may provoke multiple incidents of violence because of the large amounts of money involved, thereby setting off a cycle of retribution and further violence. Defining the turf boundaries of criminal gangs helps police prevent conflict and make arrests when it occurs.

### 4.4. Dispute Resolution

Even without established gangs in an area, informal groups may establish spatial boundaries and defend them against intruders. Portions of a school cafeteria may become the de facto territories of cliques and are

off-limits to outsiders. In coastal communities, good fishing or surfing spots may be considered territories by local residents. Graffiti marks a beach for “Locals only.” Often described as localism, this attitude can provoke individual or group violence when a stranger intrudes. Sometimes the authorities will formally divide contested space, such as by establishing separate areas for different activities. Segregation is also used to reduce conflict in wilderness recreation, where the main conflict is between motorized and pedestrian visitors. Separate winter trails are established for snowmobiles and cross-country skiers and separate summer trails for hikers and off-road vehicles.

#### 4.5. Urban Planning

It is important to respect territorial boundaries in locating city services. Clients should not be expected to cross through hostile neighborhoods to reach social agencies. A park or playground built between territories of rival gangs is likely to be underutilized and vandalized. Territoriality is included in the “minor incivilities” and “broken windows” theories of community development. Residents are less likely to maintain places that are deteriorated, litter strewn, and over which they lack control. Pride in residency and ownership are concomitants of territoriality.

Design can encourage territories, such as by closing streets in a community or using different color groupings or plantings in a housing development. This gives residents a greater sense of identity and control. Design can also be used to discourage territories in public space, such as a library, a cafeteria, or bus, in which a limited area must be shared and management does not want personal belongings spilling over into adjacent seats.

#### See Also the Following Articles

Environmental    Stress ■ Personal    Space ■ Privacy  
 ■ Residential Satisfaction and Perceived Urban Quality

#### Further Reading

- Altman, I. (1975). *The environment and social behavior*. Monterey, CA: Brooks/Cole.
- Brown, B. B. (1987). Territoriality. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 505–531). New York: Wiley.
- Gifford, R. (2002). *Environmental psychology* (3rd ed.). Colville, WA: Optimal Press.
- Taylor, R. B. (1988). *Human territorial functioning*. New York: Cambridge University Press.



# Test Anxiety

Moshe Zeidner

University of Haifa, Haifa, Israel

1. Basic and Conceptual Issues
  2. Measurement and Assessment of Test Anxiety
  3. Anxiety and Cognitive Performance
  4. Coping, Interventions, and Clinical Considerations
- Further Reading

## GLOSSARY

*cognitive-focused interventions* Emphasize the mediating role of cognitive processes in sustaining or eliminating test anxiety and refer to a wide array of therapeutic approaches directed toward modifying the worry and irrational thought patterns of test-anxious clients.

*debilitating anxiety* Refers to anxiety that is predominantly harmful to task performance.

*emotionality* Consists of perceptions of autonomic reactions evoked by evaluative stress.

*emotion-focused interventions* Aim primarily at reducing the arousal and heightened emotional reactions of test-anxious people when faced with stressful evaluative situations.

*situation-specific personality trait* In the context of test anxiety research, refers to the individual's disposition to react with extensive worry, intrusive thoughts, mental disorganization, tension, and physiological arousal when exposed to evaluative situations.

*test anxiety* The set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or loss of competence on an exam or similar evaluative situation. Test-anxious behavior is typically evoked when a person believes that his or her intellectual, motivational, and social capabilities are

taxed or exceeded by demands stemming from the test situation.

*worry* Refers primarily to cognitive concern about the consequences of failure.

This article examines test anxiety theory, measurement, performance, and clinical parameters. First, it provides a historical backdrop for test anxiety research and presents a number of conceptualizations and distinctions, delineating key facets and components of test anxiety. Next, it discusses measurement and assessment issues, focusing on self-report and alternative assessment procedures. Then, the anxiety–performance interface is discussed, noting key moderating and mediating factors. It concludes by presenting a number of treatment and intervention programs and clinical considerations when treating test-anxious clients.

## 1. BASIC AND CONCEPTUAL ISSUES

### 1.1. Overview

Tests and evaluative situations have emerged as a potent class of anxiety-evoking stimuli in our society, which bases many important decisions relating to an individual's status in school, college, and work on tests and other assessment devices. Thus, test anxiety figures prominently in the literature as one of the key villains in the ongoing drama surrounding psychological testing and is frequently cited among the factors at play in

determining a wide array of unfavorable outcomes and contingencies, including poor cognitive performance, scholastic underachievement, and psychological distress and ill health. Many individuals have the ability to do well on exams but perform poorly because of their debilitating levels of anxiety. Consequently, test anxiety may limit educational or vocational development because test scores and grades influence entrance to many educational or vocational training programs in modern society. The field of test anxiety research has prospered, in part, due to the increasing personal salience of test situations for people in modern society, making tests and their long-term consequences significant educational, social, and clinical problems for many.

Experience of test anxiety is a near-universal phenomenon across people differing in age, gender, and culture. Meta-analyses of test anxiety data from various national sites show that although mean test anxiety levels vary somewhat across cultures, test anxiety is a prevalent and relatively homogeneous cross-cultural phenomena. Furthermore, women tend to report higher levels of test anxiety than men, but the gender difference often does not translate into objective performance differences. Gender differences are attributable to differential exposure and learning experiences as well as the possibility that women are more willing to admit they are anxious.

This article first provides an overview of basic issues and conceptualizations. Then, it discusses the measurement and assessment of test anxiety, followed by a review of empirical studies of evaluative anxiety and performance. Finally, it examines interventions tailored to alleviate test anxiety and its affective, cognitive, and behavioral components.

## 1.2. Conceptions of Test Anxiety

The term test anxiety refers to the set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or loss of competence on an exam or similar evaluative situation. Test-anxious behavior is typically evoked when a person believes that his or her intellectual, motivational, and social capabilities are taxed or exceeded by demands stemming from the test situation.

Throughout its relatively brief history as a scientific construct, test anxiety has taken on a variety of different meanings as a function of both changes in the "zeitgeist" and variations in the theoretical persuasions of the individual investigators involved in this research arena. Thus, in the early days of test anxiety

research the construct was initially defined in motivational terms as drive level, goal interruption, or need to avoid failure. Test anxiety was also conceptualized as a relatively stable personality disposition that develops when parents hold exaggerated expectations and are overcritical of their children's achievement efforts. In keeping with the cognitive revolution of the late 1960s and early 1970s, test anxiety came to be viewed primarily as a cognitive-attentional phenomenon, as evidenced by a number of key chapters in Sarason's 1980 book on test anxiety. Accordingly, the highly anxious person is one who attends, in an exaggerated way, to evaluative cues, to self-generated concern about the ability to do well enough, and to feelings of physiological arousal. Recent conceptualizations have viewed test anxiety as part of a cybernetic self-control process, with anxiety reflecting the existence of a conflict between competing reference values, or as a form of self-handicapping employed to preserve one's self-merit in the face of potential failure.

The test anxiety construct was dramatically advanced by a number of important conceptual distinctions, which helped refine thinking and research in the area. One useful distinction, advanced by Charles Spielberger, is between anxiety as a relatively stable personality trait and anxiety as a more transitory state reaction to specific ego-threatening situations. Another important conceptual and methodological contribution to the test anxiety literature is the distinction made by Alpert and Haber between facilitating and debilitating anxiety. Accordingly, facilitating and debilitating anxiety are claimed to lead to task-related and task-irrelevant behaviors during test situations, respectively. Yet another conceptual contribution, advanced by Liebert and Morris, is the critical differentiation between worry and emotionality components of test anxiety. This distinction proved to be instrumental in shifting test anxiety theory and research toward a more cognitive orientation. According to this conceptualization, anxiety was viewed to be a bidimensional phenomenon, including cognitive (worry) and affective (emotionality) components. Specifically, worry was viewed primarily as cognitive concern about the consequences of failure, whereas emotionality was defined as consisting of perceptions of autonomic reactions evoked by evaluative stress. These two components are revealed to be empirically distinct, though correlated, and worry relates more strongly to test performance than does emotionality.

Theorizing by Zeidner, in his 1998 book on test anxiety, emphasized the distinction between test

anxiety as an attribute of the person and as a dynamic process. From the first perspective, dispositional test anxiety may be construed as a contextualized personality trait. Accordingly, test anxiety refers to the individual's disposition to react with extensive worry, intrusive thoughts, mental disorganization, tension, and physiological arousal when exposed to evaluative contexts or situations. The more transient state expressions of anxiety may be assessed separately from the more stable trait. From the second, process-oriented perspective, test anxiety depends on the reciprocal interaction of a number of distinct elements at play in the ongoing stressful encounter between a person and an evaluative situation, including the evaluative context, individual differences in vulnerability (trait anxiety), threat perceptions, appraisals and reappraisals, state anxiety, coping patterns, and adaptive outcomes. Events that elicit test anxiety consist of a number of distinct temporal phases, including preparation, confrontation, anticipation, and resolution. Accordingly, threat appraisals, state anxiety levels, and competence for task performance may change at different stages.

Much of the ambiguity and semantic confusion associated with the status of test anxiety as a psychological construct stems from the fact that different investigators have studied this term with quite divergent meanings. Thus, test anxiety has been used to refer to several related but logically very different constructs, including stressful evaluative stimuli and contexts, individual differences in anxiety proneness in evaluative situations (i.e., trait anxiety), and fluctuating anxiety states experienced in a test situation (i.e., state anxiety). Although the question still looms large whether test anxiety is best conceptualized as a relatively stable personality trait (individual difference variable) or an ephemeral emotional state, a widely accepted conceptualization by Spielberger proposes a hybrid term by construing test anxiety as a situation-specific personality trait. Accordingly, test anxiety refers to the individual's disposition to react with extensive worry, intrusive thoughts, mental disorganization, tension, and physiological arousal when exposed to evaluative situations.

### 1.3. Facets and Typologies

In contrast to early mechanistic views of test anxiety as a unified construct, test anxiety is currently construed as a complex multidimensional construct embodying a series of interrelated cognitive, affective, and behavioral components and reactions. The fact that anxiety is such a complex construct, encompassing worry and

self-preoccupation, physical upset, disruptive feelings, and maladaptive behaviors, makes it particularly difficult for researchers to sort out all these components. Although a number of attempts have been made to determine the dimensionality of the test anxiety construct and to tease out its key facets, little agreement has been reached among experts on the exact number of facets or components of test anxiety. However, researchers have found it particularly useful to differentiate between a cognitive facet (worry, irrelevant thinking, etc.), an affective facet (tension, bodily reaction, and perceived arousal), and a behavioral facet (deficient study skills, procrastination, avoidance behaviors, etc.). Thus, test-anxious individuals may be characterized by their thoughts, somatic reactions, feelings, and observable behaviors in evaluative situations. In any given test situation, test-anxious subjects may experience all, some, or none of these test anxiety reactions. The specific anxiety response manifested may vary, depending on the constitutional qualities and past experience of the individual, the nature of the problem to be solved, and various situational factors affecting the level of anxiety evoked.

Discussions of test anxiety in the literature are commonly guilty of a "uniformity myth," conveying the impression that test anxiety is a rather homogeneous or unidimensional category. However, under the assumption that test anxiety is a multidimensional and multidetermined phenomenon, it stands to reason that a variety of different types of test-anxious examinees may be identified. This simple fact is often overlooked when writers present theory and research relating to the "test-anxious" student—typically treated as a uniform category. Some examinees may be anxious in test situations because they have high motivation to succeed on academic tasks given the likelihood of failure, some may have poor study or test-taking skills, some may be anxious because they have low intellectual ability, some tend to be perfectionist overachievers and will be dissatisfied with anything less than a perfect score, and others are anxious because they fail to meet social expectations or fear parental punishment.

## 2. MEASUREMENT AND ASSESSMENT OF TEST ANXIETY

### 2.1. Self-Report Instruments

Self-report inventories have become the most prevalent format for assessing test anxiety because they are

considered to provide the most direct access to a person's subjective experiential states in evaluative situations; possess good psychometric properties; are relatively inexpensive to produce; and are simple to administer, score, and interpret.

Among the most widespread contemporary measures of test anxiety are Spielberger's Test Anxiety Inventory, Sarason's Reactions to Tests, Suinn's Test Anxiety Behavior Scale, and Benson *et al.*'s Revised Test Anxiety Scale.

A test anxiety scale needs to show adequate reliability and validity before it can be used in making practical decisions. Fortunately, most popular test anxiety inventories have satisfactory reliability coefficients, typically in the high .80s to low .90s. During longer intervals between assessments, personality traits such as test anxiety may change, causing lower stability coefficients. Additional factors influencing reliability are test length, test-retest interval, variability of scores, and variation within test situation.

Factor analysis, both exploratory and confirmatory, has been the method of choice in studying the structure and construct validity underlying items on test anxiety inventories. A large number of factor analytic studies have supported two to four factor structures of the construct. However, self-report scales are often plagued by a number of conventional threats to validity, including (i) response bias—that is, the tendency of a person to favor a particular response regardless of the stimulus (e.g., respond “yes” to the item “I feel calm and relaxed during important exams”); (ii) social desirability—that is, responding in ways judged socially acceptable to the self, examiner, or others expected to see the results; (iii) defensiveness—persons who are high in defensiveness may be experiencing test anxiety at some level but may be repressing conscious feelings about it; and (iv) fakeability—students may be able to fake bad (high test anxiety levels) or fake good (low test anxiety levels) on these measures.

A main problem confronting criterion validity lies in finding an agreed upon or acceptable criterion; currently, there is no infallible or perfectly objective criterion against which to validate test anxiety scores. Scores on ability tests, grade point average, observer ratings, behavior in structured evaluative situations, and the like have been employed as measures of criterion behaviors. It is noted that if anxiety is expected to be evoked in the criterion performance, then attempts to minimize anxiety in a selection instrument may in fact reduce its validity as a predictor.

## 2.2. Alternative Assessment Procedures

A variety of less frequently used assessments have been employed in assessing test anxiety, including “think-aloud” procedures (e.g., “Please list as many thoughts and feelings as you can recall having during this test”), physiological measures designed to gauge changes in somatic activity believed to accompany the phenomenological and behavioral components of test anxiety, trace measures (e.g., accretion levels of corticosteroids, adrenaline products, sugar, cholesterol, and free fatty acids), performance measures (e.g., examination scores, semester grade point averages, and latency and errors in recall of stress-relevant stimulus materials), and systematic observations of specific behaviors reflective of test anxiety (perspiration, excessive body movement, chewing on nails or pencil, hand wringing, “fidgety” trunk movements, and inappropriate laughter when subjects were engaged in exam situations).

## 2.3. Practical Uses in Applied Settings

The ultimate test for a test anxiety measure is that it shows practical utility in applied settings, and current measures leave much to be desired in this respect. Seldom do test anxiety scales inform us about the various situational and personal factors eliciting test anxiety (e.g., anxiety proneness, inadequate preparation, and overstimulation), the full range of manifestations of test anxiety (e.g., cognitive, affective, and behavioral), coping procedures and strategies, the consequences of test anxiety, or the dynamic fluctuations in test anxiety states across various phases of a stressful evaluative encounter. Future scales need to be more relevant for planning, execution, and evaluation of clinical or educational interventions through specification of the various antecedent conditions, manifestations, and consequences of test anxiety. Furthermore, current instruments are designed to measure the relative presence of test anxiety but do not inform us enough about the very high or very low test-anxious individual, which is essential for clinical decision making. Thus, future measures need to cover specific manifestations of anxiety, ranging from a total lack of concern about evaluation and minimal motivation to debilitating levels of worry, anxiety blockage, and panic. In addition, future measures need to distinguish between facilitating and debilitating arousal and cognitive processes that are realistic (e.g., worrying about an imminent difficult exam in physics, which is prompted

by a genuine threat) and those that are unrealistic (e.g., those prompted by an unlikely exam failure).

### 3. ANXIETY AND COGNITIVE PERFORMANCE

The importance of test anxiety as a key construct in understanding sources of examinee distress, impaired test performance in evaluative situations, and academic underachievement is readily apparent. This situation demands that test anxiety be better understood through systematic assessment and research and appropriately dealt with. Indeed, much of test anxiety research during the past half century has been conducted to help shed light on the negative effects of test anxiety on examinee performance and these concerns have stimulated the development of a variety of assessment methods as well as therapeutic techniques and intervention programs.

#### 3.1. Nature of Anxiety–Performance Relationship

Hundreds of studies have investigated the complex pattern of relations between anxiety and different kinds of performance. Test anxiety has been found to interfere with competence both in laboratory settings and in true-to-life testing situations in school or collegiate settings. Processing deficits that relate to test anxiety include general impairments of attention and working memory, together with more subtle performance changes such as failure to organize semantic information effectively.

Hembree's 1998 meta-analytic study, based on 562 North American studies, demonstrated that test anxiety correlated negatively, although modestly, with a wide array of conventional measures of school achievement and ability at both high school and college levels. Data collected on students from the upper elementary school level through high school show that test anxiety scores were significantly related to grades in various subjects, although the correlation was typically approximately  $-0.2$ . Cognitive measures (i.e., aptitude and achievement measures combined) correlated more strongly with the worry than emotionality component of test anxiety. Higher effects sizes were reported for low- than high-ability students and for tasks perceived as difficult than those perceived as being easy. Another meta-analysis reported by Ackerman and Heggestad in 1997 showed a mean correlation of  $-.33$  between test

anxiety and general intelligence test performance. Test anxiety was also correlated in the  $-.20$  to  $-.30$  range with other broad intellectual abilities, including fluid and crystallized intelligence, learning and memory, visual perception, and math ability.

The nature of the anxiety–competence relationship is best viewed as reciprocal in nature. Thus, high levels of test anxiety, with elevated levels of worry and cognitive interferences, absorb part of the capacity needed for attention, short-term memory, problem-solving, or other cognitive processes required for successful completion of the test. Test anxiety also produces certain aversive patterns of motivation, coping, and task strategies that interfere with learning and performance. The result is that competence and self-efficacy suffer, thus leading to further anxiety over time and generating a vicious circle of increasing anxiety and degrading competence. Future research would profit from employing nonrecursive process models in order to better capture the dynamic and cyclical nature of the anxiety–performance relationship.

#### 3.2. Mediating and Moderating Effects

##### 3.2.1. Mediating Effects

There is a large literature on test anxiety as a predictor of information processing in laboratory studies that overlaps with studies of general anxiety. The information-processing components sensitive to test anxiety relate to input (encoding and acquisition of information), central processing (e.g., memory, language processing, conceptual organization, judgment, and decision making), and output (e.g., information retrieval, response selection, and execution). Deficits related to test anxiety have been identified at various stages of processing, suggesting some general impairment in attention and/or working memory. These various performance deficits are often attributed to high levels of worry and cognitive interference.

Both cognitive interference and cognitive bias appear to be pervasive in evaluative anxiety, influencing various stages of information processing. In general, these mechanisms appear to operate much as they do in general anxiety, although evidence is lacking on cognitive bias and evaluative anxiety. A hypervigilance theory put forth by Michael Eysenck in 1992 plausibly suggests that anxiety leads to scanning of the environment for threat (generating distractibility and attentional impairment), followed by focusing of attention on sources of

threat (generating attentional bias). In addition, competence deficits may also be a consequence of poor skill acquisition. For example, deleterious effects of test anxiety may reflect not just cognitive interference but also deficits in study habits and test-taking skills.

Behavioral avoidance generated in part by performance-avoidance goals plays a key role in maintenance of evaluative anxiety and concomitant skill degradation. Test anxiety leads to procrastination, motivated by the aversiveness of the test material or fear of failure on the test. Procrastination, such as failure to complete homework assignments or study for the test, leads to failure to acquire the knowledge required. In turn, this lack of preparation leads to poor performance and anxiety in the test situation, increasing subsequent test anxiety and avoidance of study.

### 3.2.2. Moderating Effects

Studies also identify moderator variables that accentuate or reduce deficits in performance. For example, negative feedback appears to be especially detrimental to test-anxious subjects, whereas providing reassurance and social support may eliminate the deficit. However, there have been sufficient instances of nonconfirmation of predicted deficits to suggest that high anxiety does not automatically generate lower achievement outcomes.

Although anxiety is predominantly harmful to task performance (debilitating anxiety), it may sometimes have a positive effect (facilitating anxiety). One of the factors that may especially tip the scales toward debilitating effects is the presence of worry because of its tendency to produce distracting cognitive interference. The nature of the task may also play an important moderating role. Generally, test anxiety is more detrimental to attentionally demanding tasks and may even facilitate performance on easy tasks. There may also be more subtle effects related to the qualitative nature of the task.

## 4. COPING, INTERVENTIONS, AND CLINICAL CONSIDERATIONS

### 4.1. Coping

Text anxiety relates to higher emotion-focused coping (e.g., trying to control anxiety symptoms) and avoidance (e.g., trying not to think of the test) but to lower task-focused coping (e.g., focusing effort on task performance). Emotion-focused coping and avoidance both appear to predict state anxiety in evaluative situations.

Perhaps surprisingly, task-focused coping also relates to higher pretest anxiety: Exam preparations may inevitably lead to elevated anxiety. It is often difficult to categorize coping strategies as exclusively adaptive or maladaptive, but task-focused coping tends to lead to higher grades than avoidance, although the data are mixed.

### 4.2. Intervention Programs and Techniques

A bewildering array of test anxiety treatment programs have been developed and evaluated during the past three decades. Treatment fashions and orientations have swayed sharply from the clinical to the behavioral, and recently to the cognitive perspective—essentially mirroring the evolution of the behavior therapies. There is no simple organizing principle with which to categorize the plethora of therapeutic techniques and approaches that have proliferated during the past few decades. Current attempts to reduce debilitating levels of test anxiety and enhance test performance have typically focused on treatments directed toward the emotional (affective), cognitive (worry), or behavioral skills (study skills) facets of test anxiety. Thus, treatment programs typically include emotion-focused treatments, designed largely to alleviate negative emotional affect experienced by test anxious persons; cognitive-focused treatments, designed to help the test-anxious client cope with worry and task-irrelevant thinking; and study-skills training, designed to improve study and test-taking skills and enhance test performance.

A brief summary of key emotion-focused, cognitive-focused, and skill-focused treatment techniques and methods, and their reported effectiveness, is presented in [Table I](#).

#### 4.2.1. Emotion-Focused Interventions

The emotionally oriented therapies aim primarily at reducing the arousal and heightened emotional reactions of test-anxious persons when faced with stressful evaluative situations. Based on the assumption that anxiety comprises a physiological component, attempts to alleviate anxiety symptoms should prove successful, in part, if they focus on reducing levels of arousal or on altering ways in which people appraise their arousal in evaluative situations.

The basic strategy in these treatments is directed to teach the client certain skills (mainly relaxational) so that when confronted by stress-inducing evaluative

**TABLE I**  
**Focal Test Anxiety Intervention Techniques**

<i>Treatment</i>	<i>Description</i>	<i>Effectiveness</i>
<b>Emotion-focused intervention</b>		
<i>Biofeedback</i>	<p>Use of instrumentation (e.g., a physiograph) to provide a person with immediate and continuous information about one or more physiological processes (e.g., skin conductance, temperature, heart rate, blood volume pulse, respiration, and electromyograph).</p> <p>Biofeedback teaches high test-anxious persons to monitor and modify the physiological processes associated with their emotional reactions.</p>	<p>A large body of literature supports the notion of <i>increased</i> physiological control when using physiological feedback and self-regulation. However, biofeedback alone is not effective in reducing test anxiety (nor does the addition of biofeedback training improve the efficacy of other forms of treatment). Given the potential cost and inconvenience of using biofeedback training, it may not be the treatment of choice for test anxiety intervention.</p>
<i>Relaxation training</i>	<p>Recommended on the premise that maintaining a relaxed state during testing, via deep breathing and muscle relaxation exercises, would counteract a person's aroused state.</p> <p>Presumably, if a person knows when and how to apply relaxation, it will be applied directly as a counterresponse to anxiety.</p>	<p>Meta-analytic research tends to support the effectiveness of relaxation therapy. However, the effects on performance tend to be negligible.</p>
<i>Systematic desensitization</i>	<p>Test anxiety is viewed as a classically conditioned emotional reaction resulting from a person's aversive experiences in evaluative situations. Systematic desensitization proposes that anxiety reactions to test situations may also be unlearned through specific counterconditioning procedures. The test-anxious client is typically trained in a deep muscle relaxation procedure and, while relaxed, instructed to visualize an ordered series of increasingly stressful test-related scenes (an "anxiety hierarchy"). The client imaginably proceeds up the hierarchy until he or she is able to visualize the most stressful scenes on the list without experiencing anxiety. Through repeated pairings of imaginal representations of threatening evaluative situations with deep relaxation, the bond between the threatening evaluative scenes and anxiety is expected to be weakened.</p>	<p>Meta-analytic data lend support to the effectiveness of systematic desensitization in reducing test anxiety in school-children and college students. It is shown to be as effective, if not more effective, in reducing test anxiety than a variety of other treatments, including relaxation training, hypnosis, and study skills training. However, systematic desensitization fares less well when cognitive performance (e.g., academic achievement) is the criterion or outcome being assessed.</p>
<i>Anxiety management</i>	<p>Teaches highly test anxious subjects to recognize their test-related arousal responses as they are building and then to use them as cues for initiating the coping response of relaxation in evaluation situations.</p>	<p>A body of research supports the effectiveness of this technique in reducing test anxiety. Thus, anxiety management training appears to be as robust and effective, if not more so, than related interventions. Reductions in debilitating test anxiety were maintained for follow-up periods ranging from several weeks to several months.</p>

*Continues*

Continued

Treatment	Description	Effectiveness
<i>Modeling</i>	Involves the live or symbolic (e.g., through videotape) demonstration of desired coping behaviors in a stressful evaluative situation such that they can be subsequently imitated by the test-anxious person. It is assumed that exposure to models displaying adaptive behavior may play a positive role in facilitating performance. Clients are instructed to vividly imagine the stressful evaluative scene and to focus on the anxiety and associated response-produced cues (e.g., racing heart, neck and shoulder tensing, dryness of the mouth, and catastrophic thoughts). The clients are then trained to use these cues to prompt adaptive coping skills to actively relax away tension and reduce anxiety before it mounts too severely.	A body of research lends support to the effectiveness of modeling in treating test anxiety. In particular, exposure to models who are task-oriented and provide attention-directing cognitive structuring clues is beneficial to the performance of test-anxious persons. Of additional benefit is evidence in the behavior of the model that he or she is successfully coping with the worry and tension associated with test anxiety.
Cognitive-focused interventions		
<i>Cognitive-Attentional training</i>	Cognitive-attentional training provides specific training in the redirection of attention to task-focused thinking and emphasizes the inhibition of task-irrelevant thinking and nonproductive worry. The cognitive-attentional approach relates performance decrements to the diversion of attention to self-focused thinking, coupled with the cognitive overload caused by the worry component of anxiety. By redirecting attention to the task and reducing worry and task-irrelevant thinking, cognitive resources are freed, and when redirected to the task, performance is improved. Attentional training programs traditionally provide clients with instructions to attend fully to the task and to inhibit self-relevant thinking while working on a variety of academic tasks.	The beneficial effects of attentional instruction on the anxiety and cognitive performance of high test-anxious students is supported by some empirical research. Task instructions that provide examinees with information about appropriate problem-solving strategies, and away from self-preoccupied worry, may be particularly helpful to the test-anxious individual's cognitive functioning.
<i>Cognitive restructuring</i>	The rationale is that test-anxious persons will be able to master their anxiety by learning to control task-irrelevant cognitions that generate their anxiety and direct attention from their task-directed performance. The two most prominent cognitive therapeutic methods in test anxiety intervention are rational emotive therapy and systematic rational restructuring. Both forms of treatment are based on the premise that anxiety or emotional disturbance is a result of illogical or "irrational" thinking. Two key irrational beliefs that maintain test anxiety is that one <i>must</i> succeed on major exams at all cost and that success on these exams is equivalent to self-worth.	Research indicates that although cognitive restructuring reduces anxiety, there is no concomitant improvement in performance. A number of studies provide evidence showing that these technique may be effective in reducing anxiety. However, concomitant improvements in cognitive performance are observed with far less consistency.

Treatment	Description	Effectiveness
Cognitive-behavioral modification	<p>Test-anxious individuals are taught how to recognize, vigorously challenge, question, and dispute their irrational beliefs and replace their maladaptive internal dialogue with more rational structures and beliefs. Presumably, by modifying irrational beliefs and schemas, negative emotional reactions will be reduced and performance improved. Systematic rational restructuring aims at helping test-anxious clients to discover the worrisome task-irrelevant thoughts they entertain during tests, to eclipse such thoughts, and to substitute positive self-statements that redirect their attention to the task at hand.</p> <p>A multifaceted program merging both cognitively focused and emotionally focused techniques (as well as skill training in many cases), thus offering the test-anxious client the best of many worlds, so to speak. This multimodal treatment attempts to deal with the multiple manifestations of test anxiety, including negative motivational or affective tendencies, irrational thought patterns, and skills deficits, and emphasizes the application and transferring of acquired coping skills to <i>in vivo</i> test situations. Given its dual emphasis on modifying both emotional processes and irrational thoughts and cognitions, this results in a powerful approach that merges emotionally oriented and cognitively oriented techniques to alleviate clients' test anxiety and enhance their test performance. This procedure is based on the premise that reducing a person's level of test anxiety involves both anxiety reduction training and detailed cognitive restructuring of certain faulty beliefs or misconceptions concerning evaluative situations. It is also assumed that some students may require restructuring of their study habits if rewarding and successful academic work is to be ensured.</p>	<p>"Multimodal" treatment packages, such as cognitive-behavior modification, are most likely to be effective by their support for the inclusion of multiple domains related to test anxiety. These procedures are relatively effective in reducing self-reported levels of debilitating test anxiety and are equally effective, more or less, in reducing both cognitive and affective components of test anxiety test performance, on average, by about half a standard deviation in school-aged samples and elevate grade point average by approximately three-fourths of a standard deviation.</p>
Skill-focused interventions	<p><i>Study skills training</i></p> <p>Study skills training differs from the other cognitive therapies in that it does not directly focus on modifying the cognitive component of anxiety but rather centers on improving students' study and test-taking skills. Presumably, improvement of these skills should have a direct impact on performance through improved mastery of the test material and also indirectly impact performance through the reduction of worries surrounding inadequate exam preparation.</p>	<p>These programs have been successful mainly in enhancing the test performance of high test-anxious subjects with poor study skills. It is apparent that when test-anxious subjects suffer from serious study skill or test-taking deficits, alternative forms of treatment would not be expected to lead to performance gains, inasmuch as the behavioral deficit still exists. At the same time, skill training alone may lead a person to performance gains, but it may still leave the person distressed.</p>

situations in the future he or she will be able to handle them adequately. The therapies also provide opportunities for application of training either within the therapy setting or in real-life situations.

These emotion-focused procedures typically include a number of common components, such as theoretical explanations of test anxiety as a conditioned response and the “deconditioning” rationale for treatment; instructions in specific methods for reducing anxiety, such as relaxation and guided imagery; guided practice in therapeutic methods; and practice (homework and *in vivo* practice). By and large, these emotion-focused treatments rely on key behavioral learning principles (counterconditioning, reciprocal inhibition, extinction, observational and coping skill learning, etc.) and also draw from an arsenal of behavioral techniques, such as deep muscle relaxation, guided imagery, and graduated hierarchies. For example, relaxation and guided imagery is not unique to a particular test anxiety behavioral intervention method, but it is employed in several methods, including relaxation as self-control, systematic desensitization, and anxiety management training.

Procedures designed to reduce emotionality, although clearly useful in modifying subjectively experienced anxiety, by themselves appear to have little effect on cognitive performance. Overall, emotion-focused treatments appear to be relatively ineffective in reducing test anxiety unless these treatments contain cognitive elements. It may therefore be necessary to combine such approaches with therapy modes focusing specifically on cognitive change in order to reliably elicit improvement in cognitive performance.

#### **4.2.2. Cognitive-Focused Interventions**

Recently, there has been a proliferation of cognitively oriented intervention programs that emphasize the mediating role of cognitive processes in sustaining or eliminating test anxiety. In part, the documented failure of emotionally oriented behavioral therapies to markedly improve the academic performance of test-anxious students, coupled with the inconsistent relation reported between emotional arousal and test performance, has led to a greater emphasis on cognitive factors in test anxiety intervention. Indeed, reviews of the literature conclude that cognitively based treatment strategies are more powerful than direct behavioral therapies in effecting test anxiety and performance changes.

Cognitive therapy is a generic term that refers to a wide array of therapeutic approaches directed toward modifying the worry and irrational thought patterns of

test-anxious clients. Broadly speaking, cognitively oriented approaches to test anxiety intervention are quite similar in assuming that cognitive processes are determining factors in test anxiety, although they differ in terms of actual intervention procedures. A fundamental assumption shared by contemporary cognitive models of test anxiety is that cognitive processes mediate the person’s emotional and behavioral responses to stressful evaluative situations. It follows that in order to modify the negative emotional reactions of test-anxious clients to evaluative situations, therapy needs to be directed at reshaping the faulty premises, assumptions, and negative attitudes underlying maladaptive cognitions of test-anxious subjects.

Given their emphasis on modifying emotional processes, irrational thoughts and cognitions, and behavioral deficits, this results in a powerful approach that merges emotionally oriented, cognitively oriented, and behaviorally oriented techniques to alleviate clients’ test anxiety and enhance their test performance.

#### **4.2.3. Skill-Focused Intervention**

Recent thinking and research suggest that high test-anxious students with poor study and test-taking skills would probably benefit most from study skills training intervention intended to improve their study and test-taking habits and skills. This intervention is based on a cognitive-deficit or deficiency model. Whereas study skills training is directed toward improving a variety of cognitive activities that affect the organization, processing, and retrieval of information (e.g., study habits and test-taking skills), training in study skills does not directly address the specific cognitive concomitants of test anxiety. Instead, it is designed to augment other cognitive interventions.

The distinction between the various treatment orientations is quite fuzzy, and these approaches are becoming increasingly difficult to distinguish. Although there may be highly specific interventions, which have an affective (e.g., relaxation therapy) or cognitive (e.g., rational emotive therapy) orientation, most methods are normally embedded in a multidimensional context. A combination of procedures (whether combined in a truly integrative manner or in the stance of technical eclecticism) seems to best represent the true nature of the test anxiety intervention process.

### **4.3. Clinical Considerations**

This section presents a number of important considerations that should be kept in mind by both researchers

and practitioners when developing, implementing, or evaluating test anxiety intervention programs.

#### **4.3.1. Performing Careful Diagnosis of Client's Problem**

In order to tailor treatment programs to meet the specific needs and problems of the client, a logical first step is a careful diagnostic assessment and analysis of the nature of the test-anxious person's affective and cognitive problem(s). For some test-anxious subjects, provision of skills training may be the treatment of choice, whereas for others it would involve building up of self-confidence in a particular content area (e.g., math) or teaching relaxation skills. Information about the following aspects of the client's problem might be particularly useful: nature of the problem as experienced and defined by the test-anxious client, perceived severity and generality of the problem, duration and extent of test anxiety, perceived origins of test anxiety, situation-specific factors that intensify or alleviate anxiety reactions, specific consequences of test anxiety for the client, and suggested changes that the client views as potentially helpful. A careful diagnostic assessment may suggest factors other than test anxiety that underlie one's heightened emotional reactions in test situations.

#### **4.3.2. Meeting Preconditions for Therapeutic Effectiveness**

In order for a test anxiety intervention program to work, a number of preconditions need to be met. First, test-anxious individuals should possess certain relevant skills in their behavioral repertoire (e.g., problem-solving, relaxation, study/test-taking skills) to apply in appropriate evaluative circumstances. Second, test-anxious clients must be sufficiently motivated to deal directly with evaluative situations and have the wherewithal and self-efficacy to efficiently implement the coping skills they have at their disposal. Third, test-anxious persons must be provided with an adequate amount of practice and experience in applying various coping skills in true-to-life evaluative situations in order to ensure transfer of therapy from the treatment environment to the real world.

#### **4.3.3. Adjusting Treatment to the Needs of Particular Types of Test-Anxious Individuals**

Interventions and therapeutic techniques would be most effective if they could be adjusted to suit the

needs of different types of test-anxious persons. Because there are different types of high test-anxious individuals, as suggested previously, each characterized by different problems and concerns (e.g., failure in meeting personal or social expectations, low feelings of self-efficacy and failure acceptance, and poor study skills), no single treatment program would be expected to be equally effective for all clients. Thus, for some highly perfectionist test-anxious individuals therapy may focus on lowering socially prescribed performance expectations, whereas for other "failure accepting" students therapy may consist of raising performance expectancies and enhancing perceived self-efficacy. Students high in test anxiety with sound study or test-taking skills should profit from treatment focusing on test anxiety reduction. In contrast, students with defective study or test-taking skills and high test anxiety would profit from a combined intervention program to improve their skills and decrease test anxiety.

#### **4.3.4. Basing Treatment on the Broader Diagnostic Picture and Specific Goals of Therapy**

The choice of which therapy to use will be influenced not only by the diagnosis of the specific nature of the client's problem and type of test anxiety but also by the broader diagnostic picture, the immediate and long-term goals of treatment, and the therapeutic orientation adopted. For example, although relaxation may not increase the performance of the test-anxious student with study skill deficits, it may be prescribed by the therapist in order to help the student reach the immediate goal of achieving control over test anxiety as a first step toward academic problem solving. Thus, after the anxiety that interferes with learning new study skills is removed, the next step would be training the student in efficient study skills. Furthermore, there are different ways that a therapist may view his or her test-anxious clients' problems (distorted thinking styles, poor problem-solving skills, etc.), and each of these views may give rise to different treatment procedures.

#### **4.3.5. Consideration of Individual Differences**

Before implementing a particular treatment, one needs to determine to what extent the treatment may interact with particular client characteristics. For example, some interventions may reduce anxiety or successfully increase the performance of high test-anxious

individuals but have a negative effect on the anxiety performance of others who are low in test anxiety. Also, whereas some people might considerably benefit from relaxation training, experiencing a substantial decrease in anxiety, some have difficulty in acquiring relaxation skills and benefit little from relaxation training. Some clients may even experience relaxation-induced anxiety during relaxation training.

#### **4.3.6. Addressing Multiple Modalities and Loci of Therapeutic Impact**

The various components of test anxiety must be dealt with if the anxiety experienced in test situations is to be reduced and improved grades are to be obtained as a result of treatment. It is important that interventions be sufficiently complex to deal with the major facets (cognitive, affective, and behavioral) of the test anxiety experience. Indeed, a treatment would be expected to be most effective if it impacts upon the entire range of components and chain of events leading to anxious manifestations in evaluative situations (arousal, worry, meaning system, internal dialogue, behavioral acts, etc.) rather than focusing on only one aspect of the process.

#### **4.3.7. Interaction among Components of Test Anxiety**

A basic consideration is that test anxiety is more than a combination of physiological arousal, negative self-preoccupation, a deficit in stress-related coping skills, and poor study habits. It is the complex interaction among these diverse components that seems to define test anxiety. Because the cognitive, affective, and behavioral components of test anxiety all contribute to the problem of test anxiety and its treatment, it is predicted that an induced change in one system would generally be followed by a change in the other. Thus, therapeutic approaches, which often emphasize cognition, extend to the emotional life and vice versa. For example, it is likely that emotion-focused training (e.g., progressive relaxation) may make the client less anxious and result in a decrease in anxiety-focused, task-irrelevant ideation. Similarly, some forms of cognitive therapy may provide test-anxious subjects with an increased sense of perceived control, which might spill over into the emotional domain and result in lower emotional arousal in a test situation.

It is readily apparent that test anxiety intervention should be based on a careful theoretical analysis of the nature of test anxiety and its key components and manifestations. Traditionally; however, test anxiety treatment studies have mainly evolved from interest in specific behavioral treatment techniques rather than from an analysis of the nature and effects of test anxiety. Indeed, most investigators who have applied behavioral methodology to the reduction of test anxiety have generally given little attention to relating the treatment process to important theoretical conceptions. The current diversity of test anxiety treatments, although supplying the clinician with a rich variety of treatment options to choose from in rendering services, reflects a state of uncertainty marked by the lack of consensus regarding the most effective method for treating test anxiety.

### **See Also the Following Articles**

Academic Interventions ■ Achievement Tests ■ Assessment and Evaluation, Overview ■ Clinical Assessment ■ Coping ■ Educational and Child Assessment ■ Intelligence Assessment ■ Reading Interventions ■ Stress

### **Further Reading**

- Ackerman, P. L., & Egested, E. D. (1997). Intelligence, personality and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.
- Covington, M. V. (1992). *Making the grade*. New York: Cambridge University Press.
- Eysenck, M. W. (1992). *Anxiety: The cognitive perspective*. Hove, UK: Erlbaum.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*, 7–77.
- Phillips, B. N. (1978). *School stress and anxiety: Theory, research and intervention*. New York: Human Sciences Press.
- Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R., & Ruebush, B. K. (1960). *Anxiety in elementary school children*. New York: Wiley.
- Sieber, J. E., O'Neil, H. F., Jr., & Tobias, S. (1977). *Anxiety, learning and instruction*. Hillsdale, NJ: Erlbaum.
- Spielberger, C. D., & Vagg, P. R. (Eds.). (1995). *Test anxiety: Theory, assessment, and treatment*. Washington, DC: Taylor & Francis.
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum.
- Zeidner, M., & Matthews, G. (In press). Evaluative anxiety: Current theory and research. In A. Elliot, & C. Dweck (Eds.), *Handbook of competence and motivation*. New York: Guilford.



# Traffic Safety Assessment

Kåre Rumar

Sturefors, Sweden

1. Introduction
  2. Road Safety Problems
  3. The Role of Psychology in Road Safety Work
  4. Three Ways to Reduce Human Errors in Road Transport
  5. Conclusion
- Further Reading

## GLOSSARY

**accident** An unavoidable event resulting from circumstances presumed to be beyond control.

**active safety** An approach to traffic risk management that aims at preventing crashes via improved road and vehicle design, driver education and training, behavioral rules (e.g., speed limits), control and enforcement, etc.

**crash** A collision between two objects in traffic. By understanding that human control, human decision making, and human decision-making error contribute to circumstances surrounding crashes, we allow for direct involvement (research and education) in the reduction of traffic fatalities and injuries.

**passive safety** An approach to traffic risk management that aims to minimize the harmful results of crashes via seat belts, air bags, energy-absorbing front ends, crash barriers, etc.

This article describes the scope of the traffic safety problem throughout the world. More than 90% of all transport fatalities in the world occur on roads. Approximately 1 million people are killed on the roads annually, and a majority (85%) of these deaths occur in

developing countries. In-depth analyses of crashes show that human error is a major contributing factor in most crashes. Therefore, psychology should be able to contribute to effective countermeasures. The existing road safety problems are analyzed in different ways. The role of psychology is described in association with a general cube model, the role of the driver, and the severity of the road safety problem. Finally, three possible ways to reduce human error in road traffic based on psychological knowledge are described.

## 1. INTRODUCTION

Worldwide, approximately 1 million people are killed in road traffic each year, and the number increases every year. In comparison with road traffic, the other transport modes (air, sea, and rail) are relatively safe. More than 90% of all transport fatalities occur on the roads. Road safety statistics are very unreliable, except those for fatalities. Therefore, this article focuses on road traffic crashes and on fatalities. A majority of these fatalities (85%) occur in developing countries. In the developed world, the largest fatality group is occupants of motor vehicles. In the world as a whole, and especially in developing countries, pedestrians comprise the largest fatality group. Contrary to most other causes of death, road traffic fatalities affect mainly young persons. Therefore, road traffic fatalities result in the highest number of lost years of all causes of death in the developed world. The World Health

Organization estimates that road crashes are the third leading health threat considering lost years and lost working time due to injuries. Thus, road crashes are a major public health problem throughout the world.

In most countries, large crashes that occur in the air, on the sea, or on rail are studied in-depth to determine the causes and to prevent similar crashes from occurring again. These studies have been quite successful and have resulted in the prevention of many crashes and fatalities. Efforts to study the causes of all fatalities on the roads by means of in-depth accident investigations have not been as successful. This process is, for many reasons, much more difficult. For example, it is much more difficult than it is with air, sea, or rail crashes to establish the important events that occurred immediately before the crash. However, the results of investigations of road crashes show convincingly that human error is a major contributing factor in more than two-thirds of the crashes (Fig. 1). Considering that human error plays such a major role in road crashes, it is evident that psychology is a discipline that may have much to contribute to road safety work.

In this article, the term crash is used instead of the traditional term accident because accident implies that the event is totally random and cannot be avoided. Consider the statements, "It was a complete accident" or "It happened by accident." By labeling road traffic crashes as accidents, it is psychologically more difficult

to convince people to contribute to the avoidance of crashes.

## 2. ROAD SAFETY PROBLEMS

There are many ways to describe road safety problems. Here, two examples are presented.

### 2.1. Exposure–Crash Risk–Injury Consequence

A fruitful way to describe road safety problems is to use what is sometimes called the Swedish ice cube (Fig. 2). The number of people killed in an area depends mainly on three variables:

- Exposure in traffic (number of kilometers, hours, trips, etc.)
- Probability of a crash given a certain exposure (crash risk or rate)
- Probability of being killed given the crash (crash consequence)

These three variables multiply in a complex way to produce the total number of people killed. By multiplying the variables, a volume (cube) is obtained. The size of this volume describes the severity of the road safety problem (how many people are killed). Consequently, when one wishes to reduce this volume, the three variables are employed. By reducing one, two, or three of these variables, the number of fatalities will be reduced.

### 2.2. Primary, Secondary, and Tertiary Problems

Another way to describe the road safety problem is to use the various specific road safety problems as starting points. The most well-known problems are alcohol, speeding, young drivers, low usage of protective devices (e.g., seat belts and helmets), tailgating, darkness, low friction, etc. If one can find the appropriate countermeasure for each one of these problems, one may think that the overall road safety problem has been solved. Unfortunately, this is not the case. There are solutions for many of these primary problems. However, for many measures, the effect is much less than expected. Another problem is that the known countermeasures are not implemented. These primary problems (alcohol, speed, etc.) are therefore only

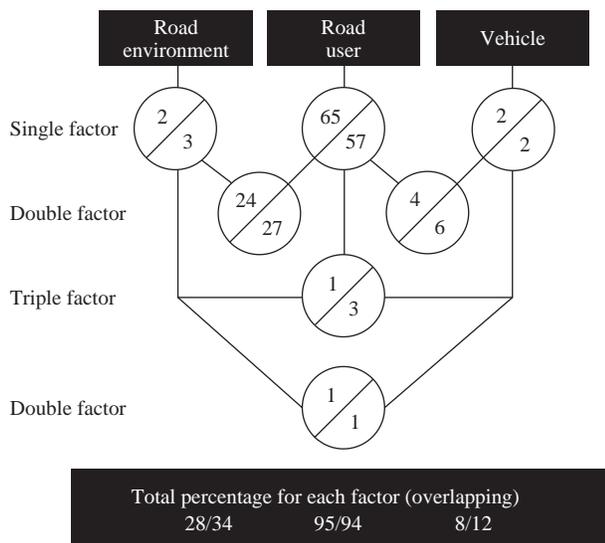


FIGURE 1 Percentages of the main factors contributing to road crashes according to the two largest crash studies (United States and the United Kingdom).

components of the overall problem. Thus, one has to look elsewhere. One has to search for the obstacles that reduce the effect of any countermeasure or that prevent known countermeasures from being implemented. Consequently, road safety problems can be described on three levels as follows:

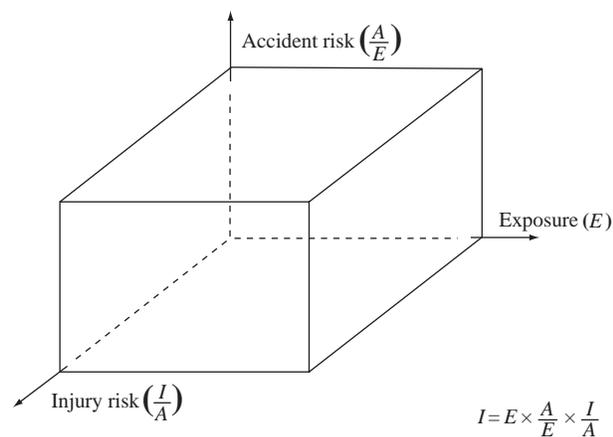
- Primary problems, which require countermeasures (alcohol, speeding, young drivers, etc.)
- Secondary problems, which reduce the effect of any countermeasure (inefficient enforcement, inadequate driver training and licensing, inadequate information, etc.)
- Tertiary problems, which prevent countermeasures from being implemented (low value of road safety measures, no defined goals, no quantitative targets, etc.)

### 3. THE ROLE OF PSYCHOLOGY IN ROAD SAFETY WORK

In this section, the two methods of defining road safety problems are used to analyze the role that psychology may play in the road safety process from problem analysis via research to implementation and follow-up.

#### 3.1. The Cube Model

The ice cube model (Fig. 2) clearly shows that traffic psychology has much to contribute to all three variable of the model. Along the exposure axis, the task could



**FIGURE 2** The magnitude of safety problems (number of human injuries and fatalities;  $I$ ) as a function of the product of three variables: exposure ( $E$ ), crash risk ( $A/E$ ), and injury consequence ( $I/A$ ).

be, for instance, to determine what makes people choose one transport mode over another. It is known that public transport is at least 10 times safer than private vehicles. What is the relative influence of price, time, comfort, freedom, company, etc. on modal choice? Results indicate that price and safety level are relatively unimportant, whereas freedom and comfort are very important.

Regarding the crash risk axis, psychology has always been one of the main contributors to understanding. This axis could also be called crash prevention or active safety, both of which could be improved by reducing the number of human errors. Traditional approaches to crash prevention are to improve roads, vehicles, and the driver or other road users.

In traditional road safety work, three E's are often mentioned, referring to engineering, education, and enforcement. The first E has already been mentioned (roads and vehicles). The other two E's are of a more psychological nature. Education may include general information, directed campaigns, education in schools, or special courses—all for the purpose of training various types of road users, such as pedestrians, bicyclists, motorcyclists, and drivers of passenger cars or heavy vehicles. Also, enforcement has many forms, such as hidden or open, intense or rare (probability), reliable or random, immediate or delayed, and manual or automatic.

Education and enforcement methods have been used since the introduction of automobiles. However, information, campaigns, education, and training have not provided the expected positive road safety results. Enforcement, on the other hand, has proved very effective. The problem is that it is very costly if carried out manually. This is why modern technology (e.g., automatic cameras and electronic identification of cars) is being used more frequently.

In recent decades, increasingly more resources have been used in human factor efforts. This means that the technical environment (roads, vehicles, and traffic legislation and control) should be adapted to human characteristics. This approach is discussed later.

Work along the third model axis tries to reduce crash consequences. This approach is often called passive safety. This approach has been very successful in recent decades. Seat belts, air bags, energy-absorbing front ends, and crash barriers are good examples of this approach. Other efforts include improved rescue service and treatment of injured persons. Along this axis, psychology has not contributed as much. However, crash protective devices (seat belts and helmets) have no effect if they are not used. The following are typical

questions for psychologists: Why do different groups of road users not use available crash protective devices more often than they do? What should be done to increase the usage rate of crash protective devices?

### 3.1.1. Specific Psychological Problems

Work along the second model axis (active safety) has not been as effective as originally hoped. One of the reasons is that the reactions of drivers under changing conditions are difficult to predict. For instance,

- How will drivers use better visibility, better friction, and improved brakes? Will they drive more safely, increase their speed, or both?
- What is the driver really doing when he or she is driving? What should a useful driver model look like?

The first problem deals with what has often been called the risk compensation hypothesis. This hypothesis states that drivers tend to stay at a constant risk level irrespective of various changes in circumstances. In order to change the risk level, the basic values of the driver have to be influenced. Experience shows that the more feedback on safety improvements the driver has, the more he or she tends to speed, and that the less feedback the driver has, the more he or she tends to drive safely. Street lighting and studded tires on ice are good examples of the first and daytime running lights are a good example of the second situation.

The second problem (designing a useful driver model) has occupied transportation psychologists for at least 70 years. It is fair to say that there is no generally useful driver model. However, there are many driver models that work quite well for limited tasks (e.g., tailgating).

### 3.2. Three Problem Levels

The second way to describe the road safety problem is to divide it into three levels (primary, secondary, and tertiary). Traffic psychology is one of the major disciplines working to find effective countermeasures to the primary problems (alcohol, speeding, young drivers, etc.). The success rate has not always been very good. Two of the major reasons for the reduced levels of effect are that enforcement levels have been inadequate and the awareness of the seriousness of the road safety problem has been too low. Driver motivation to follow recommendations and legislation has therefore been low.

Regarding the second- and third-level problems, traffic psychology has been practically the only contributor

to solutions. By refining education and training and by improving enforcement levels and strategies, traffic psychology has contributed to increasing the efficiency of most countermeasures. In both cases, intelligent transportation systems have proven to be a potentially very effective tool.

The third-level problems are of a very psychological character. The idea is mainly to increase the motivation for increased road safety. Politicians should be motivated to implement a number of road safety measures, citizens should be motivated to ask for the implementation of road safety measures, and drivers should be motivated to conform to the safety recommendations and legislation. Sweden was the first country to try to work systematically on this problem level. The Swedish "Vision Zero" has proven very effective in reducing the motivational problems, especially on the political level. Vision Zero states that nobody following the rules should be killed or impaired in road traffic. Responsibility for the design of the system lies with the road authorities, who own the system.

Third-level problems also include the way road safety work is monitored and followed up. Without feedback, mistakes may be, and have been, repeated. General information campaigns are one of the most popular road safety measures among politicians and other decision makers. The few follow-up studies that have been carried out show that the road safety effects of such general campaigns are, at best, very limited.

By establishing quantitative targets for the various sectors of road safety work, results can be followed up easily by a kind of result management. The targets must not be set only in terms of crashes and fatalities because these are unreliable indicators of road safety in the short term. In order to obtain a quick follow-up system, performance indicators are used in modern road safety work. Such performance indicators are often behavioral by nature, such as the speed distribution of cars, the time gaps between cars, and the proportion of drivers using seat belts. Here, psychology can contribute substantially.

## 4. THREE WAYS TO REDUCE HUMAN ERRORS IN ROAD TRANSPORT

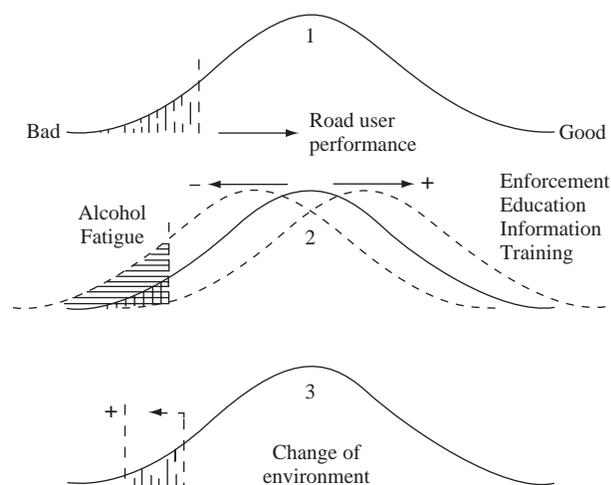
Assume that in a large majority of road crashes, human error is a major contributing factor (Fig. 1). One of the priority targets in road safety work should be to reduce human errors. To do so, the following three

approaches must be considered, to which psychology can make potential major contributions:

- Improve selection of drivers
- Modify (improve) driver behavior
- Adapt the traffic conditions and driver tasks to human characteristics

The first approach (Fig. 3) is to select drivers in such a way that those who do not perform well enough and who often commit errors will not be allowed on the roads. Currently, drivers are selected on the basis of medical, knowledge, and skill tests. These tests and their passing criteria are different in different countries. It is difficult to make these tests more stringent and tough because driving a car is considered almost a human right in most countries. The situation for pilots, for instance, is very different. Selection procedures and criteria are to a very large extent of a psychological nature (e.g., test validity and reliability).

The second approach (Fig. 3) is to try to improve the knowledge, performance, attitudes, and behavior of drivers. Typical methods include training, education, information, and enforcement. In order to carry out this improvement, considerable knowledge concerning learning psychology is required. Also, it is important to



**FIGURE 3** Three ways to reduce human error: (1) restrict road users who make many errors; (2) improve the performance of road users through information, education, training, legislation, and enforcement while ensuring that their performance is not impaired by negative factors, such as alcohol, drugs, and fatigue; and (3) adapt the environment to human characteristics so that it becomes easier to walk, bicycle, and drive.

avoid the impairment of driver behavior due to fatigue, alcohol, and drugs.

The ancient Greeks stated that “Man is the measure of everything.” In other words, if human beings commit errors in road traffic, there is something wrong with the road traffic system. The third approach (Fig. 3) means that the traffic system should be adapted to human characteristics and limitations, mainly concerning human perceptual and cognitive characteristics (the ability to perceive and interpret the traffic scene as a road user). The traffic systems—the road, road environment, road lighting, traffic control devices, vehicle dynamic characteristics (stability, suspension, braking, steering, etc.), and vehicle exterior and interior systems and details (lighting, rearview mirrors, controls, instruments etc.)—should be designed in such a way that they are easy to decode, interpret, and understand. In order to carry out such an adaptation, qualified psychological knowledge and research are required.

## 5. CONCLUSION

The previous discussion shows how both ways of describing the road safety problem lead to counteractions, which to a large extent are comparable. Therefore, it is not crucial which model is used. However it is crucial that the involved actors can understand and share the model. A common road safety model simplifies cooperation between all road safety actors. This article also notes the major role that human error plays in road crashes. This is a strong indication of the importance of traffic psychology in past and present road safety work, and it indicates that traffic psychology will be even more important in future road safety work.

### See Also the Following Articles

Accidents in Transportation ■ Aviation ■ Driving Safety ■ Models for Transportation ■ Transportation Systems, Overview ■ Travel Behavior and the Environment

### Further Reading

- American Automobile Manufacturers Association. (1998). *World motor vehicle data*. Washington, DC: Author.
- Barjonet, P.-E. (Ed.). (2001). *Traffic psychology today*. Boston: Kluwer.
- Brown, I., & Groeger, J. (Eds.). (1990). Errors in the operation of transport systems. *Ergonomics*, 33(10/11).

- Evans, L. (1991). *Traffic safety and the driver*. New York: Van Nostrand Reinhold.
- Evans, L., & Schwing, R. C. (Eds.). (1985). *Human behavior and traffic safety*. New York: Plenum.
- International Road Federation. (2000). *World road statistics*. Geneva: Author.
- Näätänen, R., & Summala, H. (1976). *Road user behaviour and traffic accidents*. Amsterdam: North Holland.
- Rothengatter, J. A., & de Bruin, R. A. (Eds.). (1987). *Road users and traffic safety*. Maastricht, The Netherlands: Van Gorcum.
- Rumar, K. (1999). *The first European transport safety lecture*. Brussels: European Transport Safety Council.
- United Nations. (2000). *Traffic safety: Problems and counter-measures*. Geneva: United Nations.
- World Bank. (2001). *Why are road crashes a problem?* Geneva: Global Road Safety Partnership.
- World Health Organization. (1999). *Injury—A leading cause of the global burden of disease*. Geneva: World Health Organization.



# Training, Cross-Cultural

David J. Dalsky

University of Mississippi, Mississippi, USA

Dan Landis

University of Hawaii, Hilo, Hawaii, USA

1. The Need for Cross-Cultural Training
  2. People Involved in Training Programs
  3. Structure of Training Programs
  4. Program Evaluation and Dimensions of Success
  5. Conclusion
- Further Reading

## GLOSSARY

**cultural assimilator** A series of critical incidents involving cultural misunderstandings that are typically used in cross-cultural training programs.

**didactic technique** A training method that includes lecture or supplying information through other means in which trainees take a passive role in learning.

**diversity training** A type of cross-cultural training whose goal is to reduce prejudice and discrimination in settings within a country.

**experiential technique** A training technique that involves a cultural simulation, role play, or related activity in which participants play an active role.

**host country** The country in which the trainee is preparing to live for a period of time.

**intercultural sensitivity** The degree of a sojourner's ability to recognize cultural differences in norms and values.

**predeparture training** Sessions of a training program that occur before the sojourner leaves for his or her host country and/or immediately upon arrival to the host country.

**reentry training** Sessions of a training program that take place just prior to the departure of the sojourner from his or her host country and/or upon his or her return home.

**sojourner** One who takes a temporary residence abroad.

Cross-cultural training involves preparing sojourners to be successful in their host country. Trainers commonly have educational backgrounds in the social or applied sciences (e.g., psychology, communications, and counseling) or a related field, and trainees are typically undergraduate study-abroad students, international business employees, diplomats, or volunteers with international assignments. Although the content of each cross-cultural training program varies widely, the essential goals are to introduce cultural differences and adaptive strategies for cultural adjustment through the use of a variety of training techniques, including lecture, discussion, and experiential learning activities that involve the active participation of trainees. A successful and effective training program will include a good balance of these techniques and will allow trainees to reflect upon their own experience in the program's activities through discussion and cultural theory integration from lecture. Depending on the participants' characteristics, training programs may be specifically designed for the host country of the trainees or generally geared for adapting to life abroad. More complete programs involve predeparture and reentry sessions, and trainers maintain contact with participants during their stay abroad. Upon completion, cross-cultural training programs are evaluated and continuously improved for the benefit of future trainees. Although the majority of programs involve sojourners, cross-cultural training programs can also be structured to help people communicate and work more effectively with members of different cultures, races, or ethnic groups living in their own country.

## **1. THE NEED FOR CROSS-CULTURAL TRAINING**

Cross-cultural training programs are typically designed to introduce sojourners to cultural differences and to prepare them for success in their host countries. Successful programs are tailored to the needs and goals of the trainees, which are assessed prior to the start of the training program. Especially when based on an underlying theory, well-designed training programs have been shown to be more effective than no training at all or the brief, less structured country-specific orientation sessions typically held by international companies, volunteer organizations, or university study abroad programs.

When the purpose of the training program focuses on improving within-country interactions, the emphasis is placed on reducing stereotypes and prejudices. This type of cross-cultural training, also known as diversity training, is becoming increasingly more common in the United States as the population continues to become more diverse and work groups consist of employees from various racial, ethnic, and cultural backgrounds.

### **1.1. Needs Assessment**

The needs assessment typically involves surveying trainees for information regarding their family, work, language, or other characteristics. Trainees with families will be concerned about locating a suitable living place, schools for their children, and a place of employment for their spouses. Details involving their place of employment must also be addressed, including the use of a foreign language in the workplace or at home with a host family, change of position in a company, transportation to and from work or school, and other specific concerns. Trainees will often be faced with quite different living experiences in their host country than those to which they are accustomed. Particularly not to be overlooked is any reduction in status that the overseas position may involve, making it even more difficult for the sojourner to adjust to his or her new role and life in the host country. Adaptive strategies to prepare for and handle these changes are introduced and practiced in cross-cultural training programs.

Diversity training program trainers consider the needs of trainees and the organization to which they belong. Work groups must be surveyed for demographic information and the trainer must be familiar with the demands and goals of the particular groups.

Organizational leaders also need to be consulted for the identification of any issues that may be of special concern for the trainer to keep in mind when structuring his or her program.

### **1.2. Goal Identification**

The identification of goals is essential for the success and effectiveness of any cross-cultural training program. Specific goals are based on the particular needs of the trainees involved and should be expressed to them by the trainer at the start of the program. If the goals of the program are not aligned with needs of the trainees, participants can be uncooperative or doubtful of the usefulness of the program. Statistics on the percentage of sojourners successfully completing their overseas assignments can also be presented in support of the benefits of cross-cultural training programs. Once participants are convinced of the need for training, they are usually more willing to play an active role in the program. An effective cross-cultural training program usually combines learning exercises and discussions with lectures on culture theory. Some of the most general goals of cross-cultural training programs include increasing awareness of cultural differences, identifying and improving attitudes, and introducing new behaviors to aid in adaptation to the host culture.

#### ***1.2.1. Increasing Awareness of Cultural Differences***

One major goal of cross-cultural training programs involves the trainees gaining an increasing awareness for the norms, values, and other specific information about their own culture and the host culture. Its importance can be demonstrated through specific examples of cross-cultural misunderstandings resulting, in part, from a lack of awareness of cultural differences. This goal of the program can be achieved through the presentation of information in lecture or video format and reinforced through activities and discussion.

#### ***1.2.2. Identifying and Improving Attitudes***

Helping the participants of a cross-cultural training program to become aware of preexisting negative attitudes or biases toward interacting with culturally different individuals can be difficult but is a very important goal. It is difficult because people do not generally like to be made aware of their own negative self-aspects, but it is important for these negative

attitudes to be exposed and identified before sojourners must interact with members of their host culture.

A cultural simulation activity is one of the best techniques to use for the participants to get a feel for what it is like to face intercultural communication problems. Through these activities, they will be immersed in a situation in which they will most likely experience anxiety, frustration, and anger. After these are identified in a discussion, coping techniques can be suggested and practiced.

### **1.2.3. Introducing New Behaviors**

Particularly when the host culture is very different from that of the trainees, one of the goals of the program may be to introduce and practice new behaviors that are appropriate and very common in the host culture. Exercises involving greetings may be especially useful for a Westerner who is planning a sojourn to an Eastern culture, for example. These exercises may be practiced in a role play preferably with a member of the host culture who is involved in the training program.

## **2. PEOPLE INVOLVED IN TRAINING PROGRAMS**

A cross-cultural trainer plays many roles during the course of a training program. He or she must be very sensitive to the needs of the trainees, sometimes acting as a teacher, counselor, and role model. Trainees must also be open to the experiences of the program and willing to express their feelings and voice their opinions. Other members of a cross-cultural training program include members of the host culture or returning sojourners who may act as role models and cofacilitators during the program.

### **2.1. Trainers**

The leaders of cross-cultural training programs typically have educational backgrounds in the social or applied sciences (e.g., psychology, sociology, anthropology, linguistics, communications, and counseling) along with cross-cultural experiences of their own. These qualifications, however, do not necessarily make for an effective trainer. A trainer with reputable credentials and cross-cultural experience, for example, may lead a program based solely on his or her intuitive feelings of what he or she thinks needs to be done.

Since this type of approach has been shown to be less effective than those grounded in psychological or communication theories, it is essential for an effective trainer to be not only well educated and experienced but also familiar with and willing to use relevant theory and research in applicable fields.

Important personal characteristics of effective trainers include empathy, patience, flexibility, open-mindedness, and a high level of intercultural sensitivity. An empathetic and sensitive trainer, for example, is quick to recognize highly ethnocentric trainees and can actively restructure his or her program to suit their needs. Responding to the specific needs of the trainees requires a great deal of patience and flexibility on the part of the trainer, especially when faced with a group that has not voluntarily chosen to sojourn or participate in the training program. After the program has been evaluated, an open-minded trainer will thoughtfully consider the feedback of his or her trainees and make appropriate adjustments for future programs.

### **2.2. Trainees**

Trainees of cross-cultural training programs are typically international business employees, diplomats, volunteers, or university students preparing to study abroad. Since the structure of a program primarily depends on the type of participants involved, business employees, for example, will go through a quite different training program than undergraduates planning to study abroad. Training techniques used for business employees will focus on the particular issues they will be facing in the workplace and with their families, whereas programs for students studying abroad will discuss issues related to host families and schools.

Individual characteristics of the trainees, including degree of ethnocentrism and intercultural sensitivity, also play a role in the structure of the program. Highly ethnocentric individuals may benefit more from experiential activities, whereas those who are less ethnocentric and have more cross-cultural experience may gain more from discussing cultural or communication theories presented in lecture. Programs with a large number of trainees may be structured so the latter individuals serve as group discussion leaders, reporting back to the entire group while the trainer provides feedback.

### **2.3. Others**

Additional participants of cross-cultural training programs may include members of the host culture who

might serve on a sort of discussion panel to answer trainees' questions or actively participate in role plays. These role plays are designed to introduce new behaviors to the trainees, providing an opportunity to "practice" with a member of the host culture and experience a cross-cultural interaction prior to their sojourn. Also, returning members of the same organization who have spent time in the same host culture may be involved in programs to answer trainees' questions, provide advice, and talk about their experiences.

### 3. STRUCTURE OF TRAINING PROGRAMS

After the goals of the program have been established, trainers must consider the type of training techniques to be used (lecture, discussion, or experiential activity), the focus of these techniques (culture-general or culture-specific), sequence, and time allocation for each. Time and resource availability can pose a difficult challenge for trainers while structuring a cross-cultural training program, and all these elements must be considered when designing a program.

One major structural issue to consider is the inclusion of a reentry component to the training program. The most complete cross-cultural training programs involve a series of predeparture training sessions before the sojourners leave for their host country, sessions immediately upon arrival to their host country, and a series of reentry sessions beginning before the sojourners return home. Reentering is often more stressful than entering a host country for the first time, which is compounded by the fact that sojourners are generally not expecting it to be a difficult transition for them. Reentry training offers a time to discuss expectations with sojourners before they leave their host country and provides a social support network of people that have been through a similar experience. Additional benefits of reentry workshops include building greater insight into self-identity and a time for reflection on what was learned during the overseas experience.

#### 3.1. Types of Training Techniques

The techniques available for use in a training program can be classified into two broad categories. Didactic techniques involve the passive role of trainees and include lectures and videotapes, whereas experiential techniques involve the active role of trainees and

include role plays, discussions, cultural assimilators, and cultural simulation activities.

##### 3.1.1. Didactic Techniques

Lectures can introduce essential aspects of psychological or communication cultural theories. A psychological theory that is particularly useful for sojourners from the West who are going to live in the East and visa versa is individualism/collectivism. The defining features of these subjective cultural dimensions can be introduced and examples of common cross-cultural misunderstandings involving them can be given to aid in the understanding of these concepts.

For example, an indirect communication style is one tendency of a collectivistic country's culture, whereas a direct style of communication is more typical in an individualistic country's culture. After a specific example is provided, possibly through the use of a cultural assimilator, trainers can then turn the focus to the underlying theory and offer it to the trainees as a kind of framework for sorting through and understanding their own cross-cultural experiences to be encountered in their host country.

Another useful theory that can be introduced in lecture is the anxiety—uncertainty management (AUM) theory. The essential components of this communication theory revolve around the core components of mindfulness, anxiety, and uncertainty. First, trainees can be taught what it means to be mindful and may be given examples to illustrate its importance in cross-cultural interactions. Next, the trainer might explain that managing anxiety and uncertainty is necessary for successful cross-cultural interactions and demonstrate this through the use of specific examples. Trainees would then be asked to generate some ideas regarding what techniques they can use to manage their own anxiety in their host country. Following this, the trainer can ask them to provide specific examples of some difficult situations they may expect to encounter along with their own strategies for managing anxiety and reducing uncertainty. If this lecture follows a culture simulation activity, trainees can be asked to reflect upon how they felt during the simulation using the AUM theory to frame their responses. The trainees can then practice using these AUM-based strategies in a cross-cultural simulation activity.

##### 3.1.2. Experiential Techniques

In cultural simulation activities, participants are asked to form groups that receive separate instructions

regarding group communication or behavioral rules. The trainees are then asked to interact with members of the other groups for awhile to accomplish some type of task until they experience emotions that typically arise in actual cross-cultural encounters. Examples of such activities most commonly used are BaFa BaFa, Barna, and Ecotonos. One important source of misunderstandings is that members of the host culture make attributions concerning their own behavior that are very different from the attributions that the sojourners make concerning the behavior of the hosts. Exposing these discrepancies through simulation activities has proven to be beneficial. After the simulation, a trainer may debrief the trainees on the purpose of the activity and follow up with a discussion based on the principles taught in the lectures of the training program.

Another experiential training technique involves the use of cultural assimilators, through which the trainees are asked to work and discuss. These sets of incidents involve common cross-cultural misunderstandings followed by four possible options for the most correct explanations. Those that are grounded in cultural theories can be integrated with information presented in lecture and combined with the participants' experience in simulation activities.

For example, an incident from one such assimilator involves a new colleague arriving from China who occupies a desk next to an American colleague. As time passes, the American begins to be bothered by the strong body odor of his new colleague and wishes to communicate this with him. The assimilator then presents a variety of possible responses. Among these are more or less direct approaches of how the American should communicate with the Chinese colleague. Culturally insensitive Western trainees are likely to choose a more direct response, such as "Present deodorant to the colleague and ask him politely to use it," rather than a more indirect approach, such as "Talk to the colleague's closer friends or family and have them communicate the message to him." This assimilator based on individualism/collectivism theory offers explanations based on the frequency of direct and indirect responses in the two cultures, thus identifying which explanation is most likely to be correct. In this example, the trainee is exposed to the importance of "saving face" in more collectivistic countries and the necessary indirect communication style that is characteristic of this value. Like most assimilators, this particular incident is available in a written format, but it is also available in an interactive computer audio/video format. Paper has traditionally been the only assimilator

format available; however, computer-based software programs have recently become available.

The role play is another example of an experiential training technique. Particularly when individuals from the host country are involved in a training session, a role play can be an excellent opportunity for the trainees to practice behaviors that may not be practiced as often or at all in their own culture. An American sojourner whose host country is Japan, for example, may be placed in a situation with a Japanese playing the part of a business colleague. They might be asked to play out a situation in which they are meeting for the first time, and the trainee would then practice exchanging business cards, bowing, and using formal Japanese when asking a "new Japanese business colleague" appropriate questions.

Within experiential and didactic training techniques, another distinction differentiates between those focusing on a specific host culture, called culture-specific, and training that focuses on broad aspects of cultural differences, called culture-general. Among the advantages of the cultural-general approach is the ease of accessibility to a set of culture-general assimilators that have been published or other sets that are in wide circulation. Culture-specific assimilators, although very useful for groups of trainees sojourning to the same country, can be difficult to obtain without having a large network of professionals in the field.

### 3.2. Time and Resource Availability

Although trainees would benefit most from training programs wide in breadth and depth, time constraints and limited resources often pose a challenge for trainers to make the most of what is available to them. An ideal program would include thorough, informative lectures combined with language training (if applicable), cultural simulation activities, and discussions to integrate all the experiences in the training program. The focus (culture-general or -specific) would be designed to suit the needs of the individuals in the group and, at a minimum, 3 full days of training is generally recommended for a training program to be effective.

## 4. PROGRAM EVALUATION AND DIMENSIONS OF SUCCESS

Anonymous written feedback from participants is perhaps the most basic means of evaluating how well the program was received by the trainees. This feedback

may be used to modify future programs and is an indication of how well the trainer performed in his or her role as the facilitator of the group. Regarding particular training techniques, the results of laboratory studies support the effectiveness of assimilators with regard to raising intercultural sensitivity and reducing prejudices. Results of longitudinal studies also support the effectiveness of cultural assimilator-based training on cross-cultural adjustment, communication, and interpersonal problem-solving ability. These three dimensions of success in the host culture are generally known as personal adjustment, interpersonal effectiveness, and professional effectiveness.

#### 4.1. Personal Adjustment

During the first month or two of the sojourn, individuals typically endure the uncomfortable experience known as culture shock. However, after this initial negative reaction to the new environment and people, those who have completed cross-cultural training programs have been found to experience less stress in their host culture regarding their personal adjustment in comparison to a group of sojourners that did not receive formal cross-cultural training. Information about cultural differences that is introduced in the training program combined with their simulated cross-cultural experiences and rehearsed behaviors allow them to anticipate difficulties and adapt more readily to situations in their host culture.

#### 4.2. Interpersonal and Professional Effectiveness

The two remaining dimensions of success in the host culture are more closely related to the sojourners' interactions with their new friends and colleagues. These are largely a function of the sojourner's ability to communicate in a second language and his or her familiarity with the host culture's norms and values. In the training program, the use of cultural assimilators, role plays, and the presentation of information related to the specific work environment can prepare a sojourner to do well on this dimension. Indeed, compared to a group with no formal training, trainees are more likely to complete their overseas work assignments and have a better social support network in the host culture.

## 5. CONCLUSION

Well-designed cross-cultural training programs can better prepare sojourners for their experience in their host country and experience upon returning to their native country. A successful program lies in the alignment of goals to the needs of the participants and the rapport created between the trainer and the trainee. An effective program depends primarily on the blend of didactic and experiential techniques involving the use of culture-general or culture-specific training or a combination of the two. It is the job of the trainer to blend the optimal mix of the two types in order to satisfy the participants of the program while maximizing its effectiveness.

### See Also the Following Articles

Diverse Cultures, Dealing with Children and Families from ■ Interpersonal Behavior and Culture ■ Social Skills Training

### Further Reading

- Brislin, R., & Yoshida, T. (1994). *Intercultural communication training: An introduction*. Thousand Oaks, CA: Sage.
- Cushner, K., & Brislin, R. (1996). *Intercultural interactions: A practical guide* (2nd ed.). Thousand Oaks, CA: Sage.
- Gudykunst, W. B., & Kim, Y. Y. (2003). *Communicating with strangers* (4th ed.). Boston: McGraw-Hill.
- Kim, Y. Y., & Gudykunst, W. B. (Eds.). (1988). *Cross-cultural adaptation: Current approaches*. Newbury Park, CA: Sage.
- Landis, D., Bennett, M., & Bennett, J. (Eds.) (2003). *Handbook of intercultural training* (3rd ed.). Thousand Oaks, CA: Sage.
- Landis, D., & Bhagat, R. (Eds.). (1996). *Handbook of intercultural training* (2nd ed.). Thousand Oaks, CA: Sage.
- Landis, D., & Brislin, R. (Eds.). (1983). *Handbook of intercultural training* (Vols. 1-3). Elmsford, NY: Pergamon.
- Nipporica Associates. (1993). *Ecotonos*. Yarmouth, ME: Intercultural Press.
- Shirts, R. G. (1974). *BaFa BaFa: A cross-cultural simulation*. Del Mar, CA: Simulation Training Systems (formerly Simile II).
- Thiagarajan, S., & Steinwachs, B. (1990). *Barnga*. Yarmouth, ME: Intercultural Press.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.



# Traits

Jesús F. Salgado

University of Santiago de Compostela, Santiago de Compostela, Spain

1. Traits as Human Characteristics of Individual Differences
2. The Big Five Dimensions of Personality
3. Criticism of the Trait Approaches and the Five-Factor Model of Personality
4. Traits, Heritability, and Genetics
5. The Big Five and Organizational Behavior Further Reading

## GLOSSARY

**agreeableness** The degree to which individuals are cooperative, warm, and agreeable versus cold, disagreeable, rude, and antagonistic.

**conscientiousness** The degree to which individuals are hard-working, organized, dependable, reliable, and persevering versus lazy, unorganized, and unreliable.

**emotional stability** The degree to which individuals are secure, nonanxious, calm, self-confident, and cool versus insecure, anxious, depressed, and emotional.

**extraversion** The extent to which individuals are gregarious, assertive, and sociable versus reserved, timid, and quiet.

**non-shared environment** Environmental factors that do not contribute to sibling similarity.

**openness to experience** Defines individuals who are creative, curious, and cultured versus practical with narrow interests.

**shared environment** Environmental factors that contribute to twin and sibling similarity.

**traits** Dimensions of individual differences in tendencies to show consistent patterns of thoughts, feelings, and behaviors.

This article describes the concept of traits as they are used for describing the human personality and the relationship between the traits and human languages. It also discusses the psycholexical approaches for describing and explaining personality. The most accepted trait-based personality model, the five-factor model, is described and some criticisms are presented. Additionally, recent evidence of heritability and genetic bases of the big five personality dimensions is discussed. Finally, advances regarding the relationship between the big five personality dimensions and organizational behaviors and processes are mentioned and the magnitude of the relations is reported.

## 1. TRAITS AS HUMAN CHARACTERISTICS OF INDIVIDUAL DIFFERENCES

It is very probable that human beings have used personality traits for describing themselves and others since they were first able to verbally communicate. This statement suggests four relevant implications: (i) Traits are conceptual attributes used for describing ourselves and others, (ii) traits are only possible because they are rooted in human language, (iii) traits are stable over time and across situations, and (iv) traits are related to human behavior.

The trait-based approaches to personality focus on the conceptualization that individuals differ from one another in a small number of dimensions that are

stable over time and across situations. In this sense, traits are defined as dimensions of individual differences in tendencies to show consistent patterns of thoughts, feelings, and behaviors that allow the characterization of human beings. Tendencies implies that traits are dispositions and not absolute determinants of human action. Consistent patterns of thoughts, feelings, and actions implies that traits should be distinguished from habits, repetitive actions, and mechanical behaviors. Furthermore, as Allport noted many years ago, personality traits are not real but rather abstractions that cannot be directly measured and must be inferred.

Different approaches to personality traits have been suggested throughout the years; the first was the lexical approach to the traits suggested by Galton in the 19th century. Galton proposed that personality characteristics are encoded in the human language and, consequently, are included in dictionaries. Following Galton's seminal work, in 1936 Allport and Odbert classified all the trait words included in the American English language as were registered in *Webster's New International Dictionary*. They identified 17,953 personality terms classified in four categories: personal traits, evaluative traits, metaphorical and temporary states, and moods. After eliminating all the words included in the last three categories, they developed a list of 4504 words that Americans used to describe themselves and others. Because this large list was not very practical from both a theoretical and an applied standpoint, in 1943 Cattell tried to reduce the number of trait words using a variety of conceptual and statistical techniques. First, he reduced the large number of words to 171 using conceptual and semantic similarity. Then, using cluster analysis, he grouped the short list of words and found 35 clusters. A factor analysis of the 35 bipolar scales produced 12 primary dimensions. Cattell added four dimensions that he had found by analyzing questionnaires and the result was the basis of his theory of personality: 16 primary personality factors. Throughout the years, a number of researchers have investigated Cattell's structure of personality and found some incongruence.

Soon after Cattell's study, Eysenck stated his theory of personality, which consisted of two independent factors, namely neuroticism (the negative pole of emotional stability) and extraversion. Later, Eysenck added a third factor, psychoticism, and the result was the well-known P-E-N system.

## 2. THE BIG FIVE DIMENSIONS OF PERSONALITY

The psycholexical approach to personality has led to the accumulation of empirical evidence supporting five personality dimensions that summarize the relations between the personality traits. This evidence was found for many different nations and with many different languages, although some minor differences exist across the national borders and the languages with different linguistic roots. These five personality dimensions, the "Big Five" factors, are extraversion, agreeableness, conscientiousness, emotional stability, and intellect. These five factors have also been labeled surgency (extraversion), neuroticism (emotional stability), openness to experience (intellect), dependability (conscientiousness), and many other terms. Currently, the most used labels are emotional stability, extraversion, openness to experience, agreeableness, and conscientiousness. The five-factor model (FFM) assumes that the personality traits are hierarchically arranged, with specific narrow traits at lower levels in the hierarchy and the broad trait dimensions at the top.

The current FFM of personality originated from works by Fiske, Borgatta, Norman, Smith, and Tuppes and Christal, who factor analytically reproduced a highly stable structure with five factors. However, they were unable to reproduce Cattell's 16 primary factors, and they found evidence for only five factors. In the FFM, emotional stability concerns the degree to which the individual is secure, nonanxious, calm, self-confident, and cool versus insecure, anxious, depressed, and emotional. Extraversion concerns the extent to which individuals are gregarious, assertive, and sociable versus reserved, timid, and quiet. Openness to experience defines individuals who are creative, curious, and cultured versus practical with narrow interests. Agreeableness concerns the degree to which individuals are cooperative, warm, and agreeable versus cold, disagreeable, rude, and antagonistic. Conscientiousness concerns the degree to which individuals are hardworking, organized, dependable, reliable, and persevering versus lazy, unorganized, and unreliable.

The five factors represent the common variance among a large set of more specific traits or facets. Saucier and Ostendorf identified hierarchical subcomponents of the Big Five personality factors and they replicated the findings in English and German. In 1997, Saucier and Ostendorf found that 18

subcomponents were necessary to provide a finely faceted model for the Big Five. Based on their results, extraversion consists of four facets—sociability, unrestraint, assertiveness, and activity; agreeableness consists of four facets—warmth, gentleness, generosity, and modesty; conscientiousness consists of four facets—orderliness, decisiveness, reliability, and industriousness; emotional stability consists of three facets—low irritability, low insecurity, and low emotionality; and openness to experience consists of three facets—intellect, imagination, and perceptiveness.

### 3. CRITICISM OF THE TRAIT APPROACHES AND THE FIVE-FACTOR MODEL OF PERSONALITY

During the 1960s and 1970s, the trait approaches to personality were challenged based on a number of criticisms, mainly the lack of empirical evidence supporting the notion of temporal stability and cross-situational consistency. The failure to find evidence supporting these assumptions drove many personality researchers to conclude that traits were a theoretical construct useless for explaining personality and individual differences. In 1968, Mischel made a particularly relevant criticism of the trait approaches. He claimed that personality test scores (presumably reflecting traits) seldom correlated higher than .30 with other behavioral criteria. This was interpreted as strong evidence that behavior depends on situations.

The first trait approaches to personality structure were also criticized for other reasons. For example, Cattell's and Guilford's systems were criticized because their factors were not generally replicable across gender, age, and methods, and many researchers have failed to find them. Eysenck's P-E-N system was criticized on the basis that the psychoticism dimension is an amalgam of agreeableness and conscientiousness.

The FFM is currently the most influential model of personality; however, it is not unanimously accepted, and some criticisms have been made. For example, Block noted that the FFM is mainly based on factor analytical results, but in his view factor analysis is not an appropriate and sufficient base to decide the theoretical constructs of personality. Block also criticized the assumptions of the lexical hypothesis and the variability of conceptions about the Big Five. Other researchers have criticized the model on the basis of

the need for many factors (e.g., Eysenck) or an insufficient number of factors (e.g., Hough). Such criticisms were answered by Costa and McCrae and Goldberg and Saucier, who noted that the Big Five were reproduced a great number of times, with different factor methods, by different researchers, with different instruments, and in different languages. Other factors were found only in isolated samples. Consequently, the model is not exhaustive, but it manages to support a more comprehensive explanation of various facets of human personality. Peabody and De Raad suggest that the Big Five personality structure cannot be as universal as some influential researchers have stated.

### 4. TRAITS, HERITABILITY, AND GENETICS

In recent years, many researchers have examined the genetic origin of human traits, and hundreds of papers have been devoted to this issue. The main conclusion of this research is that genes make a major contribution to individual differences in personality, especially when assessed by questionnaires. In the past 10 years, genetic research on personality has mainly focused on the Big Five personality dimensions, both in America and Europe. Loehlin summarized five large twin studies in which extraversion and emotional stability were examined. All the studies used self-report questionnaires that had not been developed for assessing the Big Five (with one exception). The measures of extraversion and emotional stability (neuroticism) were the most represented. The total sample comprised 24,000 pairs of twin. The findings showed heritability estimates of 49% for extraversion and 41% for emotional stability. Furthermore, the findings showed that the environmental factors were very important but that the environmental influence was mostly due to non-shared environmental effects rather than to the shared environmental influences. Shared environment is defined as environmental factors that contribute to twin and sibling similarity, whereas nonshared environment is defined as environment factors that do not contribute to sibling similarity.

The results of recent studies using questionnaires explicitly developed for assessing the Big Five personality dimensions (e.g., NEO-PI-R) confirm that the genetic influences on personality are strong. For example, using the NEO-PI questionnaire, in 1996 Jang and colleagues found that genetic influences accounted for

41, 53, 61, 41, and 44% for emotional stability, extraversion, openness, agreeableness, and conscientiousness, respectively. In 1998, Leohlin and colleagues, using 47 self-rating scales assessing the Big Five, estimated the genetic influence at 58, 57, 56, 51, and 52% for the same personality dimensions, respectively. In 1997, Riemann and colleagues, using the German self-report version of the NEO-FFI, found that genetic influences explained 52, 56, 53, 42, and 53% of the variance for the same dimensions, respectively. These studies examined all the variance sources using self-reported questionnaires, but there is also evidence from peer-reported questionnaires. In this case, the results suggest that genes account for two-thirds of the variance and nonshared environment accounts for the remaining one-third. Riemann and colleagues found that genes explained 61, 63, 81, 57, and 71% of the variance in the peer reports for emotional stability, extraversion, openness, agreeableness, and conscientiousness, respectively.

Additional information on the genetic and environmental sources of influence on personality was obtained by Borkenau *et al.* in 2001. These researchers independently examined the three sources of variance using ratings provided by judges who never met the twins but observed videotaped behaviors of one twin of each pair in 1 of 15 different settings. The aggregated video-based trait ratings, which were highly reliable (reliability mean = .94), showed that the genetic influences for the Big Five ranged from .38 to .62; that the shared environmental influences ranged from .20 to .28, with the exception of extraversion, which was not influenced; and that the nonshared environmental influences ranged from .30 to .38. In all cases, the genes were the main source of variance (Table I). An advance in

clarifying the role of the genetic influences was made by Jan *et al.* in 2001. They found that the serotonin transport gene accounted for 10% of the relationship between emotional stability and agreeableness.

## 5. THE BIG FIVE AND ORGANIZATIONAL BEHAVIOR

For many years, personality measures were considered poor predictors of organizational behaviors. This belief was mainly due to the small correlations found between personality characteristics and the organizational criteria (e.g., job performance) and in part due to the strong criticisms made by those with the situationist view of personality. However, during the past decade, many studies have shown that the Big Five personality dimensions are relevant variables for explaining many organizational behaviors and processes. Several organizational areas of interest for researchers were job performance, training success, teamwork, deviant behaviors, turnover, job satisfaction, leader emergence, leadership effectiveness, and transformational leadership behaviors. There are many cumulative results that allow to conclude that traits are relevant variables for predicting and explaining these organizational behaviors. For example, job performance is related to conscientiousness and emotional stability. Training success is mainly related to extraversion, openness, and conscientiousness. For teamwork, the dimensions with a strong relation are emotional stability, agreeableness, and conscientiousness. In the case of deviant behaviors, the best predictors of the absence of these dysfunctional behaviors are agreeableness and conscientiousness. For turnover, the most related dimensions are emotional stability, extraversion, agreeableness, and conscientiousness. Leadership behaviors were examined in relation to the emergence of leaders, the effectiveness of leaders, and the transformational leadership. In the three cases, the Big Five were shown to be relevant variables related to these organizational behaviors. For example, emotional stability, extroversion, openness, and conscientiousness were found to be related to the emergence of leadership. The effectiveness of leaders was associated with emotional stability, extraversion, openness, and agreeableness. Lastly, extraversion, openness, and agreeableness were related to transformational leadership behaviors (Table II). Thus, the dispositional variables of personality are thought to be relevant explanatory causes of many organizational behaviors and processes.

TABLE I  
Genetic and Environmental Influences on the Big Five  
Personality Dimensions<sup>a</sup>

Dimension	Genes	Shared Environment	Nonshared Environment
Agreeableness	.43	.27	.30
Conscientiousness	.38	.25	.37
Emotional stability	.50	.20	.30
Extroversion	.62	.00	.38
Openness	.39	.28	.33

<sup>a</sup>Source: Borkenau *et al.* (2001).

**TABLE II**  
**Correlations among the Big Five and Organizational Behaviors and Process<sup>a</sup>**

Personality transformational dimension	Job performance	Training success	Deviant behavior	Turnover	Teamwork	Leadership emergence	Leadership effectiveness	Leadership
Emotional stability	.13	.08	-.06	-.35	.20	.24	.22	.03
Extraversion	.13	.23	.01	-.20	.16	.33	.24	.28
Openness	.13	.24	.14	-.14	.12	.24	.24	.26
Agreeableness	.07	.11	-.20	-.22	.27	.05	.21	.32
Conscientiousness	.31	.23	-.26	-.31	.23	.33	.16	-.06

<sup>a</sup>Sources. Barrick *et al.* (2001), Judge and Bono (2000), Judge *et al.* (2002), and Salgado (2002).

**See Also the Following Articles**

Agreeableness ■ Conscientiousness ■ Emotion ■ Extroversion-Introversion ■ Openness to Experience ■ Optimism ■ Personality Assessment ■ Psychological Assessment, Standards and Guidelines for

**Further Reading**

Barrick, M., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-30.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117, 187-215.

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F. M. (2001). Genetic and environmental influences on observed personality: Evidence from the German observational study of adult twins. *Journal of Personality and Social Psychology*, 80, 655-668.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.

Goldberg, L. R. (1990). An alternative "description of personality": The Big Five factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.

Jan, K. J., Hu, S., Livesley, W. J., Angleitner, A., Reimann, R., Ando, J., Ono, Y., Vernon, P. A., & Hamer, D. H. (2001). Covariance structure of neuroticism and agreeableness: A twin and molecular genetic analysis of the role of the

serotonin transporter gene. *Journal of Personality and Social Psychology*, 81, 295-304.

Jan, K. L., Livesley, W. J., & Vernon, P. A. (1996). Heritability of the Big Five personality dimensions and their facts: A twin study. *Journal of Personality*, 64, 577-591.

Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology*, 85, 751-765.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765-780.

Loehlin, J. C., McCrae, R. R., Costa, P. T., & John, O. P. (1998). Heritabilities of common and measure-specific components of the Big Five personality factors. *Journal of Research in Personality*, 32, 431-453.

McCrae, R. R., & Costa, P. T., Jr. (1990). *Personality in adulthood*. New York: Guilford.

Peabody, D., & De Raad, B. (2002). The substantive nature of psycholexical personality factors: A comparison across languages. *Journal of Personality and Social Psychology*, 83, 983-997.

Plomin, R., DeFries, J. C., McClearn, G. E., & Rutter, M. (1997). *Behavioral genetics* (3rd ed.). New York: Freeman.

Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10, 117-125.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*, 76, 613-627.



# Transfer of Learning

Robert E. Haskell

University of New England, Old Orchard Beach, Maine, USA

1. Overview
2. Conceptual Issues and Problems
3. Categorical Distinctions
4. Models of Transfer
5. Historical Overview
6. Transfer Thinking and Reasoning
7. Higher Order Transfer
8. Instructional Issues
9. Dispositional Characteristics and Aspects of Transfer
10. Biological Origins of Transfer
11. Future Empirical Research and Theory  
Further Reading

## GLOSSARY

**analogical reasoning/transfer** Based on analogical proportionality (e.g.,  $A : to B :: C : D$ ).

**causal** A variable responsible for causing an effect or event.

**counterintuitive** Counter to what seems reasonable to expect.

**induction** A form of reasoning in which a small set of facts is used to justify a more general assertion.

**isomorphic relations** One-to-one structural relations between two phenomena that preserve the relations among their elements.

**model** A simplified working representation of a phenomenon.

**structural relations** Deeper underlying causal relations.

**taxonomy** A descriptive and systematic classification scheme.

Transfer of learning is considered to be the use of past learning in the learning of something new and the application of learning to both similar and new situations. The term transfer is derived from *trans*, meaning across or over, and *ferre*, meaning to bear, thus, to carry over.

## 1. OVERVIEW

Transfer of learning has a long history, beginning in instructional contexts and subsequently examined by laboratory research. Long considered a foundational educational concept, research on transfer of learning in formal educational settings has been less than encouraging. In addition, transfer of learning has been plagued by problems of definition and lack of coherent theory. In recent years, the concept has undergone reconceptualization from an educational and instructional concept to reflect a cognitive function undergirding thinking and reasoning. Because of its underlying basis in similarity relations, the concept of transfer of learning implies involvement in a host of other concepts, such as analogical reasoning, isomorphic relations, and generalization processes.

### 1.1. Conceptualizing Transfer

Beyond the general definition, although there is some disagreement regarding precisely how to define transfer, it is clear that it involves seemingly simple

similarity relations. The everyday basis of transfer can be exemplified by the following phrases: “it’s like . . .,” “it’s equivalent to . . .,” “it’s akin to . . .,” “for instance . . .,” “it’s the same as . . .,” “similarly . . .,” “in the same way . . .,” “it resembles . . .,” or “it’s analogous to . . .” Accordingly, transfer as a cognitive function is involved in the use of figurative language as with analogies and metaphors. In addition, although not generally seen as involving transfer, because they are seen simply as an “instance” within a category, whenever an “example” is invoked transfer already has occurred. Examples are typically generated on the basis of similarity relations between the example and what it is seen as exemplifying. It can be seen that undergirding the cognitive concept of transfer and its various exemplifications is an invariance operation that may be possibly biologically based (Fig. 1).

### 1.2. Exemplifications

Transfer is pervasive in everyday activities. New things often are understood in terms of past experiences, as being like something already familiar. When first invented, cars were seen as “horseless carriages.” With the advent of the computer age, there is now electronic “mail” and “notepads” as well as electronic “notebooks.” Transfer is indicated when experience riding a bicycle is

applied to learning to ride a motorcycle, when experience playing the piano is applied to playing an accordion, when experience ice skating is applied to roller blading, or when knowledge of one word processing program is applied to learning another. These examples are generally considered near transfer.

In formal educational settings, there is supposed to be transfer both within and among different knowledge domains and courses. For example, it is expected that there will be transfer or the application of what is learned from a mathematics course to everyday situations that require quantitative calculations.

On a higher level, transfer is manifested by recognizing that a concept in one subject area is like a concept in another seemingly unrelated subject area. For example, although it is seldom recognized, the concepts of learning in psychology (the process of acquiring knowledge or skill through practice, training, or experience), socialization in sociology (the social process whereby individuals acquire the values and behavior patterns of their society or social group), adaptation in biology (the alteration in an organism resulting largely from natural selection enabling the organism to survive), and acculturation in anthropology (the process of acquiring cultural traits or social patterns of another culture) are very similar. Recognition of these examples generally would be considered far transfer.

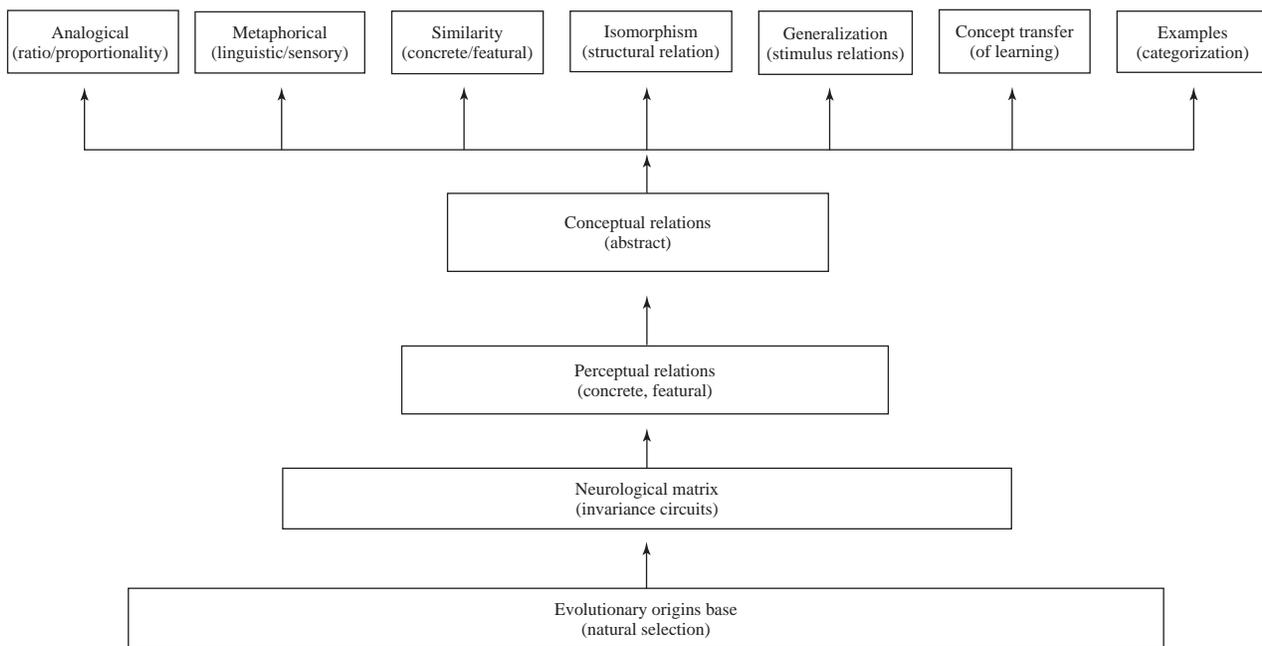


FIGURE 1 Transfer relations of invariance: domains and origins.

### 1.3. Importance

For more than 100 years, it has generally been agreed that the aim of all education—elementary, secondary, vocational, industrial, and business training—is to understand, extend, and apply what is learned in one context to similar and new situations. There continues to be a consensus that transfer is fundamental to all learning. Accordingly, transfer has become what Resnick characterized as the Holy Grail of education.

The demands of modern civilization make skill at transfer increasingly important. In a highly complex, rapidly changing, information-age society, the ability to transfer or generalize from the old to the new not only renders the world predictable and understandable but also is a necessity for adaptation to increasing technological and global advances. This is especially true where job and vocation changing are becoming the norm. Transfer is also seen by some as an important conceptual tool for interdisciplinary understanding and work.

### 1.4. Cognitive Benefits of Transfer

Although not typically stated in cognitive terms, transfer is extremely economical in terms of learning resources. For example, learning the aforementioned four concepts of learning, socialization, adaptation, and acculturation, each from a different discipline, requires a certain amount of memory storage space. Recognizing that the concepts are essentially the same, they can then be “chunked” into one larger concept, thus easing the load on memory storage. After the four concepts are seen as being alike, they become associated with each other, as well as other material associated with them, thus facilitating retrieval of the information. The four concepts then become cognitively integrated for a deeper understanding of each one. Finally, since the four concepts are not identical, certain aspects from one concept may reveal something new about the others that was not known prior to the transfer integration.

## 2. CONCEPTUAL ISSUES AND PROBLEMS

Despite the seemingly simple definition and examples of transfer given earlier, research reveals three basic problems of conceptualization involving not only the term but also its application, research design, and instructional outcomes.

### 2.1. Terminological Issue

There have been problems with the conceptualization of the term transfer, with some critics maintaining that the term is simply a metaphor since nothing actually gets transferred or “moved” from one domain or from one idea to another. The position is that all prior learning either simply enhances or, conversely, inhibits learning. In this view, the concept of transfer has no special merit or claim. Although the argument has some logical validity, if only for pragmatic and research purposes, most do not adopt this view. Nevertheless, the traditional distinction between the terms learning and transfer of learning is increasingly being questioned, with transfer being seen as primary.

### 2.2. The Question of Application Transfer

The second problem of conceptualizing transfer involves the expansion of its meaning. Although the extension of the original educational term to reflect a full-fledged cognitive function increasingly is accepted, it has been expanded in another way that does not have such acceptance. For example, when calculating the area of a  $10 \times 10$  in. square learned in a math class, which then is applied to finding the area of an actual  $10 \times 10$  ft. floor, it is generally considered to be a transfer of the classroom learning. Critics maintain that neither simple application of an abstract formula to a concrete situation nor the simple learning of finding the square feet of a classroom floor, when applied away from the physical place it was learned, should be considered transfer.

### 2.3. The Question of Facilitated Transfer

A third problem involves assessing whether transfer actually has taken place in a given instructional or research setting. Perhaps the leading critic of contemporary transfer research is Detterman, who argues that most transfer studies that demonstrate some low level of transfer in fact do not demonstrate transfer. For example, in one representative study children learned to stack toy tires on top of each other, eventually allowing a doll to reach a shelf on which other tires were to be stacked. Then, a similar problem was presented that required children to stack bales of hay on top of each other so that a farmer doll could reach a tractor. Three sets of such problems were used. It can

be seen that although the surface structure of each story in the set was slightly different, the underlying structure was transparently the same. If children were not able to solve a similar problem, the experimenter demonstrated the solution to them. Children were always asked to repeat the solution to the problem, at which time the next structurally identical problem was presented. Detterman maintains that the children did not engage in transfer but simply learned the “rules of the game” or were repeatedly shown the transfer over the three problem sets.

There are other methodological problems in studies that claim to find transfer, such as when subjects are told that previous material may be useful in the solution of a new problem, are explicitly informed about strategies and methods known to improve learning on the material to be learned, are instructed to use those strategies on that material, have a similarity of the material pointed out to them, and/or are given other hints about a similarity between the problems to be solved. Critics maintain that it is not reasonable to refer to the learning or problem solution as the result of transfer. The assumption behind such critiques is that only unaided or spontaneous transfer constitutes transfer. Based on research findings, however, spontaneous transfer seems to occur only 20% of the time in both adult and child populations (using age-appropriate tasks).

#### **2.4. Outcome Research**

Consistent findings during the past 100 years on which most researchers agree are that significant transfer in formal instructional settings seldom occurs, but it is pervasive in everyday situations; learning often is “welded” to the specific concrete context in which it is learned; when transfer does occur in formal instructional situations, it is frequently on such a basic conceptual level that many researchers do not consider it transfer at all but simple learning and application; and even when transfer occurs in informal everyday settings, it, too, is typically of a basic character scarcely distinguishable from simple learning or application of what is learned.

When more liberal assessments of the outcome literature are made, it is typically the result of less demanding conceptions or definitions of what constitutes transfer. It is not only general or significant (far) transfer that seems to occur infrequently but also the lower level application (near) transfer. It is widely agreed that the educational, social, technological, industrial, and business costs of transfer failure are considerable.

That transfer is pervasive in everyday settings is only seemingly contradictory to the negative findings in formal settings. If transfer is a general brain-based function, then it is reasonable to expect it to be pervasive in everyday settings on some level. However, the problem is to prevent transfer from being welded to the particular area in which it was learned, such as when transfer of mathematical operations occurs within mathematics but does not transfer to solving physics problems. In addition to the varying definitions, many variables have been identified that account for the failure of transfer, including insufficient learning of the original material, insufficient time encoding the original learning, inappropriate encoding of the material (e.g., rote vs deep or meaningful), insufficient learner motivation, inadequate memory span, the time lapse between original learning and the transfer task, insufficient learner background knowledge, and inappropriate instructional methods.

### **3. CATEGORICAL DISTINCTIONS**

As with any set of phenomena, it is possible to classify transfer along various dimensions. Accordingly, there have been various general taxonomic schemes suggested, as well as other instructional distinctions.

#### **3.1. Taxonomies of Transfer**

Attempts to construct a systematic taxonomy or classification system of transfer continue to be limited in scope. This is due, in part, to the problem of precisely conceptualizing and quantitatively measuring what is called the metric of similarity involved in transfer (e.g., near versus far transfer). Nevertheless, any systematic, albeit qualitative, attempts can be useful for providing a common frame of reference or classification system for understanding and researching transfer, as the systematic scheme by Barnett and Ceci shows.

One traditional and continuing broad scheme is near versus far transfer. The former refers to transferring or applying what is learned to something closely similar; the latter refers to transferring what is learned to something quite dissimilar. The problem with such taxonomies is that they are relative not only to definition but also to individual perception of similarity on the part of both the learner and the evaluator, as well as to context variables that influence what is perceived as “near” and “far.” What may be near transfer to one person may be far transfer to another. Transferring

within a domain of knowledge, however, is generally considered near transfer, and transferring across domains is considered far transfer.

In addition, there are many possible ways for classifying the different kinds of transfer, which include subject matter transfer, procedural or skill transfer, concrete transfer, abstract transfer, theoretical transfer, positive transfer, and negative transfer.

### 3.2. Positive and Negative Transfer

Transfer is considered positive when past experience leads to apprehending a valid or useful similarity relation. In general, negative transfer occurs when past experience leads to apprehending or applying an invalid similarity relation.

Negative transfer can also be the consequence of either insufficient knowledge or past learning interfering with current learning. An example of the former is when a whale is incorrectly thought to be a fish, or, conversely, when a porpoise and a deer are not seen as both being mammals. An example of the latter is when driving a vehicle with a clutch inappropriately activates the search for a clutch pedal the first time one drives a vehicle with an automatic transmission.

### 3.3. Transfer of Training versus Transfer of Learning

The long-standing distinction between learning or educating, on the one hand, and training, on the other hand, applies to instructing for transfer as well. Historically, the basic differences between training and learning are that training tends (i) to be more task specific, (ii) to be more concrete, (iii) to be less conceptual and theoretical, (iv) to be less process- and more immediate-outcome oriented, (v) to use a narrower range of knowledge base, (vi) to be less long-term outcome oriented, (vii) not to be as concerned with linkages to other knowledge bases, (viii) to result in a shallower depth of understanding, and (ix) to be less personally meaningful to the learner. Both training and educating are appropriate for their respective purposes.

Unlike in education, corporate instructional programs historically have been oriented toward training and the transfer of what has been learned in the training setting back to the workplace—a near application transfer based on physical distance. For both

understanding and promoting transfer, the business model approach tends to focus on managerial and organizational support systems, whereas an academic or educational approach to promoting transfer focuses on instructional and cognitive processes. Both approaches seem to be needed to promote significant transfer in most instructional settings (for some business organizations, within approximately the past decade there has been a shift from an industrial-age training model to an information-age learning model) (Table I).

## 4. MODELS OF TRANSFER

Throughout the years, numerous models of transfer have been suggested. Early models once thought to be invalid continue to be examined and new models continue to be developed. The most well-known are described here. Some models are concerned with instructional processes, others with understanding the cognitive mechanisms of transfer, and still others with how to promote transfer through learning strategies.

TABLE I  
Idealized Summary of Transfer Characteristics

<i>Transfer of learning</i>	<i>Transfer of training/basic learning</i>
Applying learning across domains/disciplines	Applying learning within a domain/discipline
Based on conceptual “elements”	Based on identical elements
Conceptual/underlying principle based	Concrete based
Context independent	Context dependent
Breadth and depth of scope	Narrow depth and scope
Generalization	Task specific
Figurative language/metaphorical/analogical	Literal language/domain examples
Larger and general knowledge base	Restricted to task knowledge base
Theory based	Immediate information based
Process oriented	Product oriented
Dispositional based	Instrumental/task based
Higher order/progressive transfer	Single-level transfer
Far transfer	Near transfer

### 4.1. Formal Discipline Model

This model developed out of ancient and classical educational theory, in which it was thought that the internal logic or structure of a discipline such as geometry, mathematics, or Latin somehow enabled a learner to transfer automatically what was learned in these disciplines to other areas. For example, it was thought that courses in formal logic taught people how to think in everyday situations. From early on, little rigorous research supported the formal discipline approach to transfer. Currently, with some interesting exceptions, the formal discipline approach is considered antiquated. However, at least as early as 1937, Allport, the renowned psychologist, suggested that the formal discipline model should not be dismissed entirely, and later, Ellis also suggested that the formal discipline model may not be a dead issue but simply in need of reformulation. Recent research suggests that the formal discipline model may have validity in some domains (e.g., statistical reasoning).

### 4.2. Behavioristic Models

#### 4.2.1. Identical Elements Model

The identical elements model, developed from the classic laboratory research of Thorndike and Woodworth in 1901, maintains that for transfer to occur there must be a common set of concrete identical elements between two experiences or learning sets. The identical elements model research resulted in the initial demise of the formal discipline model of transfer. Thorndike's negative findings of transfer outcomes led to a pessimistic view of the possibility of teaching for transfer and has exerted wide-ranging influence over instructional and curriculum design until present day.

A neo-identical elements model remains extant in some areas of cognitive psychology research. For example, what are called production rules (essentially "if . . . then" statements) in computational models of cognition are considered higher level versions of the identical elements stimulus-response connections of Thorndike. Much of the research on analogical reasoning (transfer) is also based on identical element relations of some kind, either concrete features or abstract relations.

#### 4.2.2. Stimulus Generalization Model

Almost from its inception, transfer research has been associated with the concept of stimulus generalization

developed from the behaviorist laboratory research of Pavlov's classical conditioning paradigm. Stimulus generalization is the evocation of a nonreinforced response to a stimulus that is very similar to an original conditioned stimulus. The stimulus generalization model presupposes identical elements that enable generalization to occur. Stimulus generalization, then, can be viewed as a basic physiological learning "theory" explanation of how transfer occurs. Despite this historical association with generalization, virtually no recent empirical or theoretical work has been conducted.

### 4.3. Cognitive Models

#### 4.3.1. General Principle Model

Just as Thorndike challenged the formal discipline model, the classic research of Judd in 1908 challenged the identical elements model by findings that suggested transfer also occurs on the basis of learners' understanding the abstract general principle(s) underlying phenomena that then can be applied to situations that do not possess obvious identical elements but that have the same or similar underlying principle. Judd's model can be considered a precursor of some current cognitive approaches to transfer because the general principle model is concerned with what is in learners' heads, not just the featural similarity between stimuli or within instructional material. It also can be considered a precursor to the metacognition approach because it recognizes that transfer can be promoted by strategies such as directing a subject's "attention" to the general principles involved.

#### 4.3.2. Information Processing Model

Historically, the concept of transfer of learning lay squarely within the applied instructional arena of education. Accordingly, until recent decades, cognitive psychology developed no theory of transfer per se. Currently, transfer as a function figures prominently in many areas of cognitive research. With few exceptions, this research has not been concerned primarily with instructional processes but rather with laying the foundation for an information processing framework that explains the internal representation, processing, and retention of information. One central area in contemporary cognitive psychology directly relating to transfer is schema research. As Phye notes, the information processing model tends to emphasize learner performance, not learner acquisition of knowledge (i.e., learning).

### 4.3.3. *The Metacognition Model*

This model is concerned primarily with promoting transfer, and not explaining its cognitive dynamics. The metacognition model makes use of self-monitoring strategies, leading learners to transfer these strategies within and across tasks or learning domains. Strategies include learners' self-monitoring to observe what they are doing or thinking while they are learning. Metacognitive research is viewed by many as the "new approach" to transfer. Strictly speaking, the metacognitive approach is not a part of cognitive psychology research but is more aligned with education.

## 5. HISTORICAL OVERVIEW

As noted previously, experimental research on transfer began early in the 20th century with psychologists Thorndike and Woodworth (1901) and Judd (1908); the first two are identified with the identical elements model of transfer and the latter with the general principle model. These two opposing models can be seen as progenitors that continue to frame much of the modern research agenda on transfer.

More than 100 years of transfer history shows that until the past few decades, (i) the identical elements model overwhelmingly predominated educational research and instruction and remains a strong influence, although to a less significant degree; (ii) in textbooks, transfer was given only cursory mention; (iii) the educational field had ceased to generate useful transfer findings or applications; (iv) transfer chapters largely had been missing from the pages of educational textbooks; (v) there was no comprehensive educational text on transfer; (vi) the concept of transfer had been included in neither cognitive psychology research nor cognitive psychology texts; and (vii) there had been an almost complete absence of conceptual and theoretical development of transfer as an educational construct.

### 5.1. Research Publications

According to Clark and Voogel's review of the educational literature, from the 1960s to the mid-1980s, research on transfer per se seemingly declined. Why this decline occurred had partly to do with the recognition that the identical elements model of transfer had ceased to be further developed and a shift in instructional focus to specific skills and competencies.

This apparent decline was misleading, however, because other fields had begun to reconceptualize transfer and to relabel it. For example, basic cognitive schemata have come to be seen by some as involving transfer. Because of an increasing recognition of the cognitive operations involving similarity relations, research on transfer has been extended to include the growing cognitive research on metaphorical and analogical transfer as well as what are called isomorphic relations, in which stories of identical structure but different content are used to test for transfer. Other research includes the cognitive function of using a variety of examples in instructing for transfer. Because cognitive research tended to take an information processing approach, which focused on how knowledge is presented and processed rather than on how knowledge is acquired in the educational or learning sense, the research was not concerned with the educational and instructional aspects of knowledge acquisition and transfer.

Historically, numerous edited books compiling articles on transfer were available, but there was no comprehensive nonedited volume. It was not until 1965 that the partially edited book by Ellis was published. It was the first attempt at a comprehensive integration of transfer research. In 1989, Singley and Anderson published a specialized nonedited computational approach to understanding transfer of skills. These books were concerned only marginally with instruction for transfer. In 1995, McKeough *et al.*'s edited volume was focused specifically on instruction for transfer based on the current research. It was not until 2000, however, that the first nonedited comprehensive overview of the research from an educational and cognitive perspective, with an integrated framework on instruction and theory, was published by Haskell.

### 5.2. Conceptual and Instructional Shifts

With few exceptions, most of the research attempting to discover the underlying mechanisms of transfer had been in cognitive psychology and not in educational research. Thus, the term transfer of learning, the original superordinate and applied educational construct, had been in danger of becoming a fragmented and subordinate process subsumed under the cognitive research labels outlined previously. Although it was repeatedly noted that educators should be exhorted simply to "teach for transfer," it was neither sufficient nor effective instructionally. Until recent decades, as Royer notes, with few exceptions

the educational field did not significantly adapt cognitive research findings to instruction.

In the early 1980s, educational researchers began to be aware of and to adapt the growing cognitive research related to transfer. This shift to a cognitive perspective is thought by some to have had a double edge, however. On the positive side, it shifted emphasis from a cognitively uninformed instructional view of transfer to a more researched-based construct, including the cognitive research on learning strategies that are thought to promote transfer. On the negative side, pure research into the cognitive mechanisms underlying transfer tended to overshadow the applied instructional approach to transfer by focusing on understanding the cognitive mechanisms of transfer. With the exception of the metacognition approach, this shift of focus from transfer as an applied instructional concept to one emphasizing a cognitive information processing approach constituted a kind of paradigm shift that is still quite evident.

## **6. TRANSFER THINKING AND REASONING**

In contrast to the traditional view of thinking and reasoning, it is becoming increasingly accepted that most everyday reasoning does not involve the manipulation of abstract rules or otherwise symbolic systems resembling a formal logic. Rather, it seems to involve pragmatic reasoning schemata, mental models, and analogical reasoning; hence, it involves transfer. Much research has demonstrated the pervasiveness and importance of analogical transfer in thinking and reasoning. The considerable research on similarity clearly demonstrates its systemic role in everyday reasoning processes. Researchers of analogical and metaphorical reasoning suggest that the perception and manipulation of similarity is the foundation of thinking and reasoning. Indeed, both Plato and Aristotle recognized this.

### **6.1. Single-Instance Reasoning and Analogical Transfer**

Research on the use of single instances in reasoning and decision making shows that people often place as much, if not more, credence in a single instance than in theory or logic. It is increasingly accepted that much of human reasoning is carried out by using schemata from everyday life as analogs that are tied to particular bodies of knowledge.

### **6.2. Legal Reasoning and Transfer**

The use of single-instance reasoning has implications well beyond its use by the layperson. A variant on single-instance and analogical reasoning is legal reasoning. The finding of similarity is the key ingredient in legal reasoning. Thus, much of legal reasoning is based on transfer. As the history of legal reasoning shows, it is based largely on reasoning from a single or a few instances or cases. Attorneys and judges often argue whether some previous case is applicable to a current case. Even reasoning on the basis of legal principle involves transfer reasoning; judges often decide cases "on principle," demanding that like cases should be treated alike.

### **6.3. Social Policy and Decision Making**

A number of studies on analogical transfer have noted that historians and governmental policymakers often make use of single instances in their experience as analogs on which to base decisions and policies about current crises. For example, in 1950, North Korea invaded South Korea. Although South Korea was thought to be of no significant strategic value, President Harry S. Truman was intrigued by certain historical incidents that had preceded World War II that he considered analogous to the North Korean invasion, particularly the Japanese invasion of Manchuria, the Italian attack on Ethiopia, and the annexation of Austria by Germany. Truman saw that the invasion of South Korea was just the beginning of a series of hostile acts like those that led to World War II, just as President George Bush, Sr., saw Saddam Hussein's invasion of Kuwait as reminiscent of Hitler's actions.

### **6.4. The Validity of Transfer Thinking**

Although widely and inescapably used in reasoning, transfer thinking should be approached cautiously. For centuries, philosophers have pointed out the dangers of reasoning based on similarity or analogy. The problem engendered by such reasoning can be seen as subject to the same problems as logical inference and induction. Recently, however, improved methods for assessing analogical transfer have been developed that apply to transfer thinking and reasoning.

## 7. HIGHER ORDER TRANSFER

Most applications of transfer are, in effect, analogical in nature, with some becoming quite complex in terms of abstract relationships and proportionality ratios. Transfer thinking is involved in even higher order thought sequences.

### 7.1. Transposition Transfer

Another phenomenon that can be seen as involving transfer is called transposition. Perceptually, for example, when animals conditioned to respond only to the darker of two gray squares are later shown an even darker gray square along with the original darker square, the animals respond to the new darkest gray square to which they had not been conditioned. This response is generally considered to occur because in relation to the first set of squares, the new darker square is perceived as the same as the original darker square. The transfer was not to the absolute specific shading, but the relative shading of the new darker square in relation to the shading of the original darker square. Notated analogically, the relation is as follows: the first darker square : the first lighter square :: the new darker square : the original darker square. Auditorially, transposition transfer can be illustrated by recognizing a musical piece played in a different key or octave as being the same even though no note is the same.

### 7.2. Progressive Transfer

Modeled on analogical and arithmetical progression, Haskell suggested a type of transfer not generally recognized that can be referred to as progressive or continuous transfer. This type of transfer takes the abstract form of arithmetical progression: 2, 4, 8, 16, 32. Stated in analogical form, it is  $2 : 4 :: 4 : 8 :: 8 : 16 :: 16 : 32$ . As a semantic example, the biological classification system can be put into this progressive form as follows: species : genus :: genus : family :: family : order :: order : class, etc. Related to analogical reasoning, Aristotle referred to this form of reasoning as continuous analogy.

### 7.3. Transfer in Science

The history of science and invention is replete with those who are skilled at transfer. For example, Darwin,

who developed human evolutionary theory, transferred the idea of farmers' selective breeding (i.e., artificial) of animals to his development of the principle of natural selection, in which natural processes were selecting genetic traits in place of the farmer. Physicist Louis De Broglie noticed that the mathematical equations of another well-known physicist, Neils Bohr, who described the orbits of an electron, were the same equations used to describe the vibrating waves of a violin string. With this transfer, De Broglie revolutionized atomic physics and laid the foundations of quantum mechanics.

As these two examples show, many advances in science are made on the basis of what, after the fact, seems to be a simple "it's like . . ." type of transfer. These examples are included in this section because theoretical thinking and a complex knowledge base were required before the higher order of transfers could be made.

## 8. INSTRUCTIONAL ISSUES

Although all educational issues have an impact on transfer, arguably there are relatively transfer-specific ones that stand out in terms of achieving transfer, including the counterintuitive nature of instructing for transfer, learning strategies, knowledge base, contexts, practice, and variability.

### 8.1. Counterintuitive Nature of Transfer

A generally unrecognized problem in understanding, achieving, and instructing for transfer is that doing so is often counterintuitive to what might be expected from instructing general learning. For example, it is generally assumed that those with a high level of theoretical knowledge of a subject are more proficient at transfer than people with only concrete knowledge. Although this is typically true, there are important counterintuitive exceptions.

For example, students who use a deep learning or conceptual approach to material, as expected, tend to score higher on more complex recognition tasks but, unexpectedly, score lower on simple recognition tasks. Moreover, students who lack prior knowledge about a skill tend to do better than those who have prior knowledge of the skill—but only under rote learning conditions. Students who learn reading material in a

verbatim or rote fashion do better than students who learn the material in a conceptual fashion on verbatim tests, but students who learn the material conceptually are better at applying their knowledge in a more generalized way.

The point is that although conceptual or theoretical knowledge is advantageous for engaging in far transfer, it may be disadvantageous for simple concrete application situations. Such counterintuitive transfer outcomes across other subject areas yield similar findings, suggesting that if the goal of instruction is transfer of explicitly taught skills, then learning instrumental and concrete procedures may be more effective than conceptual learning. However, if the goal of instruction is far transfer, then learning conceptually is superior to simple instrumental learning.

Counterintuitive findings also apply to conditions of practicing a skill. For example, Schmidt and Bjork found that providing information to learners about their performance errors during practice improves performance during training but can degrade performance on a test of long-term retention or transfer. Furthermore, increasing the task variability during practice decreases performance during training but facilitates performance on later tests of the ability to generalize training to new conditions.

## **8.2. Contexts and Cultures of Transfer**

Typically, learning has been thought of as the inherent property of the individual learner. However, it is increasingly recognized that all learning takes place in a context and that the context typically becomes a part of, and associated with, the learning and its transfer. Broadly construed, the concept of context includes not only the immediate learning situation but also sociocultural influences. More narrowly, context differs from cultural influences in that context tends to be associated with the learning situation. What Pea has called cultures of transfer, however, are networks of social norms that may either enhance or inhibit transfer.

## **8.3. Knowledge Base, Learning Strategies, Practice, Theory, and Exemplars**

After a long period of concentrating on a content-empty learning strategies approach to instruction, it

is increasingly recognized that content-specific knowledge (or declarative knowledge) is crucial for transfer. An appropriately encoded knowledge base results in expertise, and expertise is associated with superior transfer. Theoretical knowledge is also central to achieving transfer. This is true even for small children. Historically, it has been thought that young children are “perceptually bound” to the concrete features of stimuli and are unable to transfer causal relations appropriately. However, the seminal work of Brown demonstrated that young children can engage in transfer generated from a “deep structure” level of conceptual understanding even when similar surface elements between tasks contraindicate transfer.

In addition, with some exceptions, research strongly suggests that reflective practice facilitates significant transfer. The research also shows that, with only minor exceptions, one of the most important means of facilitating transfer is with the use of a variety of related examples. It is thought that processing multiple and varied examples results in the development of cognitive schemata.

After reviewing the research on transfer, Haskell suggests that for significant learning and transfer to occur, the following principles are required: Learners must (i) acquire a large primary knowledge base in the area in which transfer is required, (ii) acquire some level of knowledge base on subjects outside the primary area, (iii) understand what transfer of learning is and how it works, (iv) understand the history of the area(s) in which the transfer is wanted, (v) acquire dispositional characteristics, (vi) develop an orientation to think and encode learning in transfer terms, (vii) create contexts and cultures that support transfer, (viii) understand the theory underlying the area(s) they want to transfer, (ix) engage in hours of practice and drill, (x) provide time for the learning to incubate, and (xi) observe and read the works of those who are exemplars of transfer thinking.

## **9. DISPOSITIONAL CHARACTERISTICS AND ASPECTS OF TRANSFER**

Dispositional aspects of transfer relate to personality and positive attitudinal characteristics toward learning, as Bereiter suggests. Strictly speaking, the dispositional view can be considered another model of transfer.

### 9.1. Personality Structure and Development

Although not widely recognized or researched, as early as 1937 Allport suggested that transfer may be important for the theoretical psychology of personality. Essentially, the idea is that for most people, the complex of behaviors and ideas that make up who they are remain unorganized, unrelated, and dissociated. Presumably, the ability at transfer may be used to integrate this complex of traits. Ausubel suggested that there exists two broad cognitive styles, generalizers and particularizers, with the former engaging in more transfer than the latter.

### 9.2. Disposition for Learning

A major issue in how best to achieve significant transfer is whether to instruct primarily for specific skills or to instill in a learner a personal disposition toward learning and transfer. The former involves acquiring abstract learning strategies, including metacognitive skills; the latter involves instilling motivational and affective values or a positive temperament for learning and transfer. As judged from the direction of most research, the general consensus seems to be with the former.

### 9.3. Prejudice

Transfer is also involved in both the development and the reduction of prejudice toward others. Prejudice can be seen as inappropriate or negative transfer of perceptions and experiences regarding others. Assuming positive transfer, this involves a person's inferring that his or her experience is similar in some degree to another person's experience, thus leading to better understanding of others. Accordingly, analogical transfers of experience can be useful for both understanding prejudice and understanding those with differences.

## 10. BIOLOGICAL ORIGINS OF TRANSFER

A possibility that has been largely overlooked is that transfer may have evolutionary origins and a neurological basis. For a phenomenon as pervasive and significant as transfer, it might be expected that it would have biological origins. Although the research is inconclusive, Haskell suggests that transfer has both evolutionary and neurological bases.

### 10.1. Evolutionary and Neurological

The evolutionary argument is that being attuned to even superficial or surface similarities may have bestowed a survival advantage on animals in complex environments. Surface similarities are swift evaluations that are often needed to avoid danger. In evolutionary terms, if it looks like a hungry lion and growls like a hungry lion, then there are survival advantages to assuming it is a hungry lion. In other words, it may have had more survival value to assume that a surface similarity is meaningful than to assume it is not. For more complex thinking, it is thought that surface similarity is often positively correlated with deeper, more "causal" or structural relations among phenomena. Evolutionarily, the consequences of an invalid surface similarity typically are not as serious as ignoring a valid one. Accordingly, a neurological substrate likely evolved. Haskell suggests findings indicating neurological circuits that specialize in the generalization of learned experience to new situations via invariance-generating circuits.

### 10.2. Transfer and Intelligence

Perhaps the most controversial issue regarding significant or far transfer is its relationship to innate intelligence, as measured by IQ tests. Historically, transfer has been thought to be correlated with IQ mainly because the lower the IQ of a learner, the less transfer is typically achieved. Ability at analogical transfer, as measured by the Miller Analogy Test, has long been correlated with IQ and, some suggest, with creativity. However the issue resolves, it is clear that IQ, although seemingly a necessary condition, is not a sufficient condition for transfer.

## 11. FUTURE EMPIRICAL RESEARCH AND THEORY

Future research on transfer will include most of the areas mentioned earlier, with specific attention to (i) how to instruct effectively for transfer; (ii) understanding how contexts influence transfer; (iii) individual learner characteristics; (iv) establishing a more precise taxonomy or classification system; (v) specific research on the various kinds of transfer, especially progressive; (vi) delineating the concept as a general cognitive function; and (vii) identifying the neurological substrates of transfer. Finally, an emphasis on theory construction is needed to guide empirical research and the development of the concept of transfer.

### See Also the Following Articles

Educational Achievement and Culture ■ Learning  
 ■ Learning Styles and Approaches to Studying ■ Teaching Effectiveness

### Further Reading

- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637.
- Bereiter, C. (1995). A dispositional view of transfer. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp. 21–34). Mahwah, NJ: Erlbaum.
- Brown, A. L., & Campione, J. C. (1984). Three faces of transfer: Implications to early competence, individual differences, and instruction. *Advances in Developmental Psychology*, 3, 143–192.
- Clark, R. E., & Voogel, A. (1985). Transfer of training principles for instructional design. *Education Communication and Technology*, 33(2), 113–123.
- Cormier, S. M., & Hagman, J. D. (Eds.). (1987). *Transfer of learning contemporary research and application*. New York: Academic Press.
- Detterman, D. K., & Sternberg, R. J. (Eds.). (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex.
- Gladstone, R. (1989). Teaching for transfer versus formal discipline. *American Psychologist*, 44(8), 1159.
- Haskell, R. E. (1998). *Reengineering corporate training: Intellectual capital and transfer of learning*. Westport, CT: Quorum (Greenwood Publishing Group).
- Haskell, R. E. (2000). *Transfer of learning: Cognition, instruction, and reasoning*. San Diego: Academic Press.
- Haskell, R. E. (2002). Cognitive science and the origin of lexical metaphor: A neurofunctional shift (NFS) hypothesis. *Theoria et Historia Scientiarum*, 6, 291–326.
- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition*, 15(4), 332–340.
- Pea, R. D. (1987). Socializing the knowledge transfer problem. *International Journal of Educational Research*, 11, 639–663.
- Phye, G. D. (2001). Problem-solving instruction and problem-solving transfer: The correspondence issue. *Journal of Educational Psychology*, 93(3), 571–578.
- Read, S. J., & Cesa, I. L. (1991). This reminds me of the time when . . . : Expectation failures in reminding and explanation. *Journal of Experimental Social Psychology*, 27, 1–25.
- Resnick, L. B. (Ed.). (1989). *Knowing, learning and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207–217.
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology*, 62(6), 913–933.
- Sunstein, C. R. (1993). On analogical reasoning. *Harvard Law Review*, 106, 741–791.
- Vogniadou, S., & Ortony, A. (Eds.). (1989). *Similarity and analogical reasoning*. Hillsdale, NJ: Erlbaum.



# Translation

Richard W. Brislin, Brent MacNab, and David Bechtold

University of Hawaii, Honolulu, Hawaii, USA

1. Translation and Cross-Cultural Research
2. Back-Translation
3. Pretesting
4. Emics and Etics
5. Decentering
6. Writing Translatable English
7. Recent Enhancements of Translation Methodology in Research
8. Translation and Qualitative Studies
9. New Challenges in Translation: The Pursuit of Equivalency in Higher Order Issues
10. Future Directions of Translation and Translation Models
11. Conclusion  
Further Reading

## GLOSSARY

**asymmetrical translations** A literal translation of a research instrument that is used to identify differences between cultures.

**back-translation** Using independent bilingual translators, the act of translating material from host language to target language and then translating back into host source to determine if equivalence of meaning is achieved.

**cultural emic** A concept, value, or other artifact of culture that has particular meaning to a specific culture (e.g., dating).

**cultural etic** A concept, value, or other artifact of culture that has universal meaning across all cultures (e.g., love).

**culture shock** A sense of loss that accompanies leaving one's own culture and adjusting to another.

**decentering** A method of translation–back translation that allows for the development of a research instrument that is of equivalent meaning to both cultures under study.

**host version** The original language used in the version of material to be translated.

**item response theory** A method of comparing responses from different translated versions of questionnaires to determine if there is equivalence.

**pretesting** The use of a translated version in a pilot test to determine if the instrument and response are as expected and equivalent to the instrument and responses in the original version of the instrument.

**symmetrical translations** A conceptual translation of a research instrument that is used to identify similarities across cultures.

**synchronous automated translation systems (SATS)** Technology that can produce seamless translations with minimal time delay.

**target version** The language used in which the original version of the material is translated into.

Translation to achieve equivalence of meaning has become a significant issue for international and cross-cultural research. A failure to achieve equivalence in translation can put into question the research generally and conclusions derived from that research specifically. Beginning in the 1970s, efforts began to enhance the quality of translation in research through the use of various methods to achieve equivalence. The use of back-translation and decentering techniques were the first attempts at achieving meaning equivalence in translation. Recently, the use of dual-focus methods in

questionnaire development as well as quantitative strategies such as item response theory have been developed to further enhance translation effectiveness. Computer and technology are also becoming important tools in translation, with the development of software such as synchronous automated translation systems. These new tools are expected to enhance both speed and accuracy of translating materials from one language to another.

## 1. TRANSLATION AND CROSS-CULTURAL RESEARCH

In almost all cross-cultural research studies, methodological issues associated with translation between languages are certain to arise. Researchers have tests, questionnaires, and interview schedules, but they also have experimental protocols so that research participants know what is expected as they attempt to complete various tasks and activities. Translation issues are clearest when languages from very different linguistic families are involved, such as English and Japanese or Swedish and Chinese. However, it is also wise for researchers to know about translation issues even when they deal with dialects, such as English spoken in the United States compared with that spoken in Scotland or French spoken in France compared with that spoken in Canada.

Many early cross-cultural studies (i.e., during the 1950s and 1960s) have standardized tests as their starting point. Well-known tests, such as the Minnesota Multiphasic Personality Inventory or the Stanford-Binet Intelligence Test, were translated into other languages, administered in various countries, and interpretations made based on patterns of differences. Many of these studies are embarrassing by today's methodological standard and have been relegated to the deserved obscurity of dust-gathering journals on library shelves. Prior to 1970, there were few attempts to assess the quality of translations. Bilinguals were presented with a standardized test in one of the languages that they knew and wrote a version in their other language. Few researchers asked about issues such as the quality of the translation, equivalence of terms, familiarity of the words in the translated version, and the level of literacy needed among research participants to give meaningful responses to test items. The lack of attention to translation quality led to the publication of many problematic studies that are rarely cited today.

Even with these difficulties, lessons were learned that have proven useful in the development of cross-cultural psychology. One important lesson is that translation between languages is not a simple and direct process in which researchers can present wording in one language and expect straightforward equivalents in another language. Researchers in the 1950s and 1960s often expected such direct equivalence and had difficulties when they did not find it. Instead of direct equivalence, they found that familiar terms in their own language were difficult to translate. The importance of "punctuality" for workers was difficult to translate into languages where cultural practices do not include organizing the workday around clocks. "Friend" in American English was difficult to translate into languages where close nonfamily members are viewed as deserving of long-term obligations and collective memberships. Americans often use "friend" to describe a person they met and chatted with yesterday.

The realization that frequent nonequivalence is a common occurrence in cross-cultural research led researchers to deal with this fact in creative ways that ultimately advanced the quality and relevance of cross-cultural studies. For example, it led to the development of procedures such as the antecedent-consequent method, in which the meaning of terms such as friend could be probed. Questions of respondents in different cultures would take forms such as the following:

(Antecedent)

If you have people who \_\_\_\_, then they are friends.

If people \_\_\_\_, then they may become friends.

(Consequent)

If people are friends, they are expected \_\_\_\_.

If a person has problems, a friend will \_\_\_\_.

Responses to such inquiries will give insights into such issues as the length of time people know each other before becoming friends, activities in which friends engage, and the obligations that friendship entails. This information that surrounds a term such as friend, with rich context, is extremely useful for translators who must search other languages for the best equivalent term or phrase.

This realization concerning translation and change in attitude paralleled another change in cross-cultural studies that occurred at approximately the same time and also provided an additional term for translation analysis. Culture shock was a term first coined in 1958 by Kalvero Oberg. So persuasive did the term become that training professionals whose jobs included preparing people for life in other cultures quickly

incorporated it into their training programs. The term refers to the sense of loss and corresponding frustrations, emotional upheavals, and physiological changes that accompany leaving one's own culture and adjusting to another. Beginning in the 1960s, there was a large group of highly educated and articulate people who learned about culture shock in cross-cultural training programs. These were Peace Corps volunteers. Culture shock was at first treated as a problematic state that might mark unsuccessful volunteers. It was therefore introduced as an indicator of adjustment problems. This treatment of the term changed when the first waves of volunteers returned after their 2-year overseas assignments. Most had experienced culture shock but also felt that their assignments had been successful. Because most experienced culture shock, it was difficult to view it as a sign of maladjustment. The view of culture shock changed and (beginning in the mid- to late 1960s) it became viewed as an expected part of any overseas assignment. Translations of the term culture shock into other languages had to include this changed connotation.

## 2. BACK-TRANSLATION

Working to improve the quality of translation for research purposes, Werner and Campbell published an influential analysis of back-translation in 1970. With this method, monolingual researchers can gain some, although not perfect, insights into the quality of the translation by working carefully with bilingual translators. Back-translation is based on a set of steps. For ease of discussion, examples of translations from English are used (Fig. 1).

Although not perfect, this back-translation procedure provides a good starting point. The researcher can examine the two English language versions and if they are equivalent and if the back-translation is worded smoothly, then he or she is at "step one" of working

toward a high-quality translation. Researchers should work closely with translators and treat them like respected colleagues. Researchers can ask the first bilingual, "Was there any English language material that was difficult to translate?" They can ask the second bilingual, "Was there any material in the target language that you had difficulty with, for example, finding equivalents in English." If translators know details about the goals of a research project, they can be extremely helpful in suggesting alternative wordings and sometimes indicating ideas for new research directions. For example, assume that the original English wording asks about the respondents' relationships with their brothers. A Japanese bilingual would point out that there is no equivalent for brother. In Japanese, native speakers must make a distinction between "older brother" and "younger brother." There are distinct terms based on age distinction but no general term for brother. If not integrated into the research team, a translator might choose one of the terms as a "default option," but this could interfere with the researchers' goals. This feedback from the translator might stimulate the researchers to examine the importance of status distinctions in Japan. They would find, for instance, that age distinctions are taken seriously even at a 4-year college. Sophomores take on a *sempai* role, or the role of superior and guide, to freshman *kohai*, or neophytes.

Why is back-translation only a good first step? One reason is that certain words have near-automatic translation equivalents, but the terms may not be conceptually equivalent, as indicated by a more detailed analysis such as the antecedent-consequent method. The English "friend" will translate to "amigo," and this will translate back to the English "friend" with near certainty. However, the types of people that respondents call friend and amigo may be very different. Similarly, the English "party" may translate into the French "soiree," and this will translate back into English as "party," but are the same types of social gatherings being discussed?

Researcher starts with newly developed materials or with a standardized instrument from past research →	A first bilingual translates this into the target language  →	A second bilingual independently translates this version back into English →	The researcher can then examine the two English versions and can begin to make assumptions about translation quality
--	---	---	--

FIGURE 1 Translation quality: translation and back-translation procedure.

Another reason for back-translation's limitations is that the second bilingual, if skillful, can make sense out of a poorly worded translation and then render it into very fluent English. The researcher will read the back-translated English and might conclude that the target version must be good given that the second English version reads so well. The activities of the second bilingual are similar to those done by many readers of this article. Haven't many of us been in social situations in which we tried to help international visitors or international students? They spoke to us in broken English, but we were able to figure out what they wanted and could express their wishes in fluent English: "You want the address of a good hotel that handles foreign currencies" Or "You need to speak to the international student adviser to make sure that your visa is legitimate and that it will allow you to enroll for courses."

### 3. PRETESTING

Given that a good back-translation suggests but does not guarantee a good translated version, other methodological steps should be added. One is pretesting. The translated version should be administered to approximately 30 monolinguals, who then answer the questions, complete the standardized test, or go through the experimental procedures, depending on the nature of the research instrument. These pretest participants should be similar to the eventual respondents in the proposed final phases of the research. Pretest participants should be similar in terms of age, gender distribution, amount of formal education, status within their communities, and so forth. With 30 participants, internal consistency reliability estimates can be obtained for questionnaires and tests. Participants can be observed going through the necessary steps in an experiment and adjustments can be made where they have difficulties. With the help of research assistants (perhaps the translators, who should be integrated as members of the research team), participants can be asked what wordings were clear and unclear, what procedures they found difficult, and so forth.

### 4. EMICS AND ETICS

In any large-scale research project, some original language material will not translate well. This is to be expected, and in many ways researchers should look

forward to the challenge of interpreting difficult-to-translate material. As research efforts become more global, the cross-cultural aspects of translation have become even more critically important in developing research of the same, multilinguistic nature. Cross-cultural researchers have laid out important groundwork in demonstrating both potential pitfalls and proposing guidance for developing research across country borders. The idea of universal concepts (etic) or culture-specific concepts (emic) holds particular importance in translation methodology. Any cross-cultural or international study that does not consider the etic/emic issue when developing research instruments or conducting research will create studies of limited validity and value.

Etic approaches to research are studies conducted cross-culturally in order to extract common elements. Hofstede's well-known IBM study is a prime example of an etic approach to cross-cultural research. Etic research has the potential for reaching much broader, and culturally sweeping, conclusions that are often beyond the scope of only one cultural landscape. Etic aspects of language represent ideas that are more universal, with semantics that are more readily translated between cultures.

Emic approaches to research are studies conducted within a cultural system based on data representative of only that cultural landscape. Emic (e.g., unique and regionally specific) constructs of language can create roadblocks to both the interpretation of research results and how clearly the research parameters and ideas are interpreted by the subjects. One reason for difficulties in translation is that a culture's emics, or specific manifestations of complex concepts, have not been clearly identified.

For example, the concept of friendship is universal (an etic), but there are specific aspects of friendship that are emic. In the United States, the term friend is used very loosely and includes recently met acquaintances. In China, the term is used to refer to a smaller number of people who have long-term mutual obligations to each other. The need to differentiate between "older brother" and "younger brother" in Japanese is another example of an emic. Another example can be found in the neo-Melanesian language, which has no word for "desert" because rainfall is plentiful in Melanesia and plants grow in most places. However, there are worn-away lands that are referred to as "nothing places" or "ples nating." In translations of the New Testament, John baptizes in such places rather than in the desert. Clear failures in translation and data

collection can often involve a misunderstanding of another culture's emics, which is a possibility in most phases of cross-cultural research. For this reason, emic manifestations of concepts should be documented and reported. In cross-cultural studies, the reporting often takes the form of what aspects of a concept are shared across cultures and what concepts are culture-specific. Researchers should always remember that it is the combination of etics and emics that advances our understanding of concepts among members of our own and other cultures.

One way of recognizing emic differences, and therefore enhancing the potential of the research in achieving translation and conceptual equivalence, is through the use of back-translation. Translation techniques generally and back-translation techniques in particular have assumed even more importance as researchers attempt to partial out those concepts that are etic in nature from those concepts that are emic.

## 5. DECENTERING

When cross-cultural researchers are sensitive to the emic/etic distinction, they are careful not to impose their own emics on other cultures. Many concepts in psychology (e.g., personality, intelligence, and worker motivation) reflect the culture in which they were developed and in which research on them took place. When one of the authors gave a lecture on personality to a group of Chinese scholars, they later commented that they found many of the specific terms (achievement orientation, need for power, openness to experience, etc.) difficult to understand since there were few readily available terms in Chinese. They mentioned that they have a much richer language to describe people's relations with others, which would follow from China's status as a collectivist culture. The key point is that personality terms in English and collectivist terms in Chinese may reflect the two culture's emics. To expect an easy translation of them could be a mistake.

When researchers use decentering, they make attempts to find concepts that are readily expressible in the languages under study. No original language version becomes the "center" that has to be translated into less familiar and sometimes clumsy target language terms. Rather, the original language version is open to change as each translated/back-translated version is reviewed. With each review, researchers reword the original until a smooth target language version (as

reported by translators or independent bilinguals) and back-translation version are developed. In the case of published standardized tests, this procedure demands that researchers be willing to revise the original wording just as if they were formulating an instrument for the first time. The assumption is that more meaningful information will be obtained from a decentered instrument than from a forced translation of English that has problematic equivalents in the target language.

## 6. WRITING TRANSLATABLE ENGLISH

If researchers prepare their own materials "from scratch," there are guidelines for writing English that translators may find relatively clear and understandable. These suggestions were developed by Brislin in 1970 as part of a research project on improving translation for cross-cultural research projects. Working with 92 translators, he found that they had a relatively easier time with their task if the original English was prepared according to the following suggestions:

1. Use short sentences of less than 16 words. With longer sentences, translators have difficulties disentangling what subjects go with what verbs.
2. Employ the active voice. Some languages do not have passive voice constructions, and the active voice often allows translators to more clearly identify subjects, verbs, and objects.
3. Repeat nouns instead of using pronouns. Some languages have more pronouns than there are in English, and so the translator may be forced to make choices from alternatives that the researcher does not know about.
4. Avoid metaphors and colloquialisms. Such phrases, part of the socialization of children in a specific culture or geographical areas within a large country, often do not have equivalents in other languages.
5. Avoid the subjunctive (i.e., phrases with "could," "would," and "should"). Some languages do not have easily available terms for the English subjunctive, and so the translator is forced to make approximations.
6. Add sentences to provide context. Expanding on the previous discussion of "friend," there might be a sentence indicating how long people have known each other. This suggestion sometimes leads to longer items, but the items become multisentence items given suggestion number 1.

7. Avoid adverbs telling “where” (e.g., beyond) or “when” (e.g., frequently). These terms often have inadequate direct equivalents and they force translators to make their own decisions on questions such as “how many times is necessary for a translation of a phrase such as ‘meeting frequently with my boss?’”

8. Avoid possessive forms when possible. English speakers are familiar with ownership and have extensive experience with its description. This is not the case in cultures in which private ownership of property and goods is not as extensive and in which collectives rather than individuals control ownership.

9. Use specific rather than general terms (e.g., “wheat” rather than “edible plants”). Since cultures categorize specific terms differently, there is no guarantee that a general term will have the same connotation in other languages.

10. Avoid words indicating vagueness regarding some event or thing (e.g., “probably,” “maybe,” and “perhaps”). They do not always have translation equivalents, and they force translators to make difficult choices among imperfect alternatives in the target language.

11. Use wording familiar to translators. If translators are familiar with English terms, they can indicate the presence or absence of equivalent terms in the target language. This suggestion reinforces the previous recommendation that translators be integrated into the research team.

12. Avoid sentences with two verbs if the verbs suggest different actions. Translators often have difficulty linking the appropriate subject with the appropriate verb.

## 7. RECENT ENHANCEMENTS OF TRANSLATION METHODOLOGY IN RESEARCH

During the past 15 years, there has been a significant effort to further develop and enhance translation effectiveness, and in doing so this effort has recognized that there are more elements to effective translation than the multistep process described previously. Most of this research and theory development has focused on four basic areas: the initial development of the questionnaire, the style of translation, the use of bilingual or monolingual translators, and the use of quantitative measures to establish item equivalence. A secondary area of theory development concerns the effectiveness of this methodology in qualitative research.

### 7.1. Questionnaire Development

Research has shown that the literal translation of a data-collection instrument from one culture to another can put in doubt the reliability of the instrument and the study. Of even greater concern is the question regarding the ethics of doing research, and drawing conclusions, from the focus of the dominant culture and not from the focus of the culture being studied. Decentering, as described previously, is one approach used to minimize this problem. However, a dual-focus approach to translation has also been offered as a possible solution to these concerns.

The dual-focus method uses a two-pronged approach for developing research instruments. Initially, and most important, a horizontal collaboration of researchers from the indigenous cultures that are being studied is established to develop the original research instrument. After the initial instrument is developed, researchers focus on capturing the concepts of the phenomena being measured in the instruments rather than simply focusing on literal translations. It has been offered that the dual-focus approach to instrument and measure development is superior to other translation methods because the approach is less time-consuming and overcomes many of the concerns regarding the validity of the decentering process. The dual-focus approach may also be better suited for multilingual studies compared to the decentering approach, which has been described as being extremely complex and cumbersome in these instances.

### 7.2. Translation Style

Although most current approaches to translation have shifted from literal to conceptual translation of meaning, there is a school of thought that advocates that the nature of the translation is in part due to the nature of research being undertaken. Translation can be either asymmetrical or symmetrical. Asymmetrical translations of instruments seek to identify differences between cultures and therefore must remain true to the original language in which the instrument was designed. A literal translation is required, which in turn can create a translated version that may appear unnatural and unusual to the target population being examined. On the other hand, symmetrical translations of instruments seek to identify similarities across cultures. Therefore, the instruments must be recognizable and colloquial in content, with items equally understandable to all the cultures within which the research is being conducted.

### 7.3. The Use of Monolingual and Bilingual Translators and Subjects

Brislin found that in judging the equivalence of translated material, monolingual or bilingual judges would be equally effective in the discovery of errors of meaning. However, further research has provided additional insight as to what to consider when using translators.

Translators who are familiar with semantic differences are a necessary requirement for cross-cultural studies that collect data from subjects in different regions with different dialects. Cultural perceptions that are held by the translator are also a significant aspect to the translation process, especially when the goal in translation is symmetrical rather than asymmetrical. It is therefore considered optimal to have the team of bilingual translators acquire skills in both languages at different times and in different locations. Expertise in the content of the instrument is also an important attribute to equivalence in translation. Bilingual translators who are familiar with the content of the research and how the instrument will be used develop more effective translations.

Perhaps the most extreme approach to issues of translation and conceptual equivalence is to recruit bilinguals to participate in the research study. Depending on the nature of the research and the type of study population to be examined, using bilingual subjects can enhance data collection quality, especially when the focus of the research is conceptual and comparative.

### 7.4. Quantitative Measures to Establish Equivalence

The use of translation and back-translation, as well as decentering techniques, is designed to enhance the equivalence of translated materials with original versions. Although effective, the use of decentering techniques becomes much more difficult as different language versions are developed. In response to this problem, certain quantitative techniques can be used after translations are made to further assess the accuracy of these translations. Item response theory provides a method of comparing responses from different translated versions of questionnaires to determine if there is equivalence. This quantitative approach is most effective using dichotomous data. When using polychotomous formats, simultaneous factor analysis can be used to determine equivalence. Pearson product moment correlation, median correlation coefficients,

as well as mean scores and standard deviations are also quantitative tools that can be used to determine equivalence between source language questionnaires and target language translations.

## 8. TRANSLATION AND QUALITATIVE STUDIES

Qualitative research literature has not extensively addressed the significance of equivalent translations given the nature of the research methodology. It is assumed that the face-to-face nature of the communication allows for in-depth probing and clarification of concepts and data between researcher and informant. For this reason, little emphasis has been placed on translation as a critical component of research methodology. However, translation equivalence is a critical concern when international qualitative data are being reduced into specific categories or themes or when subtle and complex phenomena are being studied.

When categorizing data, translation equivalence errors may not necessarily occur in the assignment of translated material into major categories and themes. Rather, these translation errors can frequently appear when the major categories are reduced into more refined and sensitive subcategories and subthemes. It is here that translation equivalence has a significant bearing upon the quality of the research.

Translating complex in-depth interviews from one language to another can create inaccuracies due to lack of word equivalence as well as difficulty in identifying interview filler so as to establish accurate drop rates. Attempting to replicate grammatical style between two grammatically different languages is not only a problem in quantitative research but also problematic in qualitative research. Difficulties increase in trying to understand not only what was said but also how one was saying it, especially when what is communicated is both complex and subtle.

Because interview data can be described in much more complex detail than a question in a questionnaire, the use of multiple, independent translators and back-translation may not be an effective approach for achieving translation equivalence. It has been recommended that when translating qualitative data a single, expert translator should be used. This will allow for consistencies in interpretations of language usage and nuances.

Aside from the translation of completed interviews from one language to another, an additional concern is the use of bilinguals in conducting research interviews. Although bilinguals may have the ability to probe and simultaneously translate for common meaning, there is concern that bilinguals will inconsistently draw upon both languages when conducting these interviews. English words or phrases, for example, are often interjected into foreign language conversation. The multiple meanings that these words or phrases may have to a bilingual interviewer compared to a monolingual interviewer may also put into doubt the validity of the data and data analysis.

Like quantitative research, qualitative research can be impacted adversely by lack of equivalence in translation. Although these problems may impact different aspects of the research, they still need to be considered and addressed both in the planning and in the execution of the research study. It is also important to be aware that because problems of translation may impact different aspects of qualitative research, the use of standard translation methodologies in quantitative studies may not be appropriate.

## **9. NEW CHALLENGES IN TRANSLATION: THE PURSUIT OF EQUIVALENCY IN HIGHER ORDER ISSUES**

During the past two decades, the need for better and higher quality translations of questionnaires and other data-collection instruments has become increasingly important to cross-cultural and international researchers. As more attention is focused on the translation methodology and some issues of conceptual equivalency, new problems occur. These problems focus on the more subtle, higher order aspects of translation and conceptual equivalency.

### **9.1. Nonverbal and Other Special Translation Issues**

The importance of nonverbal cues in the overall communication process is well documented. Some estimates indicate that more than 80% of receiver information is acquired through nonverbal cues. Pitch, tempo, loudness, hesitation, facial cues, hands and hand gestures, timing, silence, symbols, and even dress can add meaning to a translated communication.

These nonverbal cues are even more important when there is significant divergence between the host culture and targeted culture during translation. Research has supported the notion that in-group advantages for cultural members are present for correctly interpreting and decoding nonverbal communication and emotion recognition. This in turn suggests that out-group members or expatriates may encounter additional challenges within these more nuanced areas of translation.

### **9.2. Professional Jargon, Metaphors, and Idioms**

Clearly, in-group/out-group misinterpretations can also occur outside of nonverbal communications. Professional jargon, idioms (a phrase or expression having a different meaning from the literal), and slang are all types of communication that require careful consideration and more than just a literal translation in order to achieve equivalence in meaning between two languages. Strategies that can be used to consider these special cases include finding ways to neutralize the text; attempting to find an exact equivalent of the original jargon, metaphor, or idiom; selecting other jargon, idioms, or metaphors that would express a similar sense; or replacing the jargon, idiom, or metaphor with an approximate literal phrase.

## **10. FUTURE DIRECTIONS OF TRANSLATION AND TRANSLATION MODELS**

The 21st century will present new challenges to achieve effective translations as the world becomes smaller and more reliant on technology. The opportunities for faster and more insightful translation of data instruments and research data will also create new problems and challenges for researchers, particularly as researchers shift their attention from identifying universal etics to recognizing culturally specific emics.

### **10.1. The Impact of Globalization**

Language can be viewed as one of the more important gateways to culture that has challenged human interaction since our earliest history, as exemplified through accounts such as the *Tower of Babel*. As an integral and prominent artifact of culture, language, and the process of understanding known as semantics,

plays a critical role in a world community that has been moving toward globalization. Recent trends such as the globalization of business and trade, worldwide communication networks, new technology, improved transportation, and geopolitical economic unions will continue to increase the incident of foreign language contact and thus the need for effective, efficient translation. These needs also drive academia toward new studies of language, semantics, and translation.

## 10.2. Technology and Translation

In developing a forward-looking view of translation, any treatment would be remiss by not discussing the use of technology. New technology, including fuzzy logic, neural networks, and parallel processing (i.e., artificial intelligence), is developing in ways that will radically alter the reality of translation as we know it today. The task of holistic, global translation, taken from an individual's perspective, is indeed daunting. There are 3000–6000 languages globally (depending on the source cited), with approximately 200 of these being spoken by 1 million people or more and approximately 100 serving as official, national languages. Given that even the best human translators are able to work effectively with only a handful of these languages, a significant gap remains. Translation technology not only promises to functionally assist the facilitation of communication between individuals, groups, and organizations but also may help to eventually preserve more fringe languages. By acting as an “intellectual crutch,” translation technology may discourage the learning of a new language, entrenching the native tongue.

Currently, textual translation programs and software are integrated and used on the Web in business and in other organizational contexts. For example, the European Union uses translation software to handle approximately 10% of its translation duties. Large, global organizations are predicting a significant increase in the total amount of material that will need to be localized to fit regional language patterns—as much as sixfold by 2005. These trends, coupled with a stagnating number of human translators, are driving forces within the translation industry, which has been growing in gross revenue by 15–20% per year. Nevertheless, we are just beginning to use technology as an effective tool for translation and as an effective leverage for managing language.

Automated translation software has been used since the Cold War era, with some of the more modern,

advanced programs able to translate in 1 hour, with 90% accuracy, what might take a human 2 or 3 days. For example, synchronous automated translation systems (SATS) represent highly advanced technology that will allow seamless translation without any perceptible time delay. SATS-based systems are also expected to handle basic aspects of idiom and will be able to alert communicators of potentially vague language that would have a high probability of multiple meaning, thus prompting for more clarification. Advanced SATS systems are projected to be mainstream and in the market by 2015. One of the main trials for this type of advanced translation system will be how effectively it can handle a variety of unique translation issues, such as interaction patterns between low- and high-context cultural societies and the relative importance of contextuality in the communication process.

## 11. CONCLUSION

During the past 30 years, the need for effective and more precise translations of research materials has become a recognized requirement for high-quality international or cross-cultural research. The challenge of achieving conceptual equivalence and meaningful understanding among peoples of different cultural backgrounds and language has taken the field of translation far beyond “finding the same word in a different language.” This article discussed the problems that researchers have identified in achieving quality translation of research materials and the approaches needed to achieve the level of desired quality. The translation/back-translation, decentering model first proposed in the 1970s provided the foundation for which virtually all future approaches and concerns regarding quality translations have been addressed. Once conceptual equivalence in language is achieved, new challenges arise regarding how to achieve this conceptual equivalence with multiple dialects and contexts.

During the past 10 years, there have been many modifications and enhancements of the translation model as it has become more widely used in fields outside of psychology. Nursing, public health, leadership, and international business studies have all recognized the general power of the model as well as identified ways, and circumstances in which, it can be further enhanced. This article also identified some of the new frontiers and future challenges awaiting researchers as they attempt to achieve greater conceptual equivalence that considers not just verbal language

translations but also nonverbal, semantic, and contextual equivalence. Finally, the article discussed the future of translation with the use of technology and what that may mean to the field.

The goal of all these efforts is simple: to enable people to better understand one another. In doing so, we will be better able to understand why people in other countries and cultures do the things that they do, and in the process we will be able to understand why we behave the way we do. Ultimately, the true purpose of effective equivalent and conceptual translation is to help us understand and, in doing so, bring us closer together.

### See Also the Following Articles

Conformity across Cultures ■ Cross-Cultural Psychology, Overview ■ Literacy, Improvement of

### Further Reading

- Babcock, R., & Du-Babcock, B. (2001). Language based communication zones in international business communication. *Journal of Business Communication*, 38(4), 372–412.
- Berry, J. (1969). On cross-cultural comparability. *International Journal of Psychology*, 4, 119–128.
- Bontempo, R. (1993). Translation fidelity of psychological scales: An item response theory analysis of an individualism–collectivism scale. *Journal of Cross-Cultural Psychology*, 24(2), 149–166.
- Brislin, R. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216.
- Brislin, R. (1986). The wording and translation of research instruments. In W. Lonner, & J. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Beverly Hills, CA: Sage.
- D’Andrade, R. (2002). Cultural Darwinism and language. *American Anthropologist*, 104(1), 223–232.
- Erkut, S., Alarcon, O., Coll, C. G., Tropp, L., & Garcia, H. A. V. (1999). The dual-focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology*, 30(2), 206–218.
- Freivalds, J. (1999). The technology of translation. *Management Review*, 88(2), 48–52.
- Lehman-Wilzig, S. (2001). Babbling our way to a new Babel: Erasing the language barriers. *The Futurist*, 35(3), 16–23.
- Serebryakova-Collins, I. (1998). Here’s a twist: Talk about translation. *Business Communication Quarterly*, 61(4), 109–111.



# Transportation Systems, Overview

Barry H. Kantowitz and John M. Sullivan

University of Michigan Transportation Research Institute, Ann Arbor, Michigan, USA

1. Introduction
  2. Operator Workload and Distraction
  3. Driver Vision
- Further Reading

## GLOSSARY

**central fovea** A central region of the retina extending about 2 degrees in diameter of visual angle and exclusively composed of densely packed cone receptors. It is the part of the retina that supports seeing fine details and color.

**driver distraction** A reduction in a driver's attention that occurs when some secondary task interferes with the primary task of maintaining safe control of a vehicle in motion.

**human factors/human factors engineering** The discipline that tries to optimize the relationship between people and technology.

**mesopic sensitivity** A mixture of scotopic and photopic vision at low light levels, between approximately  $0.01 \text{ cd/m}^2$  and  $3 \text{ cd/m}^2$ , a common level found in nighttime driving.

**National Highway Traffic Safety Administration (NHTSA)** A unit of the U.S. Department of Transportation that is responsible for reducing deaths, injuries, and economic losses resulting from motor vehicle crashes.

**photopic sensitivity** Chromatic, cone-mediated vision used at high light levels (greater than about  $3 \text{ cd/m}^2$ ).

**pilot workload** An intervening variable that modulates the tuning between the demands of the environment and the capacity of the operator.

**scotopic sensitivity** Monochromatic, exclusively rod-mediated vision occurring at very low light levels (less than about  $0.01 \text{ cd/m}^2$ ).

**underload** A potential problem in highly automated modern aircraft in which most routine operations are carried out by the plane itself, which could let the flight crew become distracted or inattentive.

**United States Department of Transportation (USDOT)** Part of the executive branch of the U.S. government responsible for ensuring fast, safe, efficient, accessible, and convenient transportation systems for the American people.

Driving is a dangerous activity that is often taken for granted without understanding the risks of being on the road in a vehicle. Psychology can improve transportation systems in two ways. First, it provides explanations to help the transportation professional understand operator behavior. Second, its models can lead to effective countermeasures to decrease risk. These points are illustrated with discussions of operator workload and distraction and driver vision.

## 1. INTRODUCTION

Driving is a dangerous activity that can be hazardous to one's health; it can even kill. The World Health Organization estimates that road crashes are the third most important worldwide threat to public health. Indeed, the April 2004 World Health Day featured automobile crashes as its major topic. While most fatalities occur in developing countries, the United States also has a deplorable safety record, with over 42,000 fatalities in

2003. Someone dies in a car crash about every 12 minutes on average. A person is injured in a car crash about every 10 seconds. Automobile crashes are the most likely cause of death for young adults in the United States.

It is difficult to understand why society tolerates this high cost of mobility. If the aviation industry had a similar safety record, no airplane would ever leave the ground. Even considering the terrorist hijacking and crashing of four airplanes on September 11, 2001, flying is still much safer than driving. The risk of a fatality driving only 18 kilometers on a rural interstate highway equals the risk of any non-stop flight, regardless of distance, on a major airline.

Fatal crashes are a very low-frequency event for any single individual; no one reading this text has died in a fatal car crash. Therefore, many people believe that a car crash is an event that involves someone else. This comfortable belief disappears when a friend or a loved one is killed or injured in a car crash.

Most car crashes are due to human mistakes rather than failure of equipment. It is difficult to measure precisely what percentage of crashes is due to driver error because the basic data come from police reports after the crash. A dead or unconscious driver cannot explain the circumstances of a crash, and a live driver may not readily admit making an error. A recent University of Michigan Transportation Research Institute (UMTRI) analysis of a fatal accident database maintained by National Highway Traffic Safety Administration (NHTSA) estimated that approximately 60% of fatal crashes have human error as a key factor. This estimate is conservative; other estimates have ranged as high as 70 to 90%. Drivers engage in all manner of risky behaviors, such as drinking alcohol, not using seatbelts, and talking on cell phones. All too often, such risk is translated into death and injury on the highway.

At least part of the crash risk is a consequence of technological progress. Prior to the 20th century, surface transportation was largely accomplished with the aid of horsepower. This placed strict limits on the speed and weight of a conveyance and the amount of time the conveyance was in service. Moreover, from time to time, a horse would display a degree of intelligence that ensured its rider returned home whether or not the rider was fully conscious. Horsemanship was also considered a skill developed over time—it was decidedly dangerous for an unpracticed rider to mount certain horses.

The modern automobile has removed many of the limitations on horse-powered travel. Automobiles can achieve higher speeds, carry more cargo, and operate

more reliably and more consistently than horses. More automobiles are on the road than were horses, and drivers of widely different levels of skill and education are licensed to operate them.

These technological advances have indirectly raised the performance demands on automobile drivers. At high speeds, the available judgment time to take an evasive action, for example, is shorter, as is the time to correct an error. With heavier vehicles, momentum (kinetic energy) is higher, so collisions are more destructive. With more roadway users of varying levels of skill, there is more potential for interaction and more complicated decisions for the driver to make. With the aid of automotive lighting and fixed roadway lighting, travel at night is now both feasible and commonplace but also raises new questions regarding human performance in varied environments.

While technology enhances the risk of human death and injury by multiplying forces well beyond that produced by an unaided human, it also offers the potential to mitigate such damage. For example, in the United States, an Intelligent Vehicle Initiative research program attempts to use new technology to improve road safety by improving the understanding of driver distraction in order to develop new methods to measure the effects of in-vehicle technology on driver performance, to create new human factors guidelines for equipment design, and to develop new integrated approaches to reducing driver distraction caused by in-vehicle information systems.

Psychology, as the discipline that purports to explain and understand human behavior, has much to contribute to improving transportation systems. Its contributions are made in two ways. First, psychology provides explanations that help transportation professionals understand driver behavior. People are not mathematically rational and often satisfice rather than optimize. Yet their behavior is consistent and can be predicted. The following sections emphasize understanding the vehicle operator. Second, psychology can use its understanding to help construct countermeasures that will improve transportation safety. While many talented and dedicated transportation engineers have made substantial contributions to traffic safety based upon their experience and observations, these professionals are seldom well-trained in the subject matter and methods of psychology. Psychology, with its focus upon human behavior, can provide additional insights for developing countermeasures. The following sections offer psychological perspectives on effective countermeasures.

The remainder of this overview article uses two topics, pilot workload/driver distraction and driver

vision, to illustrate briefly how psychology is applied to transportation systems. Both understanding and countermeasures are discussed. Greater detail about specific transportation sub-systems will be found in the articles that follow this overview.

## **2. OPERATOR WORKLOAD AND DISTRACTION**

Regardless of the type of vehicle being controlled, demands on the operator, whether a pilot of an airplane or a driver of a passenger car, must fall within his or her mental and physical limitations. In modern transportation systems, human physical limits are seldom exceeded, but the cognitive load placed on an operator can often be incongruent with human capabilities. This psychological phenomenon has been extensively studied in aviation and is currently a major research topic in driving.

### **2.1. Pilot Workload**

Airplane pilots are highly trained and must demonstrate continued skill levels and sound health on a repetitive basis to maintain their licenses. Simulators are used extensively to practice emergency skills, and a rigorous system of checks and balances ensures the safety of the flying public. This system works well; flying on a major airline is far safer than driving on a highway.

One of the reasons for this high margin of safety is decades of research on pilot workload. An airplane pilot cannot pull over to the side of the road, as can a driver, when some problem develops. Pilot workload is defined as an intervening variable that modulates the tuning between the demands of the environment and the capacity of the operator. It is easy to understand that if workload is excessive, the pilot will be unable to make correct decisions in a timely manner, putting the aircraft at risk. Modern airplanes have extensive automation to reduce pilot workload. But if workload is too low, the pilot may be “out of the loop,” perhaps even asleep, and cannot respond correctly in a timely manner. Underload can be a problem in highly automated modern aircraft, in which most routine operations are carried out by the plane itself. Pilots sometimes joke that the aircrew of the future will be one human pilot and a dog; the dog’s job will be to keep the pilot from touching the controls. Some pilots believe that their manual flying skills are being degraded by the frequent use of automated systems. Indeed, a 1988 review of pilot workload by

Kantowitz and Casper concluded, “It would be tragic and ironic if our current efforts to measure pilot workload succeeded, only to be faced with a new generation of aircraft where pilot workload was so low that nobody bothered to measure it at all” (p. 184).

Fortunately, the theoretical understanding of pilot workload derived from cognitive models has been applied to the issue of automation. Automation can be implemented in a sensible way that takes pilot capabilities into account, or in a clumsy way that makes the pilot’s tasks more difficult to accomplish. The pilot should always be able to understand what the automation is doing at any moment. Some cockpit designs make this easier than others. For example, in Boeing aircraft, the throttle moves in response to automation commands, giving the pilot useful visual feedback. However, in Airbus designs, to save weight the throttle does not move, making it more difficult for the pilot to discover what the automation is doing.

Automation tends to fail in a sharp, sudden manner compared to the graceful degradation of human pilots. When automation can no longer cope with a situation, it turns off, leaving the pilot to deal with the problem. An incident that illustrates this occurred when a pilot made an initial error in transferring fuel from the wing tanks to a center tank. If the weight of fuel is not evenly distributed, the plane can roll over. The automation did correct for this error up to a point, but eventually the imbalance became too great. At that point, the automation turned off, and the pilot suddenly realized there was a significant problem. Had the plane been under manual control, this problem would have been detected much earlier by the crew. Although automation has improved aviation safety, it can produce new problems if the needs of the human pilots are not considered most carefully.

### **2.2. Driver Distraction**

Many of the theoretical and methodological tools used to learn about pilot workload can be, and are being, applied to the problem of driver distraction. Driver distraction occurs when a secondary task, such as tuning a radio, eating a pizza, entering data for an in-vehicle navigation system, or talking on a cell phone, interferes with the primary task of maintaining safe control of a vehicle in motion.

Driver workload and distraction are often studied by comparing some measure of driving performance with and without secondary tasks. It is important to remember that workload and distraction are theoretical

constructs that cannot be directly observed. Instead, they must be inferred from changes in performance, which can be observed and measured directly. Thus, there must be some experimental conditions that contrast single-task (only driving) and dual-task (driving plus other tasks) performance before inferences about workload and distraction can be generated. The typical outcome of such experiments is that adding most secondary tasks detracts from driving performance. The practical conclusion from such results is that effective countermeasures will limit the driver's ability to perform extraneous tasks, especially when the workload imposed by driving is high. Driving workload can be increased, for example, when sharp radius curves are negotiated, when the driver is fatigued, or when traffic density is substantial.

A very clever field study was able to relate secondary-task reaction time to road safety. In 1986, Harms showed that driver workload on 100-meter segments of roadway covaried with the number of reported accidents on these segments. Decreased performance on a secondary task, given constant primary-task performance, is a common index of driver workload. While many studies have demonstrated this in driving simulators and on test tracks, Harms was one of the first researchers to link workload and safety on actual highways.

Talking on a cell phone is a prime example of a secondary task that reduces the driver's ability to control a vehicle safely. Some states have banned the use of hand-held phones on the road, but a hands-free phone is also quite dangerous. An epidemiological study found that the relative risk of injury and collisions is 38% higher for cell phone users. Studies of vehicles on test tracks have also demonstrated the distraction effects of cell phone usage. Drivers who need to use a cell phone in the car should pull over to a safe place (a parking lot rather than the side of an interstate highway) to talk.

It is certainly possible to use cognitive models of the driver to invoke new technology that will build adaptive interfaces to limit secondary tasks, especially telematic secondary tasks from devices inside the vehicle. For example, in-vehicle navigation systems know when the driver is approaching a dangerous curve, and could then turn off the driver's cell phone. It seems likely that these new interfaces, sometimes called workload managers, will appear soon in new vehicles as effective countermeasures for excessive in-vehicle information loads. A recent on-road test of such a device used sensors that monitored the roadway and vehicle dynamics to dynamically estimate driver workload. When workload was too high, cell phone calls were automatically

diverted to a telephone mailbox without asking the driver's permission. Satisfaction with this demonstration system was high, suggesting that commercial systems can be developed to improve driver safety and comfort.

Many of the new in-vehicle information systems being deployed in ground vehicles have been developed not by large automotive companies, but by smaller suppliers, who often lack substantial in-house knowledge and capability in the fields of psychology and human factors. Fortunately, researchers have responded to this need by developing human factors guidelines funded by the United States and European governments that are available to all designers. For example, one highly used set of guidelines for advanced traveler information systems contains 75 specific design rules that can be applied by users who lack formal training in human factors. (These and other guidelines can be found on the Web by first searching for Turner Fairbanks Highway Research Center and then for human factors guidelines.)

### 3. DRIVER VISION

Key ingredients for safe driving include staying on the road, maintaining safe distances to other vehicles, recognizing and avoiding obstacles in the roadway, predicting other vehicle trajectories, reading and understanding traffic signs, and obeying traffic laws. Vision plays a critical role in providing input to the earliest stages of the complex chain of decisions and actions that influence driving behavior.

There are many important links between what a driver sees and what a driver does. A driver sees lane markings in the roadway, and steers to the center of the lane; sees a stopping vehicle ahead, and judges closing speed; sees a child at the side of the road, and slows down; sees a gap open up, and merges into traffic; sees a stop sign, and applies the brake; sees a detour sign, and plans an alternate route; sees an exit sign, and merges right in preparation for an exit. Seeing plays such an important role in driving that the desire to quantify it is difficult to resist. In a 1996 paper, Sivak noted that while it is often suggested that 90% of driving is visual, no real quantitative basis actually exists for this number. What can be quantified is the change in crash risk when visibility is reduced. In a 1996 article, Owens and Sivak found that reduced visibility in darkness doubles and quadruples the risk of some types of fatal crashes compared to the same crashes in daylight.

### 3.1. Basic Visibility

In discussing driving and vision, a distinction should be made between sensation and perception. To see an object, light receptors in an observer's eye must generate sufficient neural energy to signal the object's presence. This initial transduction of light energy to neural activity is principally the domain of sensation. However, even when adequate sensory registration of an object occurs, whether an object is actually seen and recognized also depends on the context of the visual scene and the knowledge, readiness, and experience of the observer. For example, objects are more quickly and reliably detected in an uncluttered scene than in a cluttered one; expected objects are more quickly and reliably detected than unexpected ones. Perception concerns what a person finally sees after registering sensory input.

Considering sensation first, if the receptors in the eye are insensitive to a stimulus, there is nothing for the perceptual system to interpret. The light sensitivity of the eye is dynamic and is influenced by the response characteristics and distributions of two classes of photoreceptors in the retina—rods and cones. Rods are more sensitive than cones and enable vision under low light levels, but they are disabled under bright conditions. Rods are absent in the central fovea of the retina, and rod-mediated vision is monochromatic. Cones are less sensitive to light than rods and are active under bright viewing conditions. They are densest in the central fovea and decline in number toward the periphery; there are three classes of cones with differing sensitivities to wavelengths of light, enabling color vision. To partially distinguish these differences in the response of the eye, vision has been identified as scotopic (monochromatic, rod-mediated vision) at low light levels (less than about  $0.1 \text{ cd/m}^2$ ), photopic (chromatic, cone-mediated vision) at high light levels (greater than about  $3 \text{ cd/m}^2$ ), and mesopic (a mixture of scotopic and photopic vision) at light levels in between. Light levels for nighttime driving conditions are generally in the mesopic range, although many objects illuminated by headlamps are in the photopic range, and unlit objects are in the scotopic range.

The implications for night driving are that objects that are looked at directly (i.e., exclusively with cones in the central fovea) require at least  $3 \text{ cd/m}^2$  to be seen clearly. Consequently, most objects require supplemental illumination to be seen with the same clarity as in daylight. One solution would be to flood the roadway with as much light as possible. However, such a strategy is both impractical and potentially disruptive to other road users, who may be temporarily

blinded or discomforted by the light. Low-beam headlamp designs address this problem by distributing light about the roadway in a way that attempts to balance the vehicle operator's illumination needs with the need to minimize glare to other road users. Light is aimed down, below the horizon, and to the right side of the road. This limits forward seeing distance and, particularly at high speed, leaves little time to respond to an obstacle in the roadway. A second, switched, high-beam system is intended to address this limitation by sending more light down the roadway. In theory, the driver switches between the two beam systems as the driving conditions require. In practice, it was found that only about less than half of the drivers in a Michigan study used their high-beam headlamps when they should.

Perhaps this happens because low-beam headlamps appear to provide sufficient illumination on modern roadways, which are marked by reflectorized lane striping and other supporting illumination. Drivers may not be aware that there is a visibility problem at all, judging their visual competence by their ability to maintain lane position. Unfortunately, while a driver's steering and lane-keeping abilities appear to suffer little under low-beam illumination, detection of unreflectorized, low-contrast objects appears to be selectively degraded. Leibowitz, Owens, and Tyrrell have suggested that this situation leads drivers to routinely overdrive their headlamps. That is, their effective stopping distance exceeds their seeing distance. Consistent with this view is the fact that drivers involved in fatal nighttime pedestrian crashes often report not seeing the pedestrians at all.

A variety of countermeasures have been devised to extend a driver's basic visual sensory capabilities while holding the problem of glare at bay. Such countermeasures include night-vision systems that render the invisible infrared radiation reflected or emitted by a roadway object as a visible image on a display, methods of adaptively redistributing the light from headlamps to optimize visibility in a variety of driving conditions, and more widespread use of reflectorized clothing. At the same time, countermeasures to address the problem of glare appear to dominate the discussion of roadway visibility, perhaps because there is more public awareness of glare than of poor visibility.

### 3.2. Conspicuity

Even when sufficient light is available to make a target object visible, it may be overlooked. A target's

detectability is also strongly influenced by visual context. A target is detected more slowly in complex visual scenes than in uncluttered visual scenes. Visual attributes that are unique to a target object can make it more conspicuous. Pedestrians who wear reflectors on moving limbs can be recognized at greater distances in the dark than when the same reflectors are worn on the torso. Limb motion heightens conspicuity. The advantage disappears, however, if the pedestrian is embedded in a cluttered visual environment.

Driver expectation also interferes with detection. Pedestrians are detected at greater distances in conditions in which they are expected (e.g., in crosswalks and urban intersections), and at shorter distances when they are not (e.g., along rural limited-access highways at night).

Various measures have been used to heighten the conspicuity of some roadway objects in specific situations. Frequent daytime collisions involving motorcycles have led to regulations requiring use of daytime running lamps on motorcycles. Fatal collisions with tractor trailers have prompted regulations requiring use of reflectorized sheeting materials or lights to mark the rear and side areas of trailers. The problem of rear-end collisions has been addressed with the use of the center high-mounted stop lamps to more clearly portray the action of a braking vehicle. And pedestrian visibility has been enhanced with the use of reflectorized materials in shoes, backpacks, and clothing.

### See Also the Following Articles

Accidents in Transportation ■ Aviation ■ Driving Safety  
 ■ Models for Transportation ■ Traffic Safety Assessment  
 ■ Travel Behavior and the Environment

### Further Reading

Birch, S. (2001). Adaptive front lighting. *Automotive Engineering*, 109, 39–42.

Campbell, J. L., Carney, C., & Kantowitz, B. H. (1999). Developing effective human factors design guidelines: A case study. *Transportation Human Factors Journal*, 1, 207–227.

Gish, K. W., Staplin, L. K., & Perel, M. (1999). Human factors issues related to use of vision enhancement system. *Transportation Research Record*, 1694, 1–9.

Hancock, P. A., Lesch, M., & Simmons, L. (2003). The distraction effects of phone use during a crucial driving maneuver. *Accident Analysis & Prevention*, 35, 501–514.

Harms, L. (1986). Drivers' attentional responses to environmental variations: A dual-task real traffic study. In A. G.

Gale et al. (Eds.), *Vision in vehicles*. Amsterdam: North Holland.

Herslund, M. B., & Jorgensen, N. O. (2003). Looked-but-failed-to-see errors in traffic. *Accident Analysis & Prevention*, 35, 885–891.

Kantowitz, B. H., et al. (2004). Development of critical knowledge gaps and research efforts. FHWA Technical Report, in press.

Kantowitz, B. H. (1988). Defining and measuring pilot mental workload. In J. R. Comstock, Jr. (Ed.), *Mental-state estimation 1987* (pp. 179–188). Hampton, VA: National Aeronautics and Space Administration, Scientific and Technical Information Division.

Kantowitz, B. H., & Campbell, J. L. (1996). Pilot workload and flightdeck automation. In R. Parasuraman, & M. Mouloua (Eds.), *Human performance in automated systems* (pp. 117–136). Mahwah, NJ: Lawrence Erlbaum Associates

Kantowitz, B. H., & Casper, P. A. (1988). Human workload in aviation. In E. Wiener, & D. Nagel (Eds.), *Human factors in aviation*. New York: Academic Press

Kantowitz, B. H., & Simsek, O. (2000). Secondary-task measures of driver workload. In P. Hancock, & P. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 395–408). Mahwah, NJ: Erlbaum.

Laberge-Nadeau, C., Maag, U., Bellavance, F., Lapierre, S. D., Desjardins, D., Messier, S., & Saidi, A. (2003). Wireless telephones and the risk of road crashes. *Accident Analysis & Prevention*, 35, 649–660.

Langham, M. P., & Moberly, N. J. (2003). Pedestrian conspicuity research: A review. *Ergonomics*, 46, 345–363.

Leibowitz, H. W., Owens, D. A., & Tyrrell, R. A. (1998). The assured clear distance ahead rule: Implications for nighttime traffic safety and the law. *Accident Analysis and Prevention*, 30, 93–99.

Luoma, J., Schumann, J., & Traube, E. C. (1996). Effects of retroreflector positioning on nighttime recognition of pedestrians. *Accident Analysis and Prevention*, 28, 377–383.

Mace, D., Garvey, P., Porter, R. J., Schwab, R. N., & Adrian, W. (2001). *Countermeasures for reducing the effects of headlight glare*. Washington, D.C.: The AAA Foundation for Traffic Safety.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Moberly, N. J., & Langham, M. P. (2002). Pedestrian conspicuity at night: Failure to observe a biological motion advantage in a high-clutter environment. *Applied Cognitive Psychology*, 16, 477–485.

Morgan, C. (2001). *The effectiveness of retroreflective tape on heavy trailers*. NHTSA Technical Report DOT HS 809 222. Washington, D.C.: National Highway Traffic Safety Administration.

Owens, D. A., & Sivak, M. (1996). Differentiation of visibility and alcohol as contributors to twilight road fatalities. *Human Factors*, 38, 680–689.

- Piechulla, W., Mayser, C., Gehrke, H., & König, W. (2003). Reducing drivers' mental workload by means of an adaptive man-machine interface. *Transportation Research Part F*, 6, 233–248.
- Resendes, R., & Martin, K. H. (2003). Intelligent vehicle initiative 2002 annual report-saving lives through advanced vehicle safety technology. FHWA-OP-03-101. Washington, DC: U.S. Federal Highway Administration.
- Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception*, 25, 1081–1089.
- Sivak, M., & Flannagan, M. J. (2003). Flying and driving after the September 11 attacks. *American Scientist*, 91, 6–8.
- Sullivan, J. M., Adachi, G., Mefford, M. L., & Flannagan, M. J. (2004). High-beam headlamp usage on unlighted rural roadways. *Lighting Research and Technology*, 36, 59–67.



# Travel Behavior and the Environment

Tommy Gärling

Göteborg University, Göteborg, Sweden

1. Introduction
  2. Determinants of Travel
  3. Transportation Systems
  4. Measures for Reducing Private Car Use
- Further Reading

In Western industrialized societies, effective urban transportation systems have evolved. A major factor in this evolution is the versatility of the private car. However, current increasing trends in private car use in urban areas are becoming a serious threat to the environment and people, thus outweighing the benefits. How can private car use be reduced? This article reviews research in environmental and transport psychology that addresses this question.

## GLOSSARY

- adaptation of car use* Changes such as more efficient car use, suppression of car trips, or a switch to other travel modes.
- car-use reduction* Decrease in car use by households and individuals.
- environmental threats* Air pollution, noise, accidents, congestion, and extensive use of land caused by motorized traffic.
- intelligent transportation systems* Applications of information technology for the substitution of travel or for supporting travel-related decisions.
- transportation* Movement of people and goods.
- travel choice* Choices to travel or not, where to travel, how to travel, and when to travel for single trips or multiple trips or tours.
- travel demand* Demand for travel derived from biological needs, social obligations, and personal desires to engage in activities.
- travel demand management* Policy measures aimed at reducing or changing demand for travel.

## 1. INTRODUCTION

In the interest of achieving effective urban transportation with positive welfare effects, Western industrialized societies have in many ways promoted the development of the private car. Only recently has it become evident that the increasing trends in private car use have serious detrimental effects to the environment and people that outweigh the beneficial effects. How can private car use be reduced? This urgent question is being addressed by research on transportation and the environment. What answers does this research provide? This article reviews what is known from research in environmental and transport psychology about the determinants of private car use, what are likely to be effective means for reducing private car

use, and what the short- and long-term consequences may be of reducing private car use.

## **2. DETERMINANTS OF TRAVEL**

### **2.1. Travel Demand**

Travel is a derived demand. Fulfilling biological needs, social obligations, and personal desires in human environments requires that people travel. In itself, however, travel is a cost or friction. Only occasionally does travel become a personal desire. In such cases, travel is a goal, not a means for attaining a goal.

The way a society is spatially organized is an essential determinant of demand for travel. The boost in motorized travel is primarily a consequence of two historical trends. One is urban sprawl. Cities have grown in size to accommodate an increasing population, thereby increasing demand for motorized travel. Another trend is that lifestyle changes have increased the complexity of travel demand far beyond the demand for home-to-work journeys on weekdays. Existing alternative travel modes lack the versatility of private cars, which is necessary to satisfy a complex travel demand.

### **2.2. Travel Choice**

Although people's travel options are to a large degree determined by the spatial organization of a society in conjunction with the available transportation means, some degree of freedom always exists in choosing to travel or not, where to travel, how to travel, and when to travel. An extensive research agenda has provided knowledge of how travel choices are made.

Traditionally, research has focused on single-trip choices of travel mode and destination. Travel time and monetary costs are important determinants. Aversion toward wait time, uncertain delays, and changes are other recurrent research findings. An important distinction is between habitual, impulsive, and planned travel. Commuting to work is an example of habitual travel. Shopping trips are more frequently impulsive. Long-distance travel is more meticulously planned than any other travel. Prechoice information search and deliberation have been shown to be reduced when travel becomes habitual. It has also been demonstrated how easily a car-use habit is developed and generalized across situational contexts.

Car travel frequently has multiple purposes. A recent aim of research is to understand interrelated travel choices. The unit of analysis is then a tour or trip chain—for instance, choices related to all trips from home and back during a day. Modeling interrelated travel choices has been an impetus to improve the traditional mathematical–statistical tools. Several generations of computer simulations have therefore been developed. The most recent generation models how people learn about and represent the environment as well as how they learn about options with positive outcomes. In computer simulations, realism of a travel plan replaces the traditional criterion of utility maximization under generalized cost constraints. Given a complex household agenda of activities (work, shopping, and leisure activities, including chauffeuring children), speed of travel is the single most important property of the transportation system. Although speed of car travel has slowed down in many urban areas, it is still the fastest travel mode and therefore likely to be chosen.

Both traditional and recent research has tended to downplay feelings of personal freedom, joy, security, and status as determinants of the choice of the private car. Cost factors may also play a more important role than assumed. Some of the costs of driving are nontransparent and thus may be largely discounted. Conversely, the capital cost of a car is not perceived as a sunk cost that should be discounted. Although the cost of an alternative is lower compared to the variable costs of driving, the additional cost for the alternative is psychologically unnecessary.

## **3. TRANSPORTATION SYSTEMS**

### **3.1. Properties**

Systems for ground transportation in urban areas differ in several respects. The basic means of transportation is walking. Another nonmotorized travel mode is cycling. All faster travel modes are motorized. To the users, speed, frequency of service, monetary costs, and comfort are the most important properties of transportation systems. An additional user-related concept is multimodal or “seamless” transportation. The concept refers to combinations of travel modes that facilitate convenient transportation from door to door. Developing such transportation systems is promising and is likely to enhance the attractiveness of alternatives to the private car.

### 3.2. Environmental Effects

The worldwide increase in motorized transportation in urban areas is expected to continue. The number of cars is increasing at a faster rate than the human population so that in 2050 more than 3 billion motor vehicles, compared to today's 600 million, are estimated to be in operation.

Conventional gasoline-powered cars emit carbon monoxide (CO), nitrogen oxides (NO<sub>x</sub>), sulfur dioxide (SO<sub>2</sub>), volatile organic compounds (VOCs), and carbon dioxide (CO<sub>2</sub>). Car traffic produces nearly two-thirds of all CO emissions, approximately one-third of CO<sub>2</sub> emissions, more than one-third of NO<sub>2</sub> emissions, and slightly more than one-fourth of emissions of VOCs. Substances are also released from the rubber in tires. These effects are strongly associated with a variety of significant environmental damage and human health problems. Depletion of the ozone layer, acidification, and global warming are examples of the former; respiratory and cardiovascular diseases are examples of the latter.

Other adverse effects of car use include noise, accidents, congestion, and extensive use of land for roadways and parking spaces. It is estimated that approximately 20% of the urban population is exposed to annoying traffic noise. Despite many attempts to improve traffic safety worldwide, fatalities and injuries are increasing. Traffic congestion slows car traffic, resulting in huge additional societal costs of urban transportation. In an attempt to reduce if not solve the congestion problem, increasing use of land for new infrastructure has encroached on agricultural land that is needed to produce food for the two-thirds of the world population that is starving.

## 4. MEASURES FOR REDUCING PRIVATE CAR USE

### 4.1. Classification

Several measures may conceivably reduce the gravity of the adverse car-use-related effects in metropolitan areas. Some of them (e.g., increasing the capacity of road infrastructure and building clean cars) do not necessarily require a reduction in car use. However, according to a general assessment of the current state, the sole short-term solution in urban areas is to reduce car use and to change use with regard to time and

TABLE I  
Measures for Reducing Car Use

---

Prohibition or physical restriction of car traffic
Fuel rationing
Taxation of cars and fuel
Road or congestion pricing
Parking control
Decreasing speed limits
Not constructing new road infrastructure
Teleworking
Land use planning
Bus and high-occupancy vehicle lanes
Park and ride schemes
Improved transit
Improved infrastructure for walking and biking
Marketing of alternative travel modes
Public information campaigns
Social modeling

---

space, particularly on major commuter arteries during peak hours and in downtown districts.

Widely proposed measures targeting reduced demand for private car use include increasing the cost of driving; prohibiting or physically restricting car traffic; improving alternative travel modes, such as transit, cycling, and walking; and changing the relative locations of homes, workplaces, shopping facilities, and recreational facilities so that driving distances are reduced. Table I lists several measures. These differ in effectiveness, cost, technical feasibility, and political feasibility.

Examples of implemented measures to manage travel demand are the prohibition of car traffic in downtown Cambridge, United Kingdom; road or congestion pricing in Singapore; and marketing of alternative travel modes in Perth, Australia. These three measures vary from more to less coercive—that is, in the degree to which adaptation is forced on car users.

### 4.2. Effects on Private Car Use

A continuum ranging from more to less coercive is obviously related to the effectiveness of measures for managing demand for car use. Implementing coercive measures is difficult, however. Despite awareness of the adverse effects of car use, car users in general are opposed to coercive measures since they impinge on their freedom to use the private car. Improving alternative travel modes is more popular. Also, coercive

measures are believed to be unfair since they affect everyone. At the same time, car users prefer more to less coercive measures since they minimize free riding and are believed to be more effective. At the political level, conflicts of interest make it difficult to reach consensus concerning the need to implement coercive measures. These difficulties are augmented by politicians' desire to win votes to remain in power if they believe (not always correctly) that their voters are not in favor of the measures.

The effectiveness of noncoercive measures depends on how able and willing car users are to reduce car use. Several moderating or mediating factors have been identified. Providing acceptable alternative travel modes is a basic requirement. However, car use has frequently developed into a habit so that information about alternatives is no longer sought. Important changes such as providing new alternatives may therefore go unnoticed. Information campaigns targeting the general public as well as individualized information about how to change travel have nevertheless had limited success.

Alternative travel modes are difficult if not impossible to make more attractive. Therefore, in addition to providing them, car use must be made less attractive. Introducing fees for car use is a preferred measure for doing this. In this way, some of the costs for improving alternative travel modes such as transit or constructing new infrastructure for walking and biking may be covered. However, substantial cost increases appear to be required to achieve sizeable changes in car use. Implementing new pricing measures or increasing existing ones are therefore neither accepted by the public nor politically feasible.

### 4.3. Consequences of Reducing Private Car Use

Private car use results from the interrelated choices made by car owners that are dependent on their physiological needs, obligations, or personal desires to participate in activities such as work, shopping, and leisure. Car-use reduction must therefore be viewed more broadly as an adaptation to changes in travel options that potentially have consequences for engagement in activities.

Research is beginning to disentangle the process of adaptation of car use. A well-founded hypothesis states that in this process households or individuals set car-use reduction goals and form plans to attain the goals. In selecting change options, they start with those that

TABLE II  
Adaptation to Reduced Car Use

<i>Adaptation</i>	<i>Psychological cost</i>
More efficient car use (trip chaining, car pooling, and choice of closer destinations)	Additional planning
Trip suppression	Activity suppression
Switching to alternative travel mode	Increased time pressure, inconveniences

are psychologically less costly. Since less costly options are also likely to be less effective, the process entails a hierarchy of choices of options over time from less costly and effective to more costly and effective. A possible hierarchy is shown in Table II.

If change is forced on car users, the costs of smaller changes may be less flexibility and, as a consequence, an increased need for planning. Additional changes may make necessary unwanted lifestyle changes that restrict the opportunity to perform some activities. Switching to alternative travel modes that are slower and less comfortable incurs the additional cost of increased time pressure.

### 4.4. Car-Use Reduction and Information Technology

Table III lists various devices that current information technology offers as components of intelligent transportation systems that are likely to decrease the psychological costs of adaptation to reduced car use. Information technology may make both substitution of travel feasible and car use more efficient. Examples of the former are flexible work arrangements, teleconferencing, and electronic commerce. Flexible working arrangements (e.g., telecommuting and a compressed workweek) lead to an overall reduction in car use due to reduced commuting. However, any savings are often offset by new, longer car trips for noncommuting purposes, many of which were previously linked to the commuting trips. Teleconferencing has the potential to decrease travel to the extent that the technology required is sufficiently developed. However, it is often the case that long trips, as opposed to short trips, are most likely to be substituted by teleconferencing. Electronic commerce also has the potential to suppress trips, primarily service and maintenance trips such as

**TABLE III**  
**Decision Support and Substitutes Provided  
 by Information Technology**

<i>Adaptation</i>	<i>Information technology application</i>
More efficient car use	Invehicle navigation systems Pretrip/enroute congestion or accident information Car pooling or car sharing information Information about road use charges
Suppressing trips	Teleworking Teleconferences Electronic commerce
Switching mode	Online public transport time tables and routing Information about road use charges Park-and-ride information Information about traffic regulations

for paying bills and other nongrocery shopping (e.g., book and music purchases), although inroads have been made with respect to grocery shopping.

Information technology making car use more efficient includes pretrip or en route information about travel time, shortest routes, congestion and accidents, and available parking spaces, as well as in-vehicle navigation systems. Information technology thus supports choices of departure times, routes, and destinations. Drivers are remarkably sensitive to such information, even when its role is not directly and explicitly relevant to route choice. Information technology may also make information better and more easily accessible with regard to alternative travel modes such as transit, thus possibly influencing travel mode choice. Furthermore, more efficient car use can be facilitated through the increased use of car-sharing schemes, whose efficiency and attractiveness (e.g., online bookings and payments) can be much enhanced with information technology.

Other information technology applications include electronically charging drivers when road pricing schemes are implemented. An essential aim of such applications is to provide information about the expected costs, particularly if the costs vary with zone, time of day, and congestion. Road pricing is of course not likely to affect car use unless such information is accessed, understood, and acted on. If these conditions are fulfilled, however, information about

road pricing allows drivers to avoid congestion. Furthermore, providing cost information may facilitate not only the implementation of car-use reduction goals but also the setting of such goals.

The effectiveness of prohibiting car traffic from entering protected downtown areas may increase if coupled with easily accessible pretrip information about alternative travel modes and destinations (e.g., parking information or information about park-and-ride options). Such coupling may reduce the likelihood of car users driving as near as possible to the prohibited area instead of switching to transit or walking at an earlier point.

### See Also the Following Articles

Aviation ■ Driving Safety ■ Transportation Systems, Overview

### Further Reading

- Ahrentze, T., & Timmermans, H. J. P. (2000). *ALBATROSS: A learning based transportation oriented simulation system*. Eindhoven, The Netherlands: European Institute of Retailing and Services Studies.
- Axhausen, K., & Gärling, T. (1992). Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems. *Transport Reviews*, 12, 323–341.
- Bovy, P. H. L., & Stern, E. (1990). *Route choice: Wayfinding in transport networks*. Dordrecht, The Netherlands: Kluwer.
- Button, K., & Hensher, D. (Eds.). (2001–2003). *Handbooks in transport*. Amsterdam: Pergamon/Elsevier.
- Gärling, T., Eek, D., Loukopoulos, P., Fujii, S., Johansson-Stenman, O., Kitamura, R., Pendyala, R., & Vilhelmson, B. (2002). A conceptual analysis of the impact of travel demand management on private car use. *Transport Policy*, 9, 59–70.
- Gärling, T., Gärling, A., & Loukopoulos, P. (2002). Forecasting psychological consequences of car-use reduction: A challenge to an environmental psychology of transportation. *Applied Psychology: An International Review*, 51, 90–106.
- Golob, T. F. (2001). Travelbehaviour.com: Activity approaches to modeling the effects of information technology on personal travel behaviour. In D. Hensher (Ed.), *Travel behavior research: The leading edge* (pp. 145–183). Amsterdam: Pergamon.
- Hanson, S. (Ed.). (1995). *The geography of urban transportation*. Amsterdam: Elsevier.
- Novaco, R. (2001). Psychology of transportation. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences*. Amsterdam: Elsevier.
- Schade, J., & Schlag, B. (Eds.). (2003). *Acceptability of transport pricing strategies*. Oxford, UK: Elsevier.



# Trust

Ana Cristina Costa

Delft University of Technology, Delft, The Netherlands

1. Introduction
  2. Defining and Measuring Trust
  3. Developing Trust
  4. The Benefits of Trust
  5. Challenges and Dilemmas of Trust: Implications for Practice
  6. Emerging Issues in Trust Research
  7. Summary
- Further Reading

Trust is central to human life and is considered to be essential for stable relationships, fundamental for maintaining cooperation, vital to any exchange, and necessary for even the most routine of everyday interactions. In organizations, the importance of trust has been recognized at both the interpersonal and institutional levels. Two types of trust can be distinguished: interpersonal trust, which refers to trust between people, and system or institutional trust, which refers to trust in the functioning of organizational, institutional, and social systems.

## GLOSSARY

**calculus-based trust** Trust based on calculation most likely to be found at the start of new relationships or partnerships when the parties do not have any prior connections or knowledge about each other.

**identification-based trust** Trust that arises from identification and empathy with the other party's desires and intentions and that results from sharing a common identity and similar values.

**interpersonal trust** Trust that is personalized and concerns trust between people that is characteristic of primary and small group relationships.

**knowledge-based trust** Trust that derives from knowledge built from prior interactions and/or collaborations allowing partners to have enough information to predict each other's likely behavior.

**swift trust** Trust that develops when people interact more in a role-based manner than in a person-based manner.

**system or institutional trust** Trust that is abstract and refers to trust in the functioning of organizations, institutions, political systems, and societies as a whole.

## 1. INTRODUCTION

In organizational science, applied psychology, and related fields, trust has become a major focus of theory and research during recent years. Although scholars have long been interested in the study of trust, changes in the functioning of organizations and in scholarly thinking have led to a revival of interest in this topic.

Contributing to the rise of trust on the research agenda is the growing evidence of the wide and varied benefits of trust for individuals, teams, and organizations. Numerous studies have demonstrated how increases in trust result, either directly or indirectly, in more positive workplace behaviors and attitudes, better team processes, and superior levels of performance. Considerable efforts have also been expended to apply emerging trust theory to a variety of important organizational problems, with some being the

result of the increase of distrust in institutions' policies and management. Although trust might not be the ultimate solution for all problems, as organizations have moved toward flatter and more team-based forms of organized activity, interpersonal dynamics—and trust in particular—have become critical elements in achieving effective collaboration. Perhaps more now than ever before, organizations must invest in conditions that facilitate trust among members to survive and remain effective.

Trust becomes a vital concept when there are significant risks involved in trusting (i.e., vulnerability) and when there is objective uncertainty about future consequences of trusting. In organizations, uncertainty and vulnerability arise for different reasons. Consequently, trust has been studied with regard to different respects and has been approached through different perspectives. This article reviews the progress regarding the conceptualization and measurement of trust. It discusses how trust emerges and develops through different forms, and reflects on important benefits of trust at different levels of the organization. In addition, it discusses some of the most important challenges for research and practice in this area.

## 2. DEFINING AND MEASURING TRUST

Because trust is so central to human relations, many definitions have been put forward from a variety of perspectives, but to date there is no agreed-on general definition. In general, trust has been conceptualized as either a one-dimensional or a multidimensional phenomenon.

### 2.1. One-Dimensional Conceptualizations

One-dimensional conceptualizations include definitions of trust as a psychological state or as a behavioral choice. The psychological state approach defines trust in terms of interrelated cognitive processes and orientations toward beliefs or positive expectations in relation to others. Some definitions apply to contexts where these others are identifiable, for example, Rousseau and colleagues' definition of trust: "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another." In other definitions, trust is presented as a more general attitude or expectancy about other people

or about the general social system. Common to these definitions is the reference to states of "vulnerability," "confidence," and "positive expectations."

Trust as a choice behavior can be seen as the willingness to take risks by acting on the basis of words, actions, or decisions of others. Some definitions portray trust as a more or less rational decision, motivated primarily by perception of risks, concerning either the probability of successful cooperation or the possibility of transaction cost reduction. More recent definitions suggest the need to conceptualize trust not only as a calculative orientation toward risk but also as a social orientation toward other people and societies as a whole. As recent research has demonstrated, one not only "thinks" but also "feels" trust. Throughout research various behaviors have appeared to be indicative of trust, including cooperative behaviors such as open communication, acceptance of influence, forbearance from opportunism, and lack of monitoring.

Despite the differences, trust as a psychological state and trust as a choice behavior are in principle compatible approaches. In most definitions, trust is related to individual attributions about other people's intentions and motives underlying their behavior. These attributions have cognitive and emotional grounds, and they influence and are influenced by individuals' general beliefs and expectations about the treatment they will receive from others. Finally, these attributions are closely linked to the decision to engage in behaviors of trust when interacting with others.

### 2.2. Multidimensional Conceptualizations

In discussing how expectations underlying trust affect subsequent behavior, several scholars have alluded to the fact that an adequate analysis of trust begins by recognizing its multifaceted character. Lewis and Weigert suggest that trust is a highly complex phenomenon containing distinct cognitive, emotional, and behavioral dimensions. People trust on the basis of emotions and cognitive processes that discriminate others as being trustworthy, untrustworthy, or unknown. The behavior reflects the significance of these earlier dimensions and enables individuals to act on their judgments. Mayer and colleagues argue that the assessment of one's trustworthiness influences and is influenced by a general propensity to trust others that develops from past experiences and

helps to shape the general beliefs about the treatment people expect to receive from others. Several authors have described trust as a function of different interrelated components, including individual propensities, interpersonal assessments of trustworthiness, behaviors of cooperation, and lack of monitoring. Gambetta explicitly connected positive expectations with cooperation: “When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him.”

Besides defining trust in different ways, the approaches discussed here suggest that trust may be composed of different elements, each of a different nature. Individual predispositions, characteristics, and intentions of the trustee(s) and situational conditions such as the degree of risk can determine both the level and the potential form that trust takes. The relative importance of these factors is determined by the type and course of the relationships in where trust occurs.

### 2.3. Trust Measures

The conceptual diversity regarding trust is also reflected in the instruments developed to measure this concept in different contexts and levels of analysis. An overview of the most relevant measures of trust for organizational settings is given in Table I. These trust measures focus on different components and are based on different conceptualizations. For example, instruments of trust that assess another party’s trustworthiness are often related to definitions of trust as a psychological state (e.g., Conditions of Trust Inventory [CTI]), whereas measures that emphasize behaviors are associated with behavioral choice definitions of trust. Multidimensional definitions of trust, on the other hand, include perceptions of both trustworthiness and trust behaviors (e.g., Organizational Trust Inventory [OTI], Team Trust Inventory [TTI]).

Multidimensional measures have the advantage of assessing simultaneously different facets of trust in a particular relationship context. However, one-dimensional measures can be useful for explaining differences between individuals or groups with respect to one of the trust facets. Instruments that measure the propensity to trust in general (e.g., Interpersonal Trust Scale [ITS], Revised Philosophies of Human Nature [RPHN]) can be used to explain differences in trust behavior between individuals or groups in the same

situation. Some individuals tend to give most people the benefit of the doubt by trusting loved ones and strangers alike until experience shows that this trust is not warranted. Others expect only the worst of everyone around them. Although propensity to trust itself does not determine whether or not a person will trust in a specific situation, as situations become increasingly unfamiliar, its influence increases. In modern organizations, for instance, the growing need for cooperation between and within boundaries brings together people who do not necessarily know each other or have a previous history of working together. In such situations, the general willingness to trust others can strongly determine initial trust decisions.

Instruments that focus on trustworthiness and/or trust behaviors are situation specific and are particularly relevant in contexts of interpersonal and group relationships. Expectations about an individual’s trustworthiness have been consistently found to be a strong determinant of trust in contexts of specific others. In general, the assessment of trustworthiness is based on three primary criteria—benevolence, competence, and integrity—the relative importance of which can differ from situation to situation. In some situations, competence can be more important. Other situations can demand political sensitiveness, making an individual’s integrity more important.

Measuring trust behaviors in a particular context can be useful for learning about an individual’s motives and intentions and for being able to make inferences of trustworthiness. Trust behaviors reflect the willingness to be vulnerable to others whose actions an individual does not control and can also differ across contexts. For instance, in strategic alliances, trust behaviors may include financial and formal investment, whereas within teams, cooperation and lack of monitoring would be more reflective of trust.

The measures described here assess different facets of trust, and most have limited applicability to particular contexts and levels of analysis. The choice of a particular measure over another implies some thought about the approach to trust chosen, the level of analysis in which trust is measured, and the purpose for which the instrument is used. In exploring differences between trusters, propensity measures can be more relevant. Differences between trustees can be better explained through measures that focus on trustworthiness. Behavioral measures of trust reflect the trust given in a particular setting. The whole process of trust in particular contexts can be better explored through multidimensional measures.

**TABLE I**  
**Overview of Trust Instruments Used in Organizational Contexts**

<i>Instrument/Author(s)</i>	<i>Components of trust measured</i>	<i>Factors identified</i>	<i>Applicability/Level of analysis</i>
Interpersonal Trust Scale (Rotter, 1967)	General willingness to trust others	Trust in peers Trust in institutions	Assesses credibility of social agents and societal optimism  Suitable to measure individual's trust in generalized others
Revised Philosophies of Human Nature (Wrightsmann, 1964)	Propensity to trust others	Trust Cynicism	Measures expectations about the way in which people generally behave Indicates the general ability to trust others
Mayer & Davis (1999)	Propensity to trust Trustworthiness	Propensity Ability, benevolence, and integrity	Applicable for studying trust in organizations or top management
Cook & Wall (1986)	Trustworthiness	Faith in intentions and confidence in actions of peers and management	Applicable to study trust in peers and trust in management
Conditions of Trust Inventory (Butler, 1991)	Trustworthiness	Integrity, honesty, fairness, competence, consistency, loyalty, discreetness, openness, receptivity, availability, and fulfillment	Assesses trustworthiness of specific peers Is applicable to dyads in either horizontal or vertical work relationships
McAllister (1995)	Trustworthiness	Cognition based Affect based	Refers to managers' trust in peers Applies to dyads
Smith & Barclay (1997)	Trustworthiness Trust behaviors	Character, role competence, judgment, motives and intentions, relationship investment, acceptance of influence, communication, openness, and forbearance from opportunism	Distinguishes between perceived trustworthiness and trust behaviors in specific peers Applied in strategic alliances contexts
Currall & Judge (1995)	Trust behaviors	Communication openness, informal accord, task coordination, and surveillance	Applied to boundary role-persons dyads
Organizational Trust Inventory (Cummings & Bromiley, 1996)	Perceived trustworthiness Behavior intentions	Keeping commitments Honest in negotiations Does not take advantage	Multidimensional measure of trust Applicable to dyads or business units
Team Trust Inventory (Costa, 2003)	Propensity to trust Perceived trustworthiness Trust behaviors	Propensity to trust Perceived trustworthiness Cooperative behaviors Lack of monitoring	Multidimensional measure of trust for team contexts

### 3. DEVELOPING TRUST

It is a matter of common understanding that trust is not static. Trust develops over time and goes through various phases such as building, declining, and renewing. People start a new relationship with a certain level of trust, either higher or lower, depending on their own individual dispositions, past experiences, familiarity with or knowledge of the other party, contextual contingencies, and perceived incentives for cooperation. Three bases for developing trust in work relationships have been identified: calculus, knowledge, and identification.

#### 3.1. Calculus-Based Trust

Calculus-based trust is likely to be found in relationships that are new and are formed between partners or team members who do not have any prior social connections. The basis for trust is the calculation of the benefits of being trusting and trustworthy against the threat that if trust is violated, one's reputation may be damaged by the other person's network of friends or colleagues. In other words, calculation underpins the belief that the actions of another party will be beneficial and reliable rather than the opposite. Trust in this sense consists of ensuring consistency of behavior, that is, that people will do what they say they will do. Making agreements, establishing the "rules of the game," and monitoring compliance can help to build this mode of trust. If others act predictably by complying with the agreements, they will build a reputation of trustworthiness on which trust can develop.

#### 3.2. Knowledge-Based Trust

Knowledge-based trust is based on prior interactions and cooperation. It occurs when one has enough information about others to understand them and to be able to predict their likely behavior. Knowledge-based trust relies on information rather than deterrence. Trust develops largely as a function of the parties having a history of interaction that allows them to develop a generalized expectancy that each other's behavior is predictable and trustworthy. Here, trust develops through information about preferences, wishes, and behaviors of the other party that develop over time as a consequence of the parties having a history of interaction. Partners or team members are likely to develop common ways of thinking through their sharing of

experiences and information. Two key processes are responsible for achieving this stage: regular communication and interaction.

#### 3.3. Identification-Based Trust

Identification-based trust is based on identification with the other party's desires and intentions. This form of trust arises between people who share a common identity, meaning that they hold similar values, including a shared concept of moral obligation. Trust develops through empathy because the parties effectively understand and appreciate the each other's wants. This mutual understanding is developed to the point where each party can effectively act for the other. Identification-based trust develops as both parties know and predict each other's needs, choices, and preferences and also share some of those as their own. Increased identification enables people to empathize strongly with the other party and incorporate parts of the other party into their own identity as collective identity.

These three basic conditions promote the emergence of trust and influence the individual expectations about the parties' trustworthiness and their willingness to engage in trusting behaviors and, consequently, their initial level of trust. If trust must be built from scratch, making agreements and monitoring compliance of team members can help to build trust. In mature relationships, trust may so solid, and the risks involved may be so small, that monitoring is not needed to maintain optimal cooperation. Calculus, knowledge, and identification can also be seen as sequential linked stages of trust where achieving trust at one level enables the development of trust at the next level. Although not all relationships develop fully and so might not develop past the first or second stage, effective cooperation is unlikely to take place or persist if relational trust does not develop between individuals who interact intensively over a period of time. Therefore, the development from calculus-based trust to knowledge-based trust is crucial.

### 4. THE BENEFITS OF TRUST

As organizations have come to rely less on structures and formal arrangements and more on collaboration and cooperation inside and outside the firms, trust is given more emphasis as one of the fundamental motors of these processes. If trust is absent, no one will risk

moving first and everyone will sacrifice the gains of collaboration and cooperation. Although laws, rules, and contracts are still necessary conditions for the stability and prosperity of organizations, for these to prevail, they must be based on reciprocity, moral obligation, and trust. The benefits of trust have been discussed in relation to reduction of transaction costs, collaboration, and cooperation within and between organizations as well as to the effects on performance and effectiveness.

#### 4.1. Transaction Costs Reduction

Currently, competitive success has become increasingly dependent on the reduction of transaction costs as the requirements for quality have escalated internationally and markets have become more uncertain. Traditional transaction costs theory has neglected trust in its assumption that the risk of opportunism is high in the “governance” of relations. However, as many have recognized, trust is an element of every transaction that can be accounted for either by previous experience or by lack of contrary evidence. Moreover, the presence of trust in transactions is likely to generate more trust at other levels because transactions are embedded in professional and social networks, diminishing the hazard of opportunism. This does not necessarily mean that trust is a by-product of all transactions. The possibility of opportunism does exist. But, in the long run, opportunism can be very costly because it will increase the amount of control costs and most certainly will inhibit future transactions.

Reducing transaction costs through trust can be achieved through reputation. Having a trustworthy reputation can be of great “economic value” because it plays an important role in determining the willingness to enter into a business exchange with a given actor. This good repute will further lead to positive expectations in the future, enhance the level of trust, and promote the actor’s willingness to cooperate. It has been argued that trust, besides reducing transaction costs and monitoring performance costs, also eliminates the need to install control systems that are designed to obtain short-term financial results. Nevertheless, trust should not be seen as a replacement for either market or hierarchical forms of transactions. These governance mechanisms are necessary for the establishment of communities, which can be seen as an important starter for trust based on shared ethical norms and values underlying the communities.

#### 4.2. Collaboration and Cooperation within and between Organizations

It is commonly assumed that some level of trust must exist so that cooperation between partners can be achieved freely. Yet, trust as a precondition to cooperation can be subjected to different demands of intensity. Requirements for trust in organizations are dependent on the mechanisms that govern the cooperative decisions and the social arrangements in which those decisions are made.

Within interorganizational forms of collaboration, trust is important to the extent that it facilitates information exchange and reciprocity between partners, leading to organizational learning. When interfirm collaborations are forged from common memberships to a professional community, existing ties, or local community partnerships, trust seems to develop more on interpersonal bases and community networks. When alliances are forged mainly from mutual dependencies and/or calculation of resource needs, trust tends to develop from formal bases that can be more costly and time consuming. One way to facilitate collaboration between organizations is to make trust a part of the organizations’ routines and practices so that collaborations between firms can continue successfully. This means that the major source of trust should be institutional. Key individuals or groups (i.e., boundary spanners) do play an important role in interfirm forms of collaboration. However, problems of turnover and the possibility of communication breakdown on the part of these individuals make trust at this level a very fragile form of governance.

Also within organizations, the importance of trust is recognized both at the institutional and interpersonal levels. Trust based on institutional arrangements, such as laws, rules, and professional practices that support the organization as a whole, create a common ground for understanding actions and enhance patterns of behaviors that are extendable to all organizational members. These create a general climate in which trust is produced and generalized to other levels. Although trust is important to the functioning of all forms of the organization, alternative forms have clear trust requirements and managerial philosophies have clear implicit levels of trust. Failures to meet these requirements bring various consequences. For instance, in functional organizational forms, insufficiency of trust reduces efficiency; in divisional forms, it reduces effectiveness and increases costs; in matrix forms, it causes the form to fail; and in networks, it

causes organizations to fail. The tendency toward flatter and more team-based forms of organized activity shows that the importance of trust in organizations has been augmented significantly. In the network form, it has become one of the requirements for organizations' survival.

New policies emphasizing interpersonal and intergroup dynamics in the workplace have accentuated the importance of trust at an interpersonal level. Interpersonal trust is a product of rational decisions and emotional bonds and can be based on different mechanisms, depending on the degree of knowledge or familiarity among the people involved. In situations where individuals have accumulated meaningful knowledge and have established some kind of bond with one another, interpersonal trust tends to be more based on the attributions that individuals make about the other person's character and the motives and intentions underlying these actions. In situations where individuals have little information about one another or have not yet established any kind of bond with one another, trust may initially develop on the basis of individual dispositions, situational constraints, and/or institutional arrangements.

#### **4.3. Effects on Performance and Effectiveness**

Apart from the general assumption that trust is an important lubricant of the social system and a facilitator of coordinated action among individuals, several important benefits have been associated with trust. Trust has been associated with positive work attitudes and behaviors such as open communication and information exchange, acceptance of influence, citizenship behavior, commitment, satisfaction, and reduction of conflict. Trust has also been associated with the ability to enhance collaboration and mutual learning between individuals by leading to cooperation and higher individual and team performance.

Some authors have labeled these effects as spontaneous sociability. When operationalized in behavioral terms, spontaneous sociability refers to numberless forms of cooperative, altruistic, and extra role behavior in which members engage that enhance collective well-being and further attainment of collective goals. It should be noted that there is also evidence showing that trust alone is not always enough and that spontaneous sociability also depends on the individual's perception regarding the efficacy of his or her own actions. In addition,

implications of trust for performance effectiveness should be interpreted carefully. First, trust cannot be seen as one of the main indicators of performance given that tasks require specific abilities and knowledge to be performed adequately. In situations where individuals or teams do not possess adequate skills and knowledge to accomplish their tasks successfully, trust probably will not improve performance. Second, performance has been found to be dependent on numerous determinants, meaning that trust is just one of these indicators. In certain conditions, however, trust may play a more a moderated role by facilitating communication and openness, which can lead to the exchange of important knowledge or generate critical discussions that may be beneficial for the end product. The importance of such effects is again dependent on the trust requirements associated with the functioning of teams and organizations.

### **5. CHALLENGES AND DILEMMAS OF TRUST: IMPLICATIONS FOR PRACTICE**

Desirable development and maintenance of trust, however, is neither easy nor ensured. Newspaper headlines during recent years have headed stories about the collapse or fraud of several major corporations. Although reorganization processes within companies are essential for survival in the current volatile environment, they often have been accompanied by a considerable lack of trust between organizational members. Ironically, the increased need for trust in modern organizations has also made the role of checks and controls even more important. This is because violations of trust are more likely to occur when vulnerability increases, because excesses of trust can lead to miscalculations of risks or illusions of greatness. In either case, such effects can lead to drastic consequences for organizations. To deal with some of the challenges and dilemmas of trust, it is important to recognize two important issues regarding trust in organizational settings: the fragility of trust and the limitations of trust.

#### **5.1. Fragility of Trust**

It is often suggested that trust is typically created rather slowly, but it can be destroyed in an instant by a single mishap or mistake. One of the reasons for this is that trust is a phenomenon that feeds on itself. When a relationship or partnership is growing, trust builds on

evidence of trustworthiness and on a track record of being trustworthy. If trust-building actions are taken, the overall level of trust grows until it begins to even out during the maintenance stage. The building process is often slow because people tend to be reticent about trusting, particularly with regard to those who are not known. Once trust has been built, the demands for evidence of trustworthiness diminish, and this can lead to a false sense of security. In some cases, parties may actively discount information that implies untrustworthiness, at least for a while. However, if solid evidence of untrustworthiness emerges, trust is destroyed quickly and the collapse is dramatic.

Both psychological and structural processes can explain the asymmetry between creating trust and destroying trust. Because people tend to pay more attention to negative events (trust destroying) than to positive events (trust building), negative events carry greater weight in people's trust judgments. This is partly because negative events are more visible and frequently take the form of specific, well-defined incidents such as lies and discoveries of errors, whereas positive events, although sometimes visible, are more often fuzzy or indistinct. The accumulation of relevant experience with low trust can frame the interpretation of events to reinforce this belief.

Organizational structures and management policies also contribute to the fragility of trust in organizations. By becoming less bureaucratic and more focused on collaborative processes, organizations have increased the level of uncertainty and vulnerability between their members. Trust based only on formal mechanisms is insufficient for organizations to function effectively. In modern organizations, both institutional trust and interpersonal trust constitute an essential feature of effective functioning. Successful companies or work relationships develop, treasure, preserve, and nurture trust at both levels. They recognize that trust helps during both good and bad times, and they recognize that if trust is lost it might never be recovered.

### 5.2. Limitations of Trust

Trust is pervasive and indispensable to the functioning of organizations, but when it is excessive or when its limitations are not recognized, trust might not be prudent and can have adverse impacts. Because trust goes beyond the available evidence, there is always an amount of unknown risk that results from the lack of control or of complete knowledge about future outcomes and about the actions of the other party.

Excesses of trust can be harmful when they entail a large risk of economic damage and threaten the survival of organizations. From interpersonal to organizational interaction, excesses of trust can have adverse impacts on the functioning of firms. For instance, a culture of excessive trust and benevolence may become oppressive when it results in the expression of criticism being taboo and when it blocks the direct voicing of complaints, thereby preventing fast and direct solutions between partners in conflicts. It can yield "group-think," which reflects excessive cohesiveness and a shared illusion of invulnerability. This can lead to an accumulation of unsolved conflicts, biasing perceptions of evidence of incompetence or malevolence.

Excesses of trust can occur for different reasons. One is naivety, ignorance, or cognitive immaturity that makes one party unaware of risks involved in the trust situation. Here, the experience of broken trust will teach awareness. Another reason is the feeling of omnipotence, that is, an overestimation of one's power to control an untrustworthy partner with the feeling that one cannot be hurt by damage imposed by the partner. Another reason is impulsiveness deriving from carelessness or from putting a large emphasis on the current benefits relative to later adverse effects. This can be related to greed.

### 5.3. Implications for Practice

Recognition of the importance of trust for the functioning of organizations has grown considerably. Although the success of some organizations and work relationships can be more dependent on trust than others, in general, those organizations and work relationships that are successful find ways in which to build and maintain trust. Some are built on stronger foundations than are others, but the management of trust is always critical.

Leaders and managers play a central role in managing the overall level of trust as well as more specific levels of trust across organizations. Because leaders and managers are responsible for combining strategy, structure, overall operating logic, resource allocation, and governance of organizations, the levels of trust displayed by these mechanisms might well be reciprocated at other levels of organizations. When managers are perceived as competent, fair, and open regarding the sharing of information, the overall level of trust in organizations is likely to improve. Leaders determine, or at least influence, the degree of interdependence between organizational units or individuals, and this in turn determines the level of trust reliance between parties that is necessary to achieve

the work goals. Although trust does not necessarily leads to cooperation, interdependence across units, teams, or individuals creates conditions for information exchange and proximity, and these may allow room for acceptance and trust development. Managers can also try to create a climate that supports trustworthy behaviors between members and, at the same time, provides adequate guidance so that trust does not become dangerous.

The increasing emphasis on short-term and highly interdependent multiproject environments brings together people who are not necessarily part of the structure, culture, and norms of the same organization. Also in such contexts, developing trust between individuals or partners becomes essential. Lacking a common ground, a certain degree of familiarity, and prior experience in working together can influence the basis on which trust might develop. Recent research has described the so-called "swift trust" as a new form of trust that develops quickly, mostly in contexts where there is no previous history of interaction between partners, there is no prospect of long-term interaction, and people interact more in a role-based manner than in a person-based manner. The concept of swift trust suggests that members initially import trust rather than develop trust. Because there is insufficient time for these expectations to be built from scratch, they tend to be imported from other settings and imposed quickly in categorical forms. The categories invoked to speed up these perceptions can reflect roles, industry recipes, cultural cues, and occupational and identity-based stereotypes. However, this form of trust might not be enough to function effectively. For trust to develop further, parties should act trustworthy by complying with the agreements and expectations. Yet the short life span of some of the current work relationships also increases the opportunity to take advantage. Again, the involvement of supervisors and managers can be crucial in maintaining trust at an optimal level that benefits all parties involved.

Trust is a significant decision-making process under conditions of vulnerability and uncertainty. Managing trust reflects awareness of the mental processes behind the decision to trust and awareness of the risks involved in such a decision. Risk assessment involves considerations about another party's trustworthiness and about the situational factors that weigh the likelihood of positive and negative long-term effects of the trust. Because the information about another party's competence, motives, and intentions is often incomplete, people recall past experiences as their best guides. In more unfamiliar contexts, these considerations can be more dependent on individual propensities, reputation, or

external information by inquiring about another party's trustworthiness. Assessing the risk before trusting is crucial to this process because when trust is not fulfilled, the trusting party suffers an unpleasant consequence that is greater than the gain that the trusting party would have received if trust had been fulfilled. Moreover, if trust is broken, it might be destroyed completely or might require a long time to be rebuilt to its former state. Therefore, trust must be preserved and nurtured by information that reinforces perceptions of trustworthiness and trustworthy behavior.

## 6. EMERGING ISSUES IN TRUST RESEARCH

Despite the considerable development in theory building regarding trust in organizations, issues concerning the definition and its measurement at various levels of analysis remain the subject of continuous discussion. The growing globalization of business has increased the emphasis on collaboration and cooperation across organizations and even countries, raising the issue of trust across cultural borders. How trust develops between representatives from various organizations and/or countries, what basis of trust will be sustained in such contexts, and what factors (both personal and organizational) will determine the level of trust must be explored more comprehensively. Similarly, the increasing competition of markets has led many competing firms to form alliances to survive. Whether trust in such contexts can develop from a calculative basis into more mature stages to achieve effective cooperation and whether partners will trust one another with their business secrets to promote learning are questions that need to be researched in more detail. In addition, the rapid advances in information and communication technology have created virtual work environments. Understanding the role of trust in such environments, how trust might develop without face-to-face interaction, and what factors are considered to assess another party's trustworthiness has become extremely important.

## 7. SUMMARY

Trust is a central to human life as well as organizational life and is one of the pillars on which organizations function. Despite the numerous benefits of trust and the current high value placed on it, trust also has

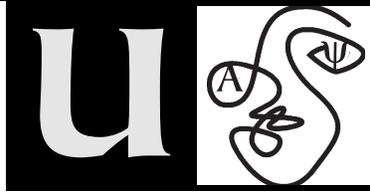
limitations. This article has focused on the definition of trust, on issues concerning the development and maintenance of trust, on the benefits of trust in organizational settings, and on challenges and dilemmas associated with trust. Emerging issues in trust research were noted, including the development of trust in a diverse workforce and across organizational borders, in competitive environments, and in virtual environments.

### See Also the Following Articles

Cooperation at Work ■ Leadership and Culture

### Further Reading

- Costa, A. C., Roe, R. A., & Taillieu, T. (2001). Trust within teams: The relation with performance effectiveness. *European Journal of Work and Organisational Psychology, 10*, 225–244.
- Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science, 12*, 450–467.
- Gambetta, D. (Ed.). (1988). *Trust: Making and breaking cooperative relations*. Oxford, UK: Basil Blackwell.
- Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50*, 569–598.
- Kramer, R. M., & Tyler, T. R. (1996). *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, CA: Sage.
- Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces, 63*, 967–985.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709–734.
- McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*, 24–59.
- Nooteboom, B., & Six, F. (Eds.). (2003). *The trust process in organizations*. Cheltenham, UK: Edward Elgar.
- Rousseau, M. T., Stikin, S. B., Burt, S. B., & Carmerer, C. (1998). Not so different after all: Across-discipline view of trust. *Academy of Management Review, 23*, 393–404.
- Williamson, O. E. (1975). *Market and hierarchies*. New York: Free Press.



# Urban Environments and Human Behavior

Gabriel Moser

Université René Descartes, Paris, France

1. Introduction
  2. The Urban Environment as a Stressful Milieu
  3. The Urban Environment as a Place of Living
  4. The Rationale of Intervention in Urban Environments
  5. Public/Private Urban Environments: The Neighborhood
  6. Urban Public Environments: The City
  7. Toward Sustainability of Urban Environments
- Further Reading

**urban stressors** The physical (noise and pollution) or social (density) environmental conditions that require various forms of coping.

It is estimated that by the year 2015 approximately half of the world's population will be living in large cities. Cities are characterized by their cultural, religious, ethnic, economic, and social diversity and their stressful living conditions. Citizens are exposed daily to a variety of environmental stressors, such as noise, pollution, or density, which make them adopt specific adjustment behaviors characterized by unfriendliness and egoism. Environmental psychology is playing an increasing role in helping shape the urban environment in order to conserve or create well-being for citizens. Socioenvironmental engineering acts on both the physical and the social aspects of the environment. It is mainly focused on the neighborhood but often extends to the relationship between the individual and the urban structure as a whole. Socioenvironmental engineering is a way to meet people's requirements for well-being and satisfaction and to achieve urban sustainability.

## GLOSSARY

- appropriation** A particular affective relationship to an object, which then becomes part of oneself.
- attachment** An emotional bond, which people often develop to their life space.
- people–environment congruity** The relation to a living place that corresponds to the aspirations of the individual and allows him or her to satisfy his or her material and social needs.
- place identity** Either a person's expressed identification to place or an aspect of his or her identity comparable to social identity.
- socioenvironmental engineering** Intervening on the physical as well as the social aspects of an environment involving all the concerned actors.
- sustainability** A development capable of satisfying the needs of the current generation without compromising the ability of future generations to satisfy their own needs.
- urban behavior** Any specific protective behavior adopted by city dwellers to cope with exposure to urban stressors.

## 1. INTRODUCTION

Since they concentrate employment opportunities, culture, entertainment, and recreation, cities have

traditionally been a pole of attraction. It is estimated that approximately half of the world's population lives in large cities. Cities have progressively become the "natural" habitat of most individuals.

Although for some inhabitants the urban environment provides an ideal level of stimulation, an optimal context for their activities, and numerous opportunities for interpersonal contacts, urban daily life conditions have become increasingly difficult. How do these reputedly bad environmental conditions affect urbanites' behavior? When examining human behavior in urban environments, the following aspects have to be taken into account:

1. The urban environment as a stressful milieu (i.e., a place of daily exposure to a variety of specific stressful environmental conditions)
2. The city as place of living (i.e., appropriated by the individual and to which he or she identifies)

Before analyzing how environmental psychology addresses life in cities, an overview of the previously mentioned aspects is provided with regard to urban behavior. How can environmental psychology help identify the processes that regulate and mediate the individual's relationship to his or her environment? The requirements of sustainability give a new impetus to environmental psychology by emphasizing individual needs and well-being. By dealing with the individual-environment relationship at different spatial levels (i.e., from dwelling to the neighborhood and the city), environmental psychology provides analyses and interventions aimed at improving urbanites' daily life at these different levels. The point of departure of this analysis is often the set of physical specifications of the environment (e.g., noise, pollution, and planning and layout of physical space) that act directly on the individual or are mediated by the social variables of the environment (e.g., crowding and population heterogeneity). However, the physical and social factors are inextricably linked with regard to their effects on individuals' perceptions and behavior.

## **2. THE URBAN ENVIRONMENT AS A STRESSFUL MILIEU**

Exposure to the various stressors and particular life conditions of urban environments results in specific urban behaviors.

### **2.1. Urban Stressors**

Urban environments display three specific aspects: ambient physical conditions (essentially noise and air pollution), social conditions (density), and overstimulation (exposure to a high number and variety of visual and other stimulations). Cities are more noisy, more dense, and more polluted.

Noise is the most common stress to which city dwellers are exposed on a daily basis. The levels of noise are associated with the size of the population. Noise is the most frequently mentioned form of stress and the one that leads to the most frequent complaints. One-fourth of urbanites are exposed to at least one loud noise at work, home, or during commuting on a daily basis.

Because of traffic, exposure to pollution is more frequent in large cities. Although the effects of air pollution on health (e.g., respiratory problems for children and elderly) are well documented, pollution has little direct effect on the behavior of urban residents.

Generally, people are more likely to perceive environmental problems when they can hear (noise), see (smoke), smell, or feel them. The relationship between exposure to air pollution and health is mediated by the perception of exposure. An important source of information is the media's interpretation of noise or pollution levels, which may influence public perceptions and attitudes.

Cities often have different forms of density due to large population concentrations. City dwellers may be exposed to social density outdoors (high number of individuals in one area) and to spatial density in their habitat (small amount of space per person) and, therefore, struggle for resources due to a low ratio of population to the amount of available resources (i.e., too many people and insufficient transport facilities, shops, and services). Furthermore, the large amount of visual and auditory stimuli to which city dwellers are exposed is a constant informational overload that increases fatigue.

### **2.2. Urban Living Conditions**

Citizens not only are exposed to environmental stressors but also deal with poor living conditions. Mobility in cities is more constrained than that in small towns: Commuting daily from residence to workplace and dealing with frequent gridlock or crowded mass transit increases urbanites' stress. They have less spare time

because of longer commuting times, and there is less time available for social interactions.

Bureaucratic formalism is also more frequent in large cities than in small towns. Compared to inhabitants of small towns, urbanites experience more frequent difficulties accessing various services due to increased competition (subway, taxis, etc.).

The level of crime is higher in large cities. Therefore, urbanites are more likely to be exposed to violence and crime. The delinquency rate is higher in city centers than in suburbs, which is commonly explained by the numerous opportunities to commit crime in the city along with a lower probability of being caught. Fear of crime (which is not necessarily correlated with objective crime rates) restricts urbanites' behavior by making them feel vulnerable. However, despite being generally more exposed to environmental stressors, the inhabitants of large cities, with the exception of newcomers, do not cite these factors as annoying them more frequently than do the inhabitants of small towns or even people living in rural habitats.

Overall, city life seems to be constraining and demanding. All analyses of environmental overload, environmental stress, and behavioral constraints note the potentially negative effects of life in large cities compared to life in small towns. When exposed to such stressful situations, city dwellers engage in coping processes. Coping is an attempt to reestablish or gain control over a situation regarded as stressful. Therefore, in order to achieve congruity between themselves and the environment, urbanites adopt protective adaptation processes and perform behavioral and/or cognitive changes, which engender specific urban behavioral norms.

### 2.3. Specific Urban Behavior

At the beginning of the 20th century, Simmel noted the social withdrawal and egoistic behavior of urbanites, as well as their detachment and disinterest toward others. In 1970, Milgram outlined three urban adaptation behaviors: coping by choosing priorities, self-protection by erecting psychological barriers around oneself, and creating rules and institutions.

Urban behavioral particularities comprise the following: segmented and functional ways of interacting with one another, anonymity and lack of involvement, indifference toward deviant and bizarre behaviors, and restriction and selectivity of responses to other peoples' demands. Environmental constraints such as gridlock,

overcrowding, and fear of crime enhance behaviors aimed at gaining control and freedom of action. The pace of life is more hectic. Certain activities are performed more quickly in large cities than elsewhere. People walk faster in large cities than in small towns. The fear of crime provokes behavioral restrictions, especially among elderly people. Withdrawal and decreasing responsiveness to the diverse demands of other people are typical reactions to stress. People filter inputs, focus attention on the most important demands, and neglect the peripheral stimuli, thus avoiding unwanted interactions. By concentrating on their own needs, urbanites not only deliberately avoid contact with others but also pay less attention to what is going on around them. The constant expression of such adjustment behavior in urban setting indicates that this is a normative behavior.

In terms of interpersonal relationships, individuals establish interaction priorities, excluding a certain number of contacts. In urban settings, interpersonal relationships are governed by rigorous rules, which enable individuals to preserve the minimum privacy needed in order to protect themselves from intrusion by others despite high-density situations. Gatherings with friends are more often planned in advance, and informal encounters are rare and limited to the home or workplace.

The characteristics of interpersonal exchanges in the urban environment not only tend to make social contacts more superficial and eliminate interactions that have no personal benefit to the individual but also lead to neglect expressions of politeness. Civility is a disinterested act. It involves a common code of conduct, which is indispensable for maintaining the social fabric, based on respect for the other. The individual holds the door open for others or says "hello" when entering a shop less often in large than in small cities.

In short, public spaces are locations for expressing community values, rules, and codes that are indispensable for individuals to maintain social distance and to protect their personal space. Social densities of large cities create both physical proximity and social distance. Thus, the regulation of social life in a densely populated society is based on adhering to routines that can be anticipated and predicted. Civil behaviors are indispensable for social interaction in the urban setting. On the other hand, incivilities may be regarded as the negation of the processes of the civilization of morals or as standards of behavior—that is, as breaches of social mores.

### 3. THE URBAN ENVIRONMENT AS A PLACE OF LIVING

People identify with the place where they live: "My home is my castle." Feeling at home is an important aspect of one's well-being and this is a universal human experience. Individuals create privileged relations to their habitats, which become their "homes," and their homes provide identity. This involves a way of building one's life within a geographically limited space. Several characteristics transform a habitat into one's "home": centrality, continuity, privacy, expression of self and identity, social relationships, atmosphere (warmth and pleasantness), and characteristics of the physical environment.

#### 3.1. Place Identity and Attachment

There are two ways of relating one's place and one's identity. The first is place identification. This concept refers to a person's expressed identification with a place, which becomes part of his or her social identity. The second way of relating one's place and one's identity is through place identity, a specific aspect of the individual's identity.

An important mechanism that supports place identity is the attachment to a specific place. Individuals often develop an emotional bond to their life space, essentially their home and the neighborhood, but often also urban places and spaces on a larger scale. Such anchoring is an ongoing process, dependent on individual time perspectives: The duration of residence is essential for the individual's appropriation of his or her life space, which in turn is indispensable for well-being. Anchoring reflects the individual's motivations, social status, family situation, and projects for the future. Urban environmental appropriation revolves around forming social and interpersonal relationships. Individuals who make emotional investments in their neighborhood are more satisfied with their interpersonal relations in that neighborhood and develop a sense of well-being.

Appropriation may be regarded as a particular affective relation to an object that may then become part of the identity of the individual. Appropriation means having control over one's living space, and it is a prerequisite of feeling "safe" and "at home": It is essential to the construction of spatial identity. Appropriation is important for the individual to be able to organize and personalize his or her life space. This may be crucial

not only in one's home but also in any other urban place in which one makes a temporal investment. Steady or transitionally occupied places produce place attachment and are often accompanied by ties to personal objects such as furniture, pictures and souvenirs, which indicate the appropriation of places.

Residential history determines how one appropriates home environments. The temporal horizon and needs that reflect one's position in one's life cycle affect one's ability to generate or failure to develop positive investments in the neighborhood. Finally, one can assume that individuals who appropriate their environment and feel at home where they live may also care more about the environment in general (i.e., more frequently exhibit ecologically beneficial behavior).

#### 3.2. From Appropriation to Urban Identity

Appropriation operates not only at the level of one's place of dwelling but also extends to urban places, such as the street, the district, the town, or even the country, and it is accompanied by social networking (family, friends, neighbors, and communities). The urban identity is essentially acquired through various territorially bonded social networks. The feeling of being at home in one's neighborhood is linked to the frequency of encounters, the nature of local relationships, and the satisfaction that they provide. It involves social integration extended to local service providers such as physicians, shopkeepers, and others, and it constitutes the framework for the different individual networks (workplace, leisure, school, etc.). Furthermore, with regard to feelings of attachment, the social relations provided by a place may be more relevant than the place itself. Taking root corresponds to a desire for stability and permanency in one's way of relating to a certain place and one's involvement with the place in the long term. The sociospatial aspects of traditional urban structures include the residential environment—the district delimited by architectural, social, and administrative boundaries. The monofunctionalism of city structures has extended the way of relating to the city as a place of daily life beyond the traditional local district.

Increased residential mobility has resulted in a shift in place investment from one's housing to one's furniture and other "belongings" that contribute to the individual's identity. Moreover, settlement identity, which refers to individual preferences for certain

types of habitat, allows mobile individuals to conserve coherence and identity across various residences.

#### 4. THE RATIONALE OF INTERVENTION IN URBAN ENVIRONMENTS

Environmental psychology analyzes individuals' and communities' perceptions, attitudes, and behaviors with regard to their explicit relationships with the physical and social contexts in which these individuals and communities live. Depending on the level of analyses, different aspects and problems are addressed.

##### 4.1. Levels of Analyses from the Dwelling to the City

In dealing with the relationship between an individual and his or her life space, environmental psychology aims to understand and act in order to fulfill people's needs and aspirations. The notions of space and place are central. For each level of spatial reference, the discipline enables the investigation of people-environment interactions, whether involving an individual, group, or society (Table I).

The interactions and/or transactions between people and their urban environment at the different environmental levels may be understood by considering the individuals' perceptions, needs, opportunities, and means of control. In addition to appropriation and anchoring, control over one's environment plays a crucial role. Here, the scale of analysis allows one to account for a significant dimension in the individual-environment relation in terms of possibilities of control and mastery over the environment in question. Such

potential control allows the individuals' aspirations to more or less master different aspects of their environment. These are important, individual, and direct with respect to the microenvironment. In one's habitat, appropriation is not shared with strangers, and the individual has full control of this place. In proximate and semipublic environments, the possibility of control and mastery is no longer individual but shared with a community. Participation at the local level is possible, and a sense of belonging may be created in certain circumstances, but control is necessarily mediated by other individuals. At the city level, control is beyond the range of the individual, and it may only be collective or societal. The possibility to control or master one's environment is an important requisite in anchoring and identity processes. How people relate to their place of living affects their well-being.

##### 4.2. Socioenvironmental Engineering in Urban Environments

Environmental psychologists intervene at the interface of the physical and social aspects of the urban environment and engage in interdisciplinary problem solving in collaboration with architects and planners.

###### 4.2.1. The Sociophysical Interface

Environmental psychologists intervene as socioenvironmental engineers, which means that they intervene as much on the social as on the physical aspect of the urban environment, examining the interrelation of these two aspects. They constantly bear in mind that an intervention on one of the aspects of the urban environment always has an impact on all the other aspects of the environment (i.e., changes in the physical aspect affect the social aspect and vice versa).

TABLE I  
Social and Physical Aspects of Urban Environmental Extensions and Type of Control

	<i>Physical aspects</i>	<i>Social aspects</i>	<i>Type of control</i>
Level 1: Micro environments	<i>Private/semiprivate spaces:</i> Habitat, housing, building	Individuals/family	Extended control (unshared)
Level 2: Proximate environments	<i>Proximate public spaces:</i> Blocks of apartments, neighborhood, district	Neighborhood community	Shared control built on consensus
Level 3: Public urban environments	<i>Public spaces:</i> Town, city	Aggregates of individuals	Mediated control (participation possible at local level)

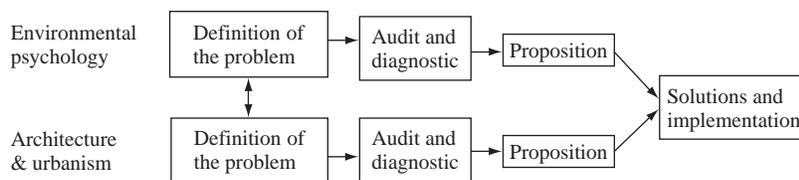


FIGURE 1 Rationale of interdisciplinary intervention.

Thus, environmental psychologists focus on the user of the environment as much as on the environment. Furthermore, each level of spatial reference has different implications that must be taken into account for any intervention.

Interventions in terms of socioenvironmental management concern public urban spaces and/or neighborhoods, and they consist of the diagnostic, audit, and intervention followed by a control after implementation of the intervention. By definition, they are implemented on existing structures. Socioenvironmental engineering also contributes to the understanding of the discrepancies between public environmental policies and individual aspirations, expectations, and beliefs. This may be important, for instance, with regard to noise or air pollution, but it also applies to the layout of the built urban environment. By identifying individuals' perceptions and needs, people–environment congruity can be enhanced.

#### 4.2.2. Cooperation with Other Disciplines

The study of the urban environment requires multidisciplinary, which is indispensable for environmental intervention. Responding to social demands often involves multiple approaches from different disciplines, each having its own characteristics and following its own disciplinary logic.

In urban matters, environmental psychologists often work in cooperation with members of other disciplines and professions, such as architects, engineers, landscape architects, and planners. Their respective analyses and interventions often have a different point of departure, focusing either on the built environment or on the people who live in it. Sometimes, a sequential gap may also appear: The architects, designers, and planners intervene at the conception and building process of the contexts in which people live, whereas social sciences typically focus on the existing structures.

Only knowledge integrated from the different disciplinary approaches may provide an adequate

response to the challenges of living in cities. This integration has to follow a specific rationale in order to be efficient (Fig. 1).

The interdisciplinary approach is a multiple and parallel approach to the same problem, where each discipline provides a solution according to its own scientific logic that is then integrated in the global response. The cooperation between social sciences and design disciplines usually results in successful problem solving of urban issues because it allows all the different aspects of people's relations to the built environment to be addressed.

## 5. PUBLIC/PRIVATE URBAN ENVIRONMENTS: THE NEIGHBORHOOD

Interventions at the proximate level deal with the immediate environment of the individual's living space. This may be a row of houses or apartment building, the immediate neighborhood, or the leisure areas in the immediate surroundings of homes, such as parks and green areas. These areas are referred to as semipublic spaces, which means that control over them is shared within a community (inhabitants of a neighborhood, users, etc.).

### 5.1. People–Environment Congruity: Feeling at Home

Concepts such as “attachment to a specific place” and “sense of community” help our understanding of how individuals and groups create bonds to a specific place. Although the “size” and the comfort of the habitable space are essential for residential satisfaction, other aspects of the living conditions also modulate its importance. Social integration in the neighborhood, the availability of satisfactory services, the existence of green space, the aesthetics of the built environment, and easy access to the public transportation network encourage the affective investment of the neighborhood and enhance people–environment congruity.

The more this congruity is important, the more the neighborhood matches one's aspirations and allows one to satisfy his or her material and social needs, and the more one develops attachment toward one's neighborhood. In certain cases, the interpersonal relationships that the individual develops within his or her place of living are more important for the feeling of belonging and attachment than the physical specifications of the urban environment. This is for instance the case in shantytowns, in which forced dislocations and allocations of far more comfortable new dwellings are regarded as traumatic by the inhabitants.

The feeling of being "at home" is closely connected to satisfaction with one's neighborhood. The degree of satisfaction with the environmental attributes of one's neighborhood (i.e., the green spaces, the aesthetics of the built framework and the degree of noise) determines both the intensity of the affectivity that one develops toward the neighborhood and one's feeling of well-being.

For individuals who have already acquired "basic" living conditions and who have an income that allows them to achieve a good quality of life, the agreeable character of the neighborhood has a modulating effect on satisfaction concerning available space in the dwelling. The feeling of being at home in one's neighborhood is linked to the frequency of encounters with others in the neighborhood, the number of close relations, the nature of local relationships, and the satisfaction with them. This explains why social relations related to a determined place may be more important to feelings of attachment than the place itself.

Besides the home and neighborhood environments, other urban environments involve congruity between people and their environment, including workplaces and institutional environments such as hospitals, prisons, and homes for children or the elderly. How can these environments be designed to meet the needs of their occupants?

## 5.2. Control of Density in a Collective Habitat

High density is an environmental characteristic often regarded as responsible for the individual's withdrawal, inattention to others, and aggressive behaviors. The most specific interventions concern the control of density and the regulation of behavior in spaces open to the public.

Certain interventions involving the design of space are effective in diminishing the feeling of being

constrained, notably in collective buildings, prisons, public libraries, and cafeterias. Some interventions may explicitly refer to the principles elaborated by Barker in his ecological theory when organizing public spaces or spaces open to the public.

The larger the collective buildings, the less social activity the residents tend to have (neighborhood relations, mutual help, and associative involvement). This is usually attributed to the management of intimacy in public spaces. For instance, inhabitants of buildings with long corridors feel threatened with regard to intimacy. An architectural intervention consisting of fractioning long corridors by introducing doors diminishes the feeling of crowding, enhances the feeling of control, and favors interactions in the spaces so delimited.

## 5.3. Management of Neighborhood Spaces

Contributions to the management of neighborhood spaces involve the "habitability" of neighborhood (districts) and the treatment of environmental problems and insecurity. Such interventions mainly focus on neighborhood revitalization and interventions designed to improve residents' satisfaction and reduce crime and feelings of insecurity.

### 5.3.1. Neighborhood Revitalization

Neighborhood revitalization has to follow two apparently contradictory principles: (i) to foster attachment and a feeling of belonging to a well-delimited and clearly identifiable area and (ii) to open the area in order to connect people to the city.

The key to successfully revitalizing certain districts is in the way in which these two contradictory exigencies are handled. Such interventions have to be conducted jointly with regard to both the social and the physical aspects of the environment and engage all the relevant actors (i.e., allow cooperation between the inhabitants, the sponsors of the project, and the municipality).

Interventions concerned with the development of urban residential environments consist of improving sidewalks or pathways for inhabitants. Their aims are (i) to encourage pedestrians to use proximate pathways in order to promote social interaction among inhabitants and foster control of different outdoor activities, (ii) to diversify the outdoor environment in order to

increase walking in the whole proximate environment, (iii) to give residents the opportunity to engage in different activities (shopping, recreational opportunities, walking, etc.), and (iv) to reduce the number of cars passing through the area and limit their speed (“quiet zones”). Such interventions often require redesigning the district as well as the built environment so that dwellings face toward roads and the installation of green spaces with multiple recreation opportunities for all residents.

Similarly, the concept of defensible space, introduced in 1972 by Newman, has often been used as a guideline for the design of proximate environments. Development designed to give maximum control over public space to inhabitants, by erecting symbolic or real barriers, restricting access to the district, designing well-illuminated pathways, and encouraging the individual appropriation of dwellings by differentiating them from each other, has given positive results. Postrenovation evaluations have found less maintenance costs and less reprehensible behaviors.

Such interventions correspond to a territorial logic. Encouraging outdoor activities and restricting vehicles allow a neighborhood to define its identity and to appropriate and therefore better control its immediate environment and the street, which enhance the feeling of belonging and increase the feeling of being safe.

### **5.3.2. Environmental Annoyances**

Dealing with environmental annoyances, such as excessive noise or undesired odors, often requires an intervention at the source and socioenvironmental management based on an intervention to improve all aspects of the neighborhood so that residents’ well-being is improved and crystallization on the incriminated annoyance is avoided.

The same type of intervention is necessary when a new industry is planned, especially if it is reputed to create new and specific annoyances. In this case, particular attention must be paid to the possible NIMBY (not in my back yard) phenomenon (i.e., preliminary rejection by inhabitants).

### **5.3.3. Insecurity and the Feeling of Insecurity**

The feeling of insecurity and subsequent fear are seldom correlated with objective dangers and are essentially based on visual cues. Fear of crime is exacerbated by an urban environment that appears to be poorly

maintained, characterized by littering, vandalism and graffiti, uncleanness, lack of illumination, and deviant behavior. Cleaning up vandalized buildings and urban furniture reduces littering, and other interventions focused on the signs of disorder and lack of control in the neighborhood provide an increased feeling of security.

A pleasant environment may also contribute to the feeling of security of inhabitants. Nasar repeatedly showed that coherence, compatibility, congruity, legibility, and clarity of the urban environment enhance the evaluative quality of cities, downtown street scenes, and residential scenes. On the other hand, a design that obviously impedes a certain behavior in the sense of defensible space may be interpreted as a sign of danger and produce a feeling of insecurity.

## **6. URBAN PUBLIC ENVIRONMENTS: THE CITY**

Regarding public urban space, environmental psychology may intervene in two ways: (i) management of the urban structure by improving the legibility of the urban environment and revitalizing the urban public space and (ii) management of the urban social diversity. The main aim of both types of intervention is to favor appropriation.

### **6.1. Management of Public Urban Spaces**

The management of urban environments mainly consists of encouraging connections between urban districts and giving urbanites the opportunity to appropriate the various public spaces for a large diversity of activities, including recreation and contact with nature.

#### **6.1.1. Homogeneity and Legibility of the Urban Space**

The urban structure is not homogeneous. Beyond the classical distinction between the center and the suburbs, several zones may be identified. One can distinguish fluid zones, to which individuals do not anchor because such zones are only designed for mobility; anchorable zones (i.e., living places) with which people can identify; and functional zones, which are similar to fluid zones (people simply “go through” them).

Furthermore, the natural fractures (e.g., hills and rivers) or artificial ones (e.g., highways and rail tracks) may shape the urban landscape and isolate districts from one another both physically and psychologically.

According to Lynch's principles of urban structure, the mental map methodology allows the problems of isolation to be diagnosed and the means of redesigning urban layouts of districts or zones to be proposed in order to improve their interconnectedness.

Legibility refers to another, mostly neglected aspect of the environment—that is, the way collective memory and cultural references shape the perception of the urban environment and, consequently, enhance people's appropriation. In identifying "landmarks" such as monuments or geographical particularities, mental maps contribute to the understanding of the legibility of the city.

### **6.1.2. Parks, Squares, and Nature in Urban Environments**

The gentrification of public spaces concerns not only the design of sidewalks and streets in order to encourage their use by pedestrians but also the interventions on parks and squares. Urban squares often suffer either monospecific or a deficit of frequentation due to the limited activities that they offer. Providing an increased number of activities and introducing water (fountains), trees for shade, and benches for resting and interaction make them attractive for passer-byes and opens them to a variety of different occasional and regular users.

Urban residents often seek nature and want to visit urban parks, gardens, and recreational areas for leisure. Green spaces and the natural environment can provide not only an aesthetically pleasing setting but also restorative experiences, including a positive effect on health. Gifford identified the following main benefits of nature: cognitive freedom, escape, the experience of nature, ecosystem connectedness, growth, challenge, guidance, sociability, health, and self-control. The most important seem to be the sense of freedom and control felt in nature, in contrast to an urban environment, which is often perceived as constraining.

Behavioral zoning, which consists of allocating specific spaces to specific activities, allows access to be controlled and canalizes the flux of visitors, thus avoiding conflicts resulting from different aspirations and contradictory needs (rest/exercise) and reducing the feeling of crowding. Determining the mean distances from habitat to where people perform different leisure activities is important in the planning of neighborhood leisure places.

Concerning children, research has consistently found a lack or absence of places for them to frequent. Gaster proposes guidelines for the planning of proximal spaces for children.

### **6.2. Management of Urban Diversity**

Individuals often group spontaneously or are grouped within the urban territory according to social, cultural, ethnic, and economic affinities and form more or less homogeneous districts that tend to singularize within the urban community in its totality. They often express a feeling of belonging and existing as a specific group through urban territorial identification.

Socioenvironmental management of urban diversity means not only encouraging the coexistence of different cultures and fostering harmonious development of all minorities but also ensuring an equitable and just repartition of services in order to provide the different communities with the infrastructures they need. Only cooperation between inhabitants and authorities as well as participation in decision-making encourage the feeling of belonging to a distinct urban community. This is achieved if two aspects are promoted: the public expression of difference and its acceptance and the construction of a common destiny through the feeling of belonging to the urban community. These objectives require intervention with regard to four aspects. First, efficient infrastructure and equitable access to different urban services (health, transport, energy, water, garbage collection, and so on) must be provided, and these services must be recognized as being a common good contributing to the well-being of everyone. Second, the needs, expectations, and aspirations of the different components of the urban population must be identified, regular contact with the different communities (and/or their representatives) of the metropolitan population must be maintained, and the different communities must be informed of any new development and/or improvement of services and infrastructures. Third, a regular and easily accessible system of communication between the town administration and the population must be developed. Finally, the expression of cultural particularities through the educational system, the media, public and cultural activities must be taken into account and encouraged, and mutual recognition among the different communities must be encouraged by promoting intercultural sports or artistic encounters.

Such principles gain significance only if they are articulated together in order to allow the simultaneous

expression of differences and a feeling of belonging to an unique urban community that has cultural identity, on the one hand, and is transversal and metropolitan, on the other hand. Such interventions encourage not only local appropriation and identification but also the sense of belonging to an identified urban community.

Regarding the city as a place of living, creating the conditions of its appropriation leads to identity building in the sense of urban citizenship, which meets the requirements of sustainability insofar as this is, according to the Cities, Identity, and Sustainability model proposed by Pol, a condition of people's engaging more readily in pro-environmental behavior.

## 7. TOWARD SUSTAINABILITY OF URBAN ENVIRONMENTS

Sustainable development and globalization are likely to have a major effect on the development of urban settlements. Both are key concepts in the economics of the 21st century, inevitably having an impact on our ways of living in general.

### 7.1. Globalization and Sustainability

By defining sustainable development as development capable of satisfying the needs of the present generation without compromising the ability of future generations to satisfy their own needs, the Brundtland Report, published in 1987, opened the way to questions related to quality of life. The reference to needs allows not only the requirement for development to be harmonious and respectful to the environment but also that the individual's well-being be recognized. People-centered analyses focusing on well-being raise the crucial questions of attachment, control, and identity or identities.

Globalization and its corollaries, global trade and communication, create pressure toward cultural uniformity in lifestyles. Globalization and the anxieties that it engenders also require the local or regional priorities and particularities, cultural differences, and therefore also specific needs to be recognized. Cultural differences and identity are essential issues with regard to the relationship between people and their environment. Identities give structure and meaning to life. They are at the origin of both conflict and cohesion, and they strongly affect patterns of both exclusion and

solidarity. At both the interindividual and the societal level, gender, religious, and cultural identities generate new forms of solidarity, often on a transnational or transcultural level. Identities are also frequently related to the environment. They shape relationships between people and environment at different spatial levels, depending on appropriation and spatial anchoring.

### 7.2. Sustainable Cities

The requirement for cities to be sustainable raises further challenges. Cities host increasingly culturally heterogeneous populations. The implementation of sustainable development through Local Agenda 21 initiatives will only be possible if the local communities reach a consensus. The current trend aims for proactive consensus-building approaches that attempt to involve people in the decision-making process. Meanwhile, many Third World communities are involved in self-help initiatives, but they are rarely able to participate in wider urban planning or politics. Does the concept of citizenship explicitly impose a single, egalitarian political identity on the disparate and separate identities likely to exist within a population of any significant size? How may the intracommunity differences be recognized and how may they converge in a common urban citizenship (Fig. 2)?

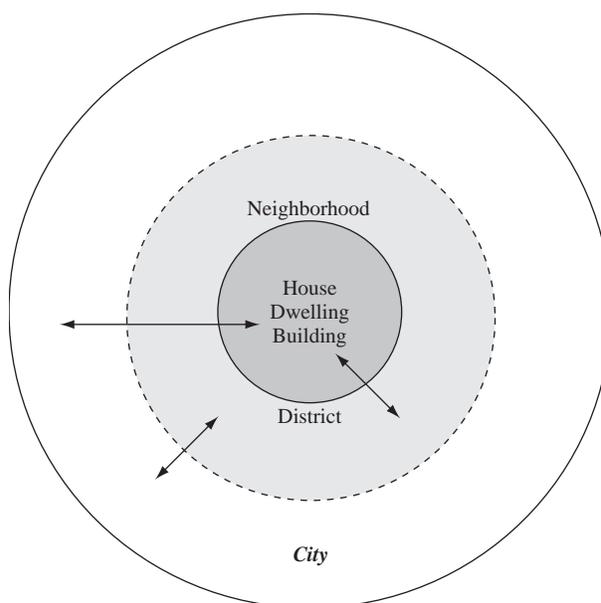


FIGURE 2 From dwelling to citizenship: extensions of appropriation.

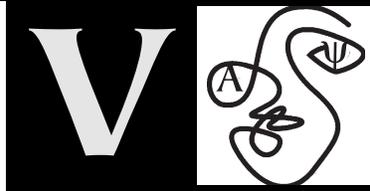
In other words, the challenge for environmental psychology is to provide tools not only able to foster the appropriation at the dwelling or the neighborhood level but also to encourage the appropriation of the city and help the individual to identify with the urban community as a whole. Placing the citizen at the center of public life is an important challenge. Achieving sustainable cities requires strong civic cultures and a new politics of cohesion and cooperation.

### See Also the Following Articles

Environmental Design and Planning, Public Participation in ■ Environmental Stress ■ Person–Environment Fit ■ Privacy ■ Residential Satisfaction and Perceived Urban Quality ■ Territoriality

### Further Reading

- Altman, I. (1975). *The environment and social behavior*. Monterey, CA: Brooks/Cole.
- Altman, I., & Low, S. M. (1992). *Place attachment. Human behavior and environment: Advances in theory and research* (Vol. 12). New York: Plenum.
- Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.
- Canter, D. (1977). *The psychology of place*. London: Architectural Press.
- Gifford, R. (1997). *Environmental psychology: Principles and practice* (2nd ed.). London: Allyn and Bacon.
- Kaplan, S. (1995). The restorative benefits of nature—Toward an integrative framework. *Journal of Environmental Psychology*, 15(3), 169–182.
- Krupat, E. (1985). *People in cities: The urban environment and its effects*. Cambridge, MA: Cambridge University Press.
- Low, S. M., & Altman, I. (1992). *Place attachment*. New York: Plenum.
- Milgram, S. (1970). The experience of living in cities: A psychological analysis. In F. F. Kortan, S. W. Cook, & J. I. Lacey (Eds.), *Psychology and the problems of society*. Washington, DC: American Psychological Association.
- Moser, G. (1992). *Les stress urbains*. Paris: Colin.
- Moser, G., Pol, E., Bernard, Y., Bonnes, M., Corraliza, J., & Giuliani, V. (2002). *People, places & sustainability*. Göttingen, Germany: Hogrefe & Huber.
- Moser, G., & Uzzell, D. (2003). Environmental psychology. In T. Millon, & M. J. Lerner (Eds.), *Comprehensive handbook of psychology. Vol. 5: Personality and social psychology* (pp. 419–445). New York: Wiley.
- Moser, G., & Weiss, K. (2003). *Espaces de vie: Aspects de la relation homme–environnement*. Paris: Colin.
- Preiser, W., & Ostroff, E. (2001). *Universal design handbook*. New York: McGraw-Hill.
- Proshansky, H. M., Ittelson, W. H., & Rivlin, L. G. (1970). *Environmental psychology: Man and his physical setting*. New York: Holt, Rinehart and Winston.
- Simmel, G. (1957). The metropolis and mental life. In K. H. Wolff (Ed. and Trans.), *The sociology of Georges Simmel*. Glencoe, IL: Free Press.
- Stokols, D., & Altman, I. (1987). *Handbook of environmental psychology*. New York: Wiley.
- Wicker, A. W. (1979). *An introduction to ecological psychology*. Monterey, CA: Brooks/Cole.



# Values and Culture

Peter B. Smith

University of Sussex, Brighton, United Kingdom

1. Introduction
  2. Levels of Analysis
  3. Individual-Level Dimensions of Values
  4. Culture-Level Dimensions of Values
  5. Conclusion
- Further Reading

Comparisons of values have provided one of the most useful ways of understanding cultural differences. This is because there is less variation between different cultures in the meaning of particular values than there is in the meaning of specific behaviors. When individuals are studied, there is evidence of universal linkages between particular values and certain self-concepts, attitudes, and behaviors. However, within different nations the array of those holding particular values is not the same. By studying culture-level value differences, the context that influences individuals within each culture can be better understood.

## GLOSSARY

**country-level analysis** Studies in which values data for each nation are aggregated so that each country's data become a single data point.

**ecological fallacy** The belief that because two variables are correlated in a particular way at one level of analysis, they must show the same relationship to each other at another level of analysis.

**individual-level analysis** Studies in which variations in values between individuals within a population are examined.

**response bias** A tendency to consistently use the extremes (or the midpoint) of scales measuring agreement or disagreement with questionnaire items, regardless of item content.

**value dimension** A continuum summarizing the relative endorsement of values or value types that contrasts some value types with other value types.

**values** Abstract generalized beliefs about desirable goals and about ways of achieving those goals.

**value type** A group of individual values shown empirically to cluster together.

## 1. INTRODUCTION

The study of values has steadily become more central to cross-cultural psychology as it has developed over the past several decades. This is largely a consequence of the way in which many researchers have chosen to conceptualize cultural differences. Early researchers placed emphasis on identifying differing behavior patterns and treating these as indicative of cultural difference. However, no generalizable patterns of behavioral difference were identified. Recently, it has been noted that similar behaviors may be interpreted in quite different ways within different cultural groups. Thus, cultural difference may depend more directly on the ways in which behaviors are interpreted, evaluated, and given meaning than on their simple presence, absence, or

frequency. If this perspective is adopted, then the study of values becomes of central importance since a person's values comprise an abstract generalized template, in terms of which all events, experiences, and possibilities are likely to be evaluated.

Individuals may endorse many values and their actions in any particular situation may well depend on the relative strength and relevance of each value to that situation. Thus, if values are to provide a secure theoretical basis for cross-cultural psychologists, they must be surveyed comprehensively, rather than in terms of some specific value that achieves a fashionable status among researchers.

Values are typically measured by providing respondents with a list of single words or phrases representing abstract values, such as honesty, authority, kindness, and so forth, and asking how strongly each value is endorsed, using rating scales. This procedure follows the pioneering work by Rokeach in the United States, who asked respondents to rank a list of values in order of their importance. However, some of the researchers whose work is surveyed in this article relied in part on respondents' statements evaluating specific behaviors and expressing attitudes toward hypothetical events, from which they inferred the presence of particular underlying values.

Before measures of values can be used to understand cultural difference, there are two key methodological problems that must be resolved. First, it is necessary to take account of the different ways in which respondents in different cultures respond to questionnaires. Members of some cultural groups show different types of response bias than do others. For instance, Hispanic respondents give more extreme ratings than Caucasian Americans. There may also be differing norms as to whether one should present oneself modestly, as in Japan, or assertively, as in the United States. Several procedures for correcting this type of bias have been proposed. The most widely used are centering, in which each score is subtracted from the respondent's overall mean, and within-subject standardization, in which centered means are also divided by the standard deviation. Some critics argue that these types of standardization actually eliminate variance that is substantive rather than artifactual. An alternative strategy is test hypotheses both with and without standardization. If hypotheses are more strongly supported after standardization, the argument is strengthened that bias was artifactual.

The second problem to be overcome is to ensure that the particular values included in a questionnaire are interpreted in equivalent ways within each of the cultures being sampled. This requirement has often been neglected. Equivalence can be tested by determining

whether endorsement of the various values is intercorrelated in a similar way within each of the cultures. For instance, if endorsements of "love," "friendship," and "honor" show a similar pattern of relationships in each culture, the chances are increased that these concepts have equivalent meanings across cultures.

## 2. LEVELS OF ANALYSIS

Most psychological studies concern the behavior of individuals. Values are also measured by seeking responses from samples of individuals. It may seem paradoxical that cross-cultural psychologists should have found it useful to survey values in order to understand cultural differences, when these differences are not simply between individuals but between large aggregates of individuals. The distinction between culture-level analysis and individual-level analysis is crucial to cross-cultural psychologists, and it is frequently not well understood. Hofstede was among the first to note its importance. In identifying the four value dimensions along which he classified national cultures, he used data aggregated to the culture-level, not individuals' responses. Thus, each nation's data became one case, not many. Hofstede maintains that his dimensions are applicable only to the interpretation of cultural phenomena, and not to the behavior of individuals; however, many others have used his concepts for individual-level analyses.

There is thus a prevailing cultural milieu in which each individual resides. Any one individual may have values that are congruent with that milieu or that diverge from it. Furthermore, two values that can be found together within a cultural milieu may be such that any one individual could not endorse both. For instance, a culture may have prevailing values favoring both authority and humility, but any one individual will have to either seek authority or endorse humility. To assume that variables that are positively correlated at one level of analysis will also be similarly correlated at the other level is to commit the so-called ecological fallacy.

The relations of culture-level values to other variables may also differ from those found at the individual level. For instance, scores on Hofstede's culture-level individualism variable correlate positively across 45 nations with mean scores for job satisfaction. In contrast, individual-level analysis in Hong Kong revealed a negative correlation between high job satisfaction and individualist values. Therefore, when surveying studies of values across cultures, levels of analysis must be distinguished.

### 3. INDIVIDUAL-LEVEL DIMENSIONS OF VALUES

The principal attempt to identify individual-level dimensions of values that are replicable cross-culturally is that by Schwartz, whose approach derives from that pioneered by Rokeach. Schwartz grouped 56 single values into value types according to their common goals. He reasoned that the basic human values likely to be found in all cultures are those that represent universal requirements of human existence (biological needs, requisites of coordinated social interaction, and demands of group functioning) as conscious goals. Drawing on the values identified by previous researchers and writers, he grouped them into the following motivationally distinct value types:

1. Power—social status, dominance over people and resources
2. Achievement—personal success according to social standards
3. Hedonism—pleasure or sensuous gratification
4. Stimulation—excitement and novelty
5. Self-direction—independence of thought and action
6. Universalism—understanding, tolerance, and protection for the welfare of all people and nature
7. Benevolence—preserving and enhancing the welfare of people to whom one is close
8. Tradition—respect and commitment to cultural or religious customs and ideas
9. Conformity—restraint of actions and impulses that may harm others and violate social expectations
10. Security—safety and stability of society, relationships, and self

Schwartz then postulated that actions taken in pursuit of each value type have consequences that may conflict or may be compatible with the pursuit of other types. For example, pursuing achievement values may conflict with pursuing benevolence values: Seeking personal success may prevent enhancing the welfare of others. The theoretical pattern of conflict and compatibility among value types is portrayed in Fig. 1. Competing value types are placed opposite one another, with more complementary types in closer proximity.

Schwartz confirmed this structure empirically, revealing that the 10 value types can be summarized as the two bipolar dimensions also shown in Fig. 1: Openness to Change (including self-direction and

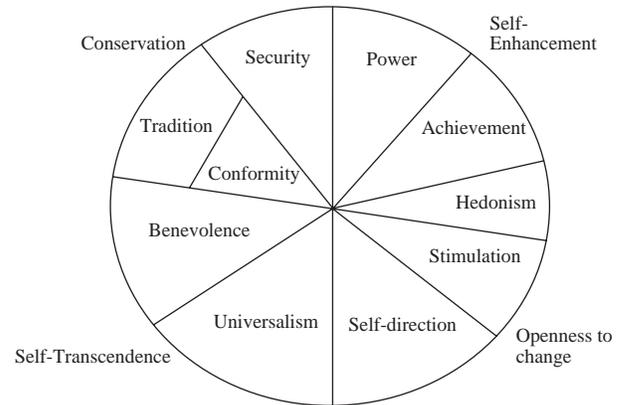


FIGURE 1 Schwartz's value types.

stimulation) versus Conservation (security, conformity, and tradition) and Self-Enhancement (power and achievement) versus Self-Transcendence (universalism and benevolence). Hedonism includes elements of both Openness to Change and Self-Enhancement.

Schwartz's data were collected using a survey comprising 56 values selected to represent each of the value types. The Schwartz Value Survey (SVS) has been administered in more than 60 nations, sampling students and urban schoolteachers. Respondents rate the importance of values as guiding principles in their lives. Correspondence between theorized and actual structure of value types was assessed by comparing two-dimensional spatial representations of the inter-correlations among the values to the theoretical spatial arrangement in Fig. 1. Separate analyses have been reported for more than 100 samples. These analyses support the near universality of the 10 value types and of their structural relations. The two bipolar dimensions are found in almost every sample. The Portrait Values Questionnaire comprises verbal portraits of persons exemplifying the 10 value types. The structure of values found earlier has proved replicable using this measure within samples from Israel, Uganda, Italy, and South Africa. In addition to the uniformity found in the structure of values, the SVS data set also reveals that certain value types are universally more strongly endorsed than others. In particular, benevolence and self-direction are favored over power and tradition.

Further analyses of SVS data have indicated some specific variations in the meaning of individual values

between nations. For instance, in Japan “true friendship” correlated strongly with security values, whereas in most other samples it correlated better with benevolence values. Australians had a rather different conception of loyalty and responsibility than respondents from other nations. In Spain, “honor” was associated with family and social interdependence, whereas in The Netherlands it was associated with self-achievement and autonomy. These variations in the overall results are valuable in identifying what may be distinctive about particular cultures or regions. However, the principal benefit from the individual-level Schwartz project is that it provides a valid measure with which to test whether values have similar correlates in different cultures.

### 3.1. Correlates of Individual-Level Values

Young people in a wide variety of nations attribute greater importance to Openness to Change and less importance to Conservation than their elders. More educated people show the same value patterns as younger people. A meta-analysis of samples from 47 countries found that women more strongly favor security and benevolence values, whereas men favor self-direction, stimulation, hedonism, achievement, and power values. Differences are strongest for the benevolence and power value types. However, the magnitude of gender differences is much greater in some cultures than in others. The Chinese Value Survey data from 23 nations also found numerous culture by gender interactions. Even the largest gender differences are smaller than differences in values typically found in relation to age, education, and national culture.

### 3.2. Attitudes and Behavior

If individuals' values can be shown to consistently predict attitudes and behaviors in a variety of different cultures, then the case is strengthened for using variations in individual values to interpret cultural differences. If the relations between individual values and other phenomena are found to vary between cultures, then other ways should be sought to explain differences between cultures. For instance, it might be found that acting consistently with one's personal values is esteemed in some cultures, whereas conforming to social expectations is preferred in others.

A country-level meta-analysis of Asch conformity studies conducted in 17 nations did find such variations. These variations might enhance or weaken the values-behavior link. At the individual level, values have been compared with the Crowne-Marlowe measure of need for social approval in Finland and Israel. Despite the differences between these nations, need for social approval was consistently positively correlated with conformity values and negatively with hedonism and stimulation.

Religious commitment has been studied among Muslims, Jews, and Catholic, Orthodox, and Protestant Christians in 13 countries. Religiosity is consistently correlated most positively with tradition values and most negatively with hedonism and stimulation. The correlations are stronger in Western Europe than in the former communist states of Central and Eastern Europe. Religiosity and universalism values are more negatively correlated in the Orthodox and Muslim samples than elsewhere. Thus, there is evidence for consistent linkages between values and religiosity, but also some effects attributable to variations in church-state relations.

Consistency has also been found between SVS values and support for particular political parties. In Hungary, Israel, The Netherlands, and Spain, the values in opposing positions in the SVS predicted support for opposing political parties. In the United Kingdom, Japan, New Zealand, and the United States, universalism values correlated positively with anti-nuclear activism, whereas power values correlated negatively. Correlations for self-direction with universalism were mostly positive and those for security and conformity mostly negative. These results indicate the importance of universalism and self-direction versus power, security, and conformity for predicting political attitudes and behavior.

SVS scores have been compared with a self-concept measure among students in Singapore and the United States. Consistent links were found between power values and vertical individualism, self-direction values and horizontal individualism, conformity values and vertical collectivism, and benevolence values and horizontal collectivism.

Preferred styles of managing conflict among management students in India, the Philippines, Hong Kong, and the United States suggested that respondents' individual-level Conservation scores mediated their preference for avoiding, whereas Self-Enhancement and Openness to Change mediated preference for competing.

The studies reviewed in this section indicate that there is a growing body of evidence of cross-national consistency between endorsement of individual values and a variety of demographic indicators, self-concept, attitudes, and social behaviors. This enhances the possibility of explaining some cultural differences in terms of the differing spectrum of values endorsed by individuals from different areas of the world. However, such explanations will always be no more than partial since they ignore the likely effects of culture-level variables.

#### 4. CULTURE-LEVEL DIMENSIONS OF VALUES

The average values that are found within a particular nation, or within discrete cultural groupings that comprise a nation, define a context within which individuals are born and socialized. No national culture is entirely homogeneous, and contemporary nations are characterized by a rich variety of ethnic, educational, occupational, professional, and regional subcultures. Consequently, it is normal rather than exceptional that there should be divergence between individuals' values and the average values found to characterize those individuals' nations as a whole. Although psychologists have mostly preferred explanations of human behavior in terms of variables derived from the person, attention to culture-level variables opens up some of the more neglected situational determinants of behavior. The values espoused by most of those around us will mould and temper the ways in which we act on our own values.

It might be argued that nations are products of history and politics, and that many nations would fail to satisfy a definition of culture that uses shared evaluations of behavior as a defining criterion. However, modern nations are characterized by numerous mechanisms by which they seek to create and maintain cultural homogeneity, including education, legislation, language policies, and the mass media. Almost all researchers of culture-level phenomena have equated cultures with national cultures.

There have been numerous attempts to characterize national cultures. The current discussion highlights those that have focused most directly on values.

The pioneer study was Hofstede's well-known study of IBM employees from 50 nations and three regions; Hofstede provided an updated account in 2001. Of the four culture-level dimensions identified by Hofstede,

individualism–collectivism has proved by far the most influential. This article gives greater attention to the three remaining dimensions. The four dimensions were determined by aggregating responses within each nation to each questionnaire item and then factor analyzing the resulting country scores.

Since the original questionnaire was not designed on the basis of any explicit theory, it is perhaps not surprising that the resulting factors comprised responses to a mixture of items. High scores on Power Distance indicated, for instance, that respondents reportedly feared to express disagreement with their manager, perceived their boss as autocratic or paternalistic, and did not want a boss who is consultative. High scores on Individualism versus Collectivism were defined by emphases on work goals of freedom to work in one's own way, sufficient time for personal life, and challenging work versus work goals of good working conditions, opportunities for training, and full use of skills and abilities. High scores on Uncertainty Avoidance reflected high reported job stress, beliefs that company rules should not be broken, and expectations of continuing to work for IBM for a long time. High scores on Masculinity versus Femininity reflected emphases on earnings, recognition, advancement, and challenge in one's work versus good relations with the boss and others, living in a desirable area, and having high job security. The items defining Individualism–Collectivism and Masculinity–Femininity are the only ones that can be considered as direct measures of values, but Hofstede asserts that the other dimensions do reflect underlying values. He has sought to refocus attention on the more neglected of his dimensions by the publication of a substantially revised edition of his 1980 book, which gives equal attention to all dimensions. In the new edition, his arguments are supported by citation of the large number of studies relevant to each dimension that have been published during the past two decades. He compares country scores on a variety of indices with value scores for the relevant nations. However, few of these studies have actually attempted to replicate the Hofstede measurements. The validity of all such comparisons rests on the validity of measurement in the original study and on the degree to which those measures continue to accurately represent particular nations.

Michael Bond was the first to test the replicability of Hofstede's dimensions. He constructed a Chinese Value Survey (CVS) intentionally based on Eastern rather than Western values by asking Chinese scholars to nominate values "of fundamental importance in Chinese culture."

Sampling students from 23 countries, a country-level factor analysis yielded four orthogonal factors, three of which correlated well with the Hofstede scores for the 20 overlapping nations. No CVS factor correlated with Uncertainty Avoidance. Bond called his fourth, unique CVS factor Confucian Work Dynamism, and Hofstede adopted this as an added value dimension of cultural variation missing from his theory. Hofstede interpreted it as opposing a dynamic orientation toward the future to a more static orientation toward the past and present and therefore renamed it Long-Term Orientation. He suggests that Western cultures are preoccupied with the search for Truth so that the emergence of an Uncertainty Avoidance factor from a survey designed by Western psychologists is to be expected. In contrast, he suggests that Pacific Asian cultures are preoccupied with the search for Virtue, which might explain a greater emphasis on Long-Term Orientation or Confucian Work Dynamism among the values identified by Chinese scholars.

Schwartz's individual-level analyses have already been described. By aggregating his individual-level data, he has also identified culture-level variations in values. Only those 45 values demonstrated to have relatively stable cross-cultural meanings in studies within each nation were included. Multidimensional scaling was again used. At the culture level, he argues that the content of cultural value dimensions reflects alternative solutions to three basic societal issues:

1. Relations between individual and group: To what extent are persons autonomous versus embedded in groups?
2. Ensuring responsible social behavior: How can people be motivated to consider others' welfare and coordinate with them?
3. The role of humankind in the natural and social world: Is it more to submit, to fit in, or to exploit?

The results consistently yielded seven value types. These can be summarized as three bipolar dimensions, reflecting the three issues detailed previously: Conservatism versus Autonomy (divided between intellectual autonomy and affective autonomy), Hierarchy versus Egalitarianism, and Mastery versus Harmony. Nation scores were computed by combining the mean ratings on the values designated a priori as representing each value type. Scores were also computed for each dimension (one polar value type minus the opposing value type). The scores derived from teacher samples correlated well with those obtained from student samples across 40 nations.

A further survey was conducted by Trompenaars, who sampled the work-related values and attitudes of approximately 15,000 organization employees from more than 50 countries. Many of his questionnaire items were based on the earlier theories of the U.S. sociologist, Talcott Parsons. Trompenaars provided country means for responses to individual questionnaire items and suggested several cultural value dimensions, but he undertook no empirical analyses to assess the validity of these dimensions. Smith, Dugan, and Trompenaars selected the items from his questionnaire that were suitable for the empirical analysis of dimensions. These items measured preferences for using universalistic versus particularistic standards in relationships, for granting status on the basis of ascribed versus achieved qualities, for orientation to the interests of the self versus the collective, and for different styles of organizational management. After standardization of individual responses, item means for 43 nations were entered into a multidimensional scaling analysis. Two clear dimensions emerged: Conservatism versus Egalitarian Commitment and Loyal Involvement versus Utilitarian Involvement.

Inglehart has also identified two culture-level value dimensions based on factor analysis of aggregated data from public opinion surveys conducted within 43 nations. The first, Well-Being versus Survival, shows a strong positive correlation with Hofstede's individualism scores and strong negative correlations with Power Distance scores. The second dimension, Secular-Rational versus Traditional Authority, shows similar but weaker correlations.

#### 4.1. Relations among Culture-Level Dimensions

There are five studies that have each sampled values from at least 23 nations. These studies vary in methodological strength. Only Schwartz derived his items from both Western and non-Western sources, controlled for meaning equivalence, and demonstrated replicability of his dimensions across samples. Nonetheless, consistencies across the different studies may point to the most central and reliable dimensions of cultural values. Smith and Schwartz examined the correlations between those country scores that were common to each pair of surveys. They conclude that there is substantial convergence between results, even though different authors use different names for

dimensions of variance in values. The correlations point to a close positive association between two basic dimensions identified in different ways by different authors:

1. The preferred cultural view of individual–group relations (autonomous versus embedded)
2. The preferred cultural mode of motivating responsible social behavior and allocating resources (negotiation among equals versus acceptance of unequal, hierarchical roles).

These dimensions are virtually identical to the first two culture-level dimensions identified by Schwartz. However, it is unlikely that all the ways in which values vary across cultures have been identified.

#### 4.2. Studies Related to Culture-Level Value Dimensions

As noted previously, categorizing cultures in terms of nations involves substantial arbitrariness. Within samples from large numbers of nations, it is possible that discernible clusters of nations can be identified, within which similar values prevail. Hofstede reports a hierarchical cluster analysis of his data set of 53 nations and regions. He lists 12 “culture areas,” but some of these are less clearly meaningful than the 7 areas identified in his 1980 book, which included only the 40 nations with large samples: Nordic (e.g., Norway and Sweden), Anglo (e.g., the United States, the United Kingdom, and Australia), Germanic, Near Eastern, less developed Asian, less developed Latin, and more developed Latin. Japan stood alone as a more developed Asian nation. Schwartz conducted multidimensional scaling analyses using nation scores on his seven culture-level value types using data from both his teacher and his student samples. Six relatively similar culture areas emerged from both analyses: West European, Anglo, East European, Islamic (e.g., Turkey and Indonesia), East Asian, and Latin American. Japan was again unique. The similarities between the Hofstede and Schwartz culture clusters, and some of the differences, simply reflect the sampling, such as the lack of former Soviet bloc nations in the IBM study and underrepresentation of Islamic countries in both. Other differences may reflect cultural change during the past 20 years. One study including 43 countries found two clear culture areas, one for Central and Eastern Europe and the other for northern Europe. Somewhat less clear Far Eastern, Latin, and Anglo

areas were also discernible. The convergence of these results with those reported by Hofstede and Schwartz is substantial.

We can therefore have some confidence in how to characterize the value preferences of regional groupings throughout the world. It is less clear how great can be the contribution of psychologists to explaining why these differences are as they are. The principal tools that can be employed are correlations and regressions, and at the country level these are most useful when large numbers of nations have been sampled. Furthermore, inferring causation from any relationships found is notoriously unreliable. To take a widely cited example, country gross national product correlates  $+0.85$  with Hofstede’s Individualism and  $-0.65$  with his Power Distance, but these figures do not indicate whether wealth is a cause or a consequence of such values or, indeed, whether the high correlations are a side effect of some unspecified third factor. We know that economic growth during the latter part of the 20th century was correlated  $+0.70$  with scores for Confucian Work Dynamism, but this correlation does not hold for the most recent decade, maybe because political and economic factors are now different.

A statistical technique (regression equations) permits the prediction of a value profile from other variables, such as attributes of the countries that have been studied. This technique has been used to test explanations of value profiles in Western and in Eastern Europe. Hofstede computed regressions that summarized what he called system-level indicators of each of his value dimensions. Moderate latitude, high population, and low wealth explain 56% of variance in countries’ Power Distance scores. Low wealth explains 8% of variance in Uncertainty Avoidance. High wealth and extreme latitude explain 78% of variance in Individualism. High population explains 8% of variance in Masculinity scores.

#### 4.3. Correlates of Culture-Level Values

An alternative procedure is to examine data on the extent of a specific psychological phenomenon across nations and determine whether country-level value scores can account for the variance that has been found. This is a hazardous enterprise since the country scores are derived from samples that differ from those whose behavior is being studied. A statistical technique (meta-analysis) permits the integration of data from

many studies. It has shown that there is greater conformity in countries that score high on Hofstede's Collectivism and low on Schwartz's Intellectual and Affective Autonomy. However, it could be that the behaviors identified as conformity in cultures high on Individualism and Autonomy may be thought of as sensitivity to others in cultures high on Collectivism and Conservatism.

Culture-level values have also been linked to work-related variables. Among managers in 21 nations, role overload is predicted by high Power Distance, and role ambiguity is predicted by low Power Distance. Schwartz found work to be more central to people's lives where Hierarchy and Mastery score high. He also found that the tendency to view work as an obligation is greater in cultures high on Conservatism and Hierarchy. Across 47 nations, it has been found that the sources of guidance on which managers relied most in handling a range of work events could be predicted from country scores derived from Hofstede, Trompenaars, and Schwartz. Hofstede cites numerous additional significant correlations between his dimensions and data derived from opinion surveys, objective indices, and psychological studies. Some of these studies also sampled substantial numbers of nations. It is clear that at the culture level, as at the individual level, value indices have substantial predictive validity.

## 5. CONCLUSION

This article has emphasized the importance of studying values both at the level of individuals and at the aggregated level of nations. Evidence suggests substantial predictive validity at both levels. We can best explain cultural differences if we focus on both levels, while taking care to treat them separately. Difficulty in conceptualizing the interplay of personal and situational influences has bedeviled many areas of psychological investigation. The study of values does not provide a complete basis for understanding cultural differences, but it has proved one of the most promising initiatives. Progress has been made in identifying dimensions of variance, rather than studying single values, and in deriving these dimensions from theory.

As previously noted, nations do not necessarily provide the best basis for cross-cultural comparisons. Many people believe that there are major regional and/or ethnic differences in values within their

country. Multiple samples from single countries are needed to test the validity of these beliefs. Even when within-nation differences are found, it is important to compare their magnitude with cross-national differences. Significant within-nation differences in SVS scores in China, Japan, and the United States are dwarfed by the much larger differences found between these nations. However, value diversity is likely to be greater in some nations than others. High SVS national value consensus is predicted by high socioeconomic development but low political democratization.

The world is currently experiencing rapid social change, driven by globalization, economic development, migration, and the information revolution. These changes may well entail marked shifts in the values endorsed by successive generations. Traditional ways in which values are transmitted from one generation to the next may be under threat. Consistency between results obtained by Hofstede and Schwartz suggests that these changes may be relatively gradual, but researchers have only recently started to map value changes systematically.

## See Also the Following Articles

Attitudes ■ Ideological Orientation and Values ■ Interpersonal Conflict

## Further Reading

- Chinese Culture Connection. (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology*, 18, 143–164.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage.
- Inglehart, R. (1997). *Modernization and post-modernization: Cultural, political and economic change in 43 societies*. Princeton, NJ: Princeton University Press.
- Schwartz, S. H. (1992). Universals in the structure and content of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, 25, 1–65.
- Smith, P. B., & Schwartz, S. H. (1997). Values. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology* (2nd ed., Vol. 3, pp. 77–118). Needham Heights, MA: Allyn & Bacon.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.



# Vocational Assessment in Schools

Edward M. Levinson, Angela B. Christy, and Marcy R. Maus

Indiana University of Pennsylvania, Indiana, Pennsylvania, USA

1. Introduction
  2. Best Practices
  3. Roles for Professionals
  4. Conclusion
- Further Reading

school, different skills are assessed with different assessment techniques, and students are provided with the skills and training necessary to make a successful transition from school to work and to community living.

## GLOSSARY

*transdisciplinary school-based vocational assessment (TVA)* A comprehensive assessment conducted within a school setting whose purpose is to facilitate educational and vocational planning in order to allow a student to make a successful adjustment to work, post-secondary education, and community living.

*transition* Movement of a student from high school to post-secondary training and from home to more independent living.

School-based vocational assessment should incorporate a transdisciplinary approach that includes school and community professionals in the evaluation and assessment of the vocational and educational needs of students, especially those with disabilities. Several steps are involved in program development, including the following: planning and development, implementation, and evaluation and improvement. Once developed, programs should ensure that assessments are conducted at three levels: elementary, middle school/junior high, and high school. As the students progress through

## 1. INTRODUCTION

### 1.1. Importance

Students continue to progress through schools expecting that some job or career awaits them upon graduation from high school. They attend classes to enhance their skills in defined subject areas but ultimately fall short when it comes to possessing the skills necessary to obtain and uphold a gratifying job.

According to the U.S. General Accounting Office, one-third of young workers ages 16–24 do not have the skills they need to perform entry-level, semiskilled jobs. Additionally, nationwide studies conducted by the National Career Development Association and the Gallup Organization suggest that adults who consciously plan and prepare themselves for a job comprise only one-third of those currently working. Consequently, the remaining two-thirds of employed adults enter their jobs as a result of chance, the influence of others, or because they took the only job available to them. Most important, only half of these workers are content with their jobs.

In particular, individuals with disabilities are at especially high risk for failure or dissatisfaction with

their jobs. When compared with their nondisabled peers, individuals with mild disabilities experience a higher rate of unemployment and underemployment, lower pay, and greater dissatisfaction with their jobs. Hence, this population of students is especially in need of vocational assessment services.

Moreover, there is a high drop-out rate among students with disabilities; as a result, they have not received the vocational training necessary to make a smooth transition from school to work. Studies comparing special education drop-out rates with control group drop-out rates (normative data) have consistently demonstrated that students with disabilities leave school more often than students without disabilities. Students with learning disabilities and emotional disabilities are at an especially high risk of dropping out of school. Some studies have reported drop-out rates exceeding 40 and 50%, respectively, for these populations.

This high drop-out rate may be attributed to several factors. Many schools focus their curriculum on college-bound students rather than those who plan to enter the workforce following high school. Statistics indicate that only 15% of freshmen entering high school graduate and obtain a degree from a 4-year college within 6 years of their high school graduation. As a result, focusing resources on college-bound students leaves the majority of non-college-bound students lacking in the skills necessary for success in the workforce. The "21st Annual Report to Congress" reported that students with disabilities were more likely to stay in school and even more likely to be employed after high school if they were provided with vocational training.

The Carl D. Perkins Vocational Education Act and the Individuals with Disabilities Education Act (IDEA) both address vocational assessment in schools. The Perkins Act stipulates that parents must be informed about vocational education available for their children 1 year before they are eligible. The act also requires that parents be provided with information about vocational education, such as eligibility requirements and details about the programs offered. Once enrolled, students should be assessed in several areas, including interests, abilities, and special needs as well as any other special services needed to assist in the transition from school to work. IDEA mandates that services aiding in the transition process be established for students with disabilities to ensure a successful transition from school into work and the community. This law requires that plans for transitioning be initiated by age 14.

The process of establishing vocational assessment programs requires effective planning and a structured,

team-oriented approach. The focus of this article is on the role of school psychologists in vocational assessment since they are the applied psychologists working in schools. School psychologists play an important role in providing transition services to students. In order to effectively and competently conduct a comprehensive vocational assessment, school psychologists need to be aware of career development theory and be knowledgeable about interest and aptitude assessment as well as other empirically based vocational assessment practices.

## 1.2. Vocational/Career Development Theory

Vocational/career developmental theory should be the basis for developing school-based vocational assessment programs. Vocational development theory also assists educators in making developmentally appropriate recommendations for students.

Vocational developmental theory suggests that individuals progress through stages. The growth and exploration stages span the school years and are each characterized by three substages. Within the growth stage, children progress through the fantasy substage (ages 0–10 years), in which they use their imagination to take on different career roles; the interest substage (ages 11 and 12 years), in which they consider different careers and areas of interest; and the capacity substage (ages 13 and 14 years), in which they become aware of career demands and the academic training required of careers following high school. The exploration stage consists of the tentative, transition, and trial substages. The tentative substage (ages 15–17 years) is a period during which the adolescent considers various career options and chooses a tentative career goal. During the transition substage (ages 18–21 years), an individual chooses a career and works toward acquiring the skills needed to enter into that career. The trial substage (ages 22–24 years) involves the attainment of a job that fits their career choice. One purpose of vocational assessment is to make certain that students are successfully progressing through these vocational developmental stages and to determine whether students require assistance in reaching the goals associated with each stage.

## 1.3. Skills Needed for Employment

In addition to understanding various aspects of one's job, there are several other characteristics that promote success in the work environment. For instance, employees

need to work well with others (social skills) and to make informative/effective decisions (decision-making skills) in the workplace.

An inclusive vocational assessment evaluates an individual's level of functioning in several domains. It should include measures of one's psychological, social, educational, physical, and vocational abilities. Reviewing records, conducting observations, and administering paper-and-pencil tests assist psychologists in gathering relevant information in these domains. In addition, performance tests, work samples, and situational assessments are also used in a comprehensive vocational assessment program. A performance test focuses less on language and more on an individual's ability to perform a specific task related to a job. Work samples consist of three phases (demonstration, training, and assessment) and expose the student to natural job responsibilities. Situational assessment measures an individual's interests, abilities, and work habits in actual or contrived work environments.

## 2. BEST PRACTICES

### 2.1. Systemic vs Individual Planning and Program Development

A school-based vocational assessment program is designed to meet the educational needs of all students, especially students with disabilities. Development and implementation of a program should be composed of three phases: planning and development, implementation, and evaluation and improvement.

Planning and development is composed of several steps. First, a representative team of school and community personnel likely to be involved in the assessment process or otherwise engaged in the program is formed. Second, a needs assessment is conducted, which involves identifying school and community resources/services to include in the program as well as recognition of possible complications associated with vocational planning. Following a needs assessment, an assessment model should be developed that incorporates local resources and reduces possible obstacles to programming. Fourth, formal interagency agreements and action plans should be arranged between school and community agencies. The purpose of these agreements is to identify what services will be provided by each agency, which students will receive services, and when services will be initiated. Last, if it

is necessary to secure additional funds, the team finds sources of financial support for the program.

The implementation phase should begin with the hiring of a vocational assessment program coordinator. The coordinator is responsible for supervising the implementation of the program and locates sites where evaluations will be conducted. In addition, this individual should develop a procedures manual, decide on and purchase necessary materials, and train involved personnel. He or she is also responsible for conducting in-service workshops to familiarize school and community personnel with the assessment program. The final stages of the implementation phase involve a pilot testing of the program with a target group of students followed by revisions and full implementation of the program.

The first two steps of the evaluation phase acknowledge areas of the program in need of evaluation and identify standards for the evaluation process. Then, a program evaluator is hired to conduct the evaluation. Lastly, the vocational assessment team plans and implements necessary improvements. Program evaluations and revisions are conducted periodically to improve program functioning.

### 2.2. Assessment Models

There are several levels of assessment in school-based vocational assessment programs. Each level has a different purpose, answers different questions, uses different evaluation techniques, and gathers different kinds of information. The type and reason for assessment at a given level are based on developmental objectives and referral questions. Programs are generally composed of two or three levels of assessment. Level 1 assessments are conducted during the elementary school years. These evaluations focus on the child's understanding of self, interpersonal skills, and decision-making skills. In addition, the assessments include vocational and career exploration activities and are intended to improve self-awareness. Level 2 assessments occur in the middle school or junior high school years. The evaluation focuses on student interests, abilities, work habits, and career maturity. The vocational assessment team consults student records, conducts interviews and observations, and evaluates students with standardized norm-referenced assessment tools. The goal of level 2 assessments is to maintain and encourage career exploration and to assist students in formulating tentative career goals. Based on the results, vocational training programs are recommended for students based on their interests and abilities. In addition, program adaptations and

support services are recommended to increase student success in their selected program. Level 3 assessments are conducted in high schools. In order to evaluate high school students, the team utilizes a variety of assessment tools including work samples and situational assessments to determine the training needed to attain post-high school education or employment goals. Evaluation at this level also assists in determining the skills an individual needs to make a successful transition from school to work, post-secondary educational training, and community living.

Transdisciplinary school-based vocational assessment (TVA) is one program model that can be used as a framework for the planning and establishment of a school-based program. The model is equally applicable for students with or without disabilities. TVA uses assessments and interventions as well as involvement from school and community professionals to plan and develop a program. The model consists of four phases and involves two levels of assessment. Phase 1 is the development of a school-based vocational assessment program, in which the school and community team plans, organizes, and implements the program. During phase 2, students in grades 6–9 are administered a level 1 vocational assessment to determine educational and vocational goals, which are then incorporated into students' Individual Education Plans (IEPs). The data from the evaluation are also used to assist team members in identifying vocational training opportunities and future residential living options, making necessary curriculum changes, and connecting students to available school or community resources. Phase 3, designed for students in grades 10–12, incorporates a level 2 vocational assessment that is designed to assist the team in making revisions to a student's program and IEP. Phase 4 is designed to monitor training and placement in a job, a post-secondary school, and/or a residential living facility. Follow-up and continual support from the team are provided during this phase in order to aid in a successful transition from school to work and community living.

### **3. ROLES FOR PROFESSIONALS**

There are several roles that the school psychologist can assume when developing or implementing a vocational assessment program. These vary based on existing vocational services in the district and the expertise of other school personnel involved in the assessment process. Ideally, the school psychologist is one of several school- and community-based professionals on the

vocational assessment team. As a team member, he or she collects information about a student's ability, academic functioning, and social performance and considers its implications from a vocational perspective. Additional measures of vocational interests and aptitudes, an evaluation of student work, an observation in the work environment, and student interviews may also be conducted when assessing the vocational needs of an adolescent. The school psychologist may also use his or her knowledge of measurement principles to aid in the selection of appropriate assessment tools and techniques to be used in the evaluation process. Lastly, the school psychologist's understanding of learning theory and adolescent psychology allows him or her to serve as a consultant for others involved in the vocational assessment and planning process.

If a school-based vocational assessment program has not already been developed, the school psychologist may take on the responsibility of developing or assisting in the development of a program. Since a comprehensive vocational assessment program involving a team of professionals may not be possible in some schools, the school psychologist may be the sole individual involved in conducting vocational assessments.

### **4. CONCLUSION**

Unfortunately, conventional vocational and transition assistance has not resulted in desirable outcomes for students with disabilities. The high unemployment and underemployment rates as well as the high drop-out rates associated with this population suggest a need for a change in focus from an academic to a vocational orientation. Vocational assessment programs in schools are valuable in preparing students with the skills necessary to acquire and maintain a satisfying job and to make a successful transition to adult life. Vocational assessment supplies team members with the information necessary to help students develop educational and vocational goals and assists students in receiving the training necessary for a successful transition from school to work and community living. The program allows for active participation from school and community professionals, and it includes an evaluation of the student's psychological, academic, physical, vocational, and social functioning.

School psychologists play a vital role in the vocational assessment process. Their psychometric expertise is useful in the selection of appropriate evaluation tools, and their knowledge of measurement ensures the accurate interpretation of data. Furthermore, their

understanding of learning and behavior theory and adolescent psychology is valuable in the development and implementation of the program.

### See Also the Following Articles

Career Counseling ■ Holland's Theory (Vocational Personality Types) ■ Indecision, Vocational ■ Vocational Interests

### Further Reading

- Brown, D., & Minor, C. (Eds.). (1989). *Working in America: A status report*. Alexandria, VA: National Career Development Association.
- Clark, G. B. (1998). *Assessment for transitions planning: A guide for special education teachers and related service personnel*. Austin, TX: PRO-ED.
- Hoyt, K. B., & Lester, J. L. (1995). *Learning to work: The NCDA Gallup survey*. Alexandria, VA: National Career Development Association.
- Levinson, E. M. (1993). *Transdisciplinary vocational assessment: Issues in school based programs*. Brandon, VT: Clinical Psychology Publishing.
- Levinson, E. M. (1995). Best practices in transition services. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology III*. Washington, DC: National Association of School Psychologists.
- Levinson, E. M. (1998). *Transition: Facilitating the post-school adjustment of students with disabilities*. Boulder, CO: Westview.
- Levinson, E. M., & Brandt, J. (1997). Children and career development. In G. Bear, K. Menke, & A. Thomas (Eds.), *Children's needs II*. Washington, DC: National Association of School Psychologists.
- Levinson, E. M., & Murphy, J. (1999). Transition and school psychology. In S. DeFur, & J. R. Patton (Eds.), *Transition and school-based services: Interdisciplinary perspectives for enhancing the transition process*. Austin, TX: PRO-ED.
- Phelps, L. A., & Hanley-Maxwell, C. (1997). School-to-work transitions for youth with disabilities: A review of outcomes and practices. *Review of Educational Research*, 67, 197–226.
- Power, P. W. (2000). *A guide to vocational assessment* (3rd ed.). Austin, TX: PRO-ED.
- Smith, F., Lombard, R., Neubert, D., Leconte, P., Rothernbacher, C., & Sitlington, P. (1995). *Position paper of the Interdisciplinary Council on Vocational Evaluation and Assessment*. Available at: <http://www.vewaa.org/council.html>.
- Super, D. E. (1957). *The psychology of careers*. New York: Harper.



# Vocational Interests

Fred H. Borgen

Iowa State University, Ames, Iowa, United States

1. History: Eighty Years of Interest Measurement
  2. Two Types of Scales
  3. Major Interest Inventories
  4. Quality and Use of Interest Measures
  5. Trends in Interest Measurement
  6. Conclusions
- Further Reading

## GLOSSARY

**career self-efficacy** One's belief that one can successfully perform the behaviors required in a particular career-related domain, such as mathematics or sales.

**contrasted-groups scale** A scale that is built by identifying items responded to differently by a criterion group and by a contrast group (or norm group). The best known examples are the clinical scales of the MMPI and the occupational scales of the Strong Interest Inventory.

**Holland's hexagon** John Holland suggested that there are six major interest dimensions or types that can be arrayed in order around a hexagon, with interests that are nearer on the hexagon (e.g., investigative and artistic) being more similar psychologically and interests that are across the hexagon (e.g., investigative and enterprising) being the most different psychologically.

**homogeneous content scale** A scale designed to have items of similar content that correlate with each other. All interest inventories measuring Holland's RIASEC dimensions or types use content scales that represent the broad content within each type. There are also more narrowly focused content scales, such as the Basic Interest Scales of the

Strong Interest Inventory. Content scales have high internal consistency reliability.

**occupational scale** First created by Strong, a contrasted-groups scale containing items empirically differentiating people in an occupation from a general group of people; high scores show relative similarity to the distinctive likes and dislikes of people in that occupation. Such scales typically have a heterogeneous mix of items; hence, their linguistic content is difficult to interpret.

**RIASEC** Acronym for Holland's six interest dimensions or types: realistic, investigative, artistic, social, enterprising, and conventional.

**vocational interest** An emotional reaction to a career-related stimulus, ranging from distinctly liking to distinctly disliking.

Beginning more than 80 years ago, the assessment of vocational interests is one of the great success stories of applied psychology. Millions of vocational interest inventories are used every year, usually by young people exploring career and educational options and choices. Increasingly, they are also used by more experienced workers who are evaluating their current situations or exploring career transitions. Interest inventories have a long and vigorous history in psychology, marked by intensive and groundbreaking research, technical innovation, theoretical consolidation, and adaptation to changing societal needs. Today, interests play as important a role as ever in assessing individuality, especially in concert with other domains, such as abilities, personality, and self-efficacy. Hundreds

of vocational interest inventories exist, but the following five (in alphabetical order) can be considered the most widely used and well researched: the Campbell Interest and Skill Survey, the Kuder Occupational Interest Survey, the Self-Directed Search, the Strong Interest Inventory, and the UNIACT interest inventory.

## **1. HISTORY: EIGHTY YEARS OF INTEREST MEASUREMENT**

Two dominant themes in the long history of vocational interests are E. K. Strong's resolute empiricism and John Holland's seminal theory. Strong mapped the interests of people in various occupations, creating a large and still growing database. Holland focused the interpretation of increasingly complexity of interest data with a theory that makes interests more understandable and accessible. The traditions originated by Strong and Holland continue to profoundly influence research and reflection about the meaning and measurement of vocational interests.

### **1.1. Strong's Empirical Success**

The first widely used interest inventory appeared in 1927 with Strong's construction of the Vocational Interest Blank at Stanford University. Strong's innovative approach was intensely empirical, constructing occupational scales consisting of those interest items that men in a specific occupation answered differently than "men-in-general." Strong was passionate in cataloguing interests, often following the careers of people over several decades. His classic 1943 book best exemplifies his data-driven approach to interest assessment. Strong's insights in this book presage developments that might be considered cutting edge today. Despite the limits of the technology of his day, Strong used pioneering methods such as factor analysis to examine the structure of interests. From this he constructed a classification of occupations that foreshadowed Holland's famous RIASEC hexagon.

### **1.2. Holland's Influential Theory**

Holland has been the leading theorist in vocational psychology during the past 40 years. His novel approach asserts that interests are an expression of personality. He seems to say that interests and personality are synonymous, yet at the same time he is well-known

for a widely used interest inventory, the Self-Directed Search. Moreover, his interest inventories are similar in many ways to others that do not claim to measure personality. A current appraisal might be that interests and personality are treated as separate domains by most theorists and counselors, although there is evidence that they overlap in many ways.

Holland proposed that there are six primary dimensions of vocational interests, characterized by his famous RIASEC acronym: (i) Realistic interests involve building and mechanical activities; (ii) investigative interests encompass science and mathematics; (iii) artistic interests include art, music, dramatics, and writing; (iv) social interests entail helping and teaching; (v) enterprising interests involve managing, selling, and persuading; and (vi) conventional interests include organizing and evaluating data. Holland derived these dimensions from his own empirical work and from the research literature, including Strong's 1943 book. These dimensions quite adequately map the primary broad dimensions of vocational interests.

Holland and his associates' insight of the early 1970s that the RIASEC dimensions (or types) array themselves on a hexagon made Holland's approach very popular in both theory and applications. The hexagonal arrangement permits a Holland calculus based on relations around the hexagon. Distance on the hexagon translates into psychological distance: For example, scientists (I types) are across the hexagon from salespeople (E types), and these people differ on an array of psychological dimensions, including interests. The RIASEC hexagon provides an efficient summary of the major dimensions of interests and their relations to each other. Therefore, a number of popular interest inventories have incorporated the RIASEC dimensions. Many of these inventories measure level of interest on these six dimensions, and many use the Holland themes as an organizing principle. Holland also generates a number of theoretical propositions from the hexagonal relationships, and these have stimulated considerable research during the past four decades.

### **1.3. Interest as a Construct often Poorly Defined**

In one sense, interest measurement has been a victim of its own success. Strong's empirical occupational scales garnered much respect and utility so their conceptual limitations were forgiven. Conceptual definition of interests did not frame Strong's occupational scales.

The scales were more in the tradition of dustbowl empiricism, with an occupational scale defined by the items that empirically differentiated the people in that occupation. Consequently, we do not have a well-developed theory of vocational interests that explains their meaning as constructs and their origins and action. Rather, there are nascent developments of such a theory. Recent work, exemplified by that of Savickas and Spokane, has attempted to imbue interest measures with enhanced meaning. Holland's theory is a major step toward understanding interests conceptually. The rise of content scales is another major step toward scales with direct meaning and interpretability. Finally, social cognitive career theory is a third major step toward understanding interests in the context of other variables, such as career self-efficacy, and developmental processes.

## 2. TWO TYPES OF SCALES

### 2.1. Occupational Scales (Contrasted Groups)

Constructing occupational scales by using the contrasted-groups method is Strong's most respected innovation and most enduring legacy. Occupational scales have remained the centerpiece of the Strong Interest Inventory for nearly eight decades. They are the basis for the lasting respect given the Strong inventory for its intense and careful empiricism tracking the likes and dislikes of satisfied workers in national samples of specific occupations. The Strong inventory has been routinely revised by collecting large samples of men and women in diverse occupations.

In the latest revision of the Strong inventory, Harmon *et al.* followed the steps Strong outlined for constructing occupational scales. As had been done from the beginning, separate occupational scales were created for women and men based on the somewhat different interests of women and men in general and sometimes within specific occupations. Thus, Harmon *et al.* created a new occupational scale for female physicists and a new scale for male physicists. Using the female physicists as an example, the steps can be described. The rationale for an occupational scale is to differentiate people in that occupation from a general reference group. High scores on a scale indicate that the person shares the likes and dislikes that differentiate people employed in that occupation from people in general. Thus, in this contrasted-group method, the responses of 348 female physicists

were compared to those of other women on each of the 317 items in the 1994 Strong. Those items for which they differed—either in the like or the dislike direction by 27% or more—became the 56 items in the female physicist occupational scale. Some things that female physicists like more than other women are obvious, such as science and math. Women physicists also like some things that might not be intuitively obvious, such as mechanical activities and statistics. They also distinctively dislike some things more than most other women, namely business, fashion, and extroverted activities.

Because of its method of construction, and also because of the type of the people in most occupations, an occupational scale has some distinctive features that differ from many psychological scales. First and foremost, it has heterogeneous content—a mix of relative likes and dislikes that distinguish people in that occupation. Second, as a corollary to the heterogeneous content, an occupational scale is difficult to interpret in content terms. It can be interpreted in an important actuarial sense by saying that a high score signifies the person shares the distinctive likes and dislikes of people in that occupation. That is a very important piece of information but not one that conceptually disentangles the types of likes and dislikes that lead to a high or low occupational score. This problem led to the development of homogeneous content scales.

### 2.2. Homogeneous Content Scales

The first popular interest inventory to use homogeneous content scales was the Kuder Preference Record. This pioneering work by Frederic Kuder dates back to the 1930s. Today, content scales are at least a part of nearly every major interest inventory. Some, such as Holland's Self-Directed Search, use exclusively content scales and do not use occupational scales at all. Content scales are designed to have directly interpretable content in a specific domain, such as mechanical interests. Content scales are made up of items that tend to correlate in a general group of people, and thus, by definition, they are homogeneous. Content scales are similar to the factors that might be derived from factoring a set of interest items, but most inventories do not use this approach to create factor dimensions. Instead, content scales are often constructed to fit a rational or conceptual map of important and relatively independent interest domains. Often, items are included in an inventory to tap a particular domain, and an inventory could be revised

by adding items likely to create a new content scale. Content scales are typically built by examining item intercorrelation matrices of the tentative items to refine the item set to enhance homogeneity or internal consistency reliability of the scale. Holland's RIASEC dimensions are the best example of broad content scales, and the majority of popular inventories have measures of these six dimensions. Many inventories also have basic interest scales, which are content scales with more specific focus than the Holland dimensions. Thus, an inventory might have an artistic Holland scale yet also include basic interest measures of its components of fine arts, music/dramatics, and writing.

### 3. MAJOR INTEREST INVENTORIES

Five of the most popular interest inventories are the Campbell Interest and Skill Survey (CISS), the Kuder Occupational Interest Survey (KOIS), the Self-Directed Search (SDS), the Strong Interest Inventory (SII), and the UNIACT Interest Inventory. The SDS and UNIACT inventories represent the most simplified approach to interest measurement because their cores are primarily the six Holland RIASEC dimensions. The most complex inventories, because of the number and kinds of scales, are the Strong and the Campbell inventories. The Kuder inventory is somewhat less complex and represents an intermediate level of complexity. The utility of these inventories has been extensively supported by research. Interest scales effectively differentiate and predict important career behaviors, such as career choice and persistence, college major, and educational aspirations.

#### 3.1. Campbell Interest and Skill Survey

Campbell's CISS is the newest of the major interest inventories. Earlier, Campbell was an important innovator with the Strong, so it is not surprising that the CISS is like the Strong in core ways: occupational scales, Holland-like scales, basic interest scales, and an organization by the Holland-like scales. It differs from Holland's scales in that it renames the RIASEC dimensions in more ordinary language and it adds an additional general content scale for adventuring. Even more distinctive in the CISS is the addition of self-ratings of skills to match the interest dimensions. For example, there is a scale for mathematics interests but also a parallel scale for mathematics skill. Such parallel

scales are plotted together so that discrepancies between interests and skills in the same domain are highlighted. Campbell's use of separate skill ratings has been emulated in other assessment inventories.

#### 3.2. Kuder Occupational Interest Survey

Kuder was one of the early pioneers of interest measurement. The KOIS is his approach to occupational interest scales, one that is quite different from Strong's. Kuder measured an individual's direct similarity to the average interests of people in an occupation, whereas Strong's occupational scales measure items that differentiate people in an occupation from those in a general group of other occupations.

#### 3.3. Self-Directed Search

The SDS follows the Holland mantra for simplicity and ease of use. Its items are self-scored to yield scores on the six RIASEC dimensions. These scores are then translated into a RIASEC code that indicated the three highest scores in the profile.

#### 3.4. Strong Interest Inventory

The SII, the oldest of today's inventories, has been revised nearly every decade to update the occupational scales that are the inventory's centerpiece. Since the beginning the SII has had a data-driven base, and it is widely viewed as the "most scientifically grounded" tool for interest assessment. The 1994 version of the SII contains 211 occupational scales constructed separately for women and men. Starting in the 1960s, the authors of the Strong added content scales—first the Basic Interest Scales, then the Holland General Occupational Theme Scales, and, recently, the Personal Style Scales. Beginning in 1996, the Strong has been available with the Skills Confidence Inventory, which measures confidence in each of the six RIASEC dimensions.

#### 3.5. UNIACT Interest Inventory

Unlike the other four inventories, the UNIACT is not commercially available to the public. It is part of the assessment tools available to U.S. high schools through the American College Testing Program. It uses six content scales to measure the six RIASEC dimensions

and then maps the scores on a world-of-work map, which is an extension of Holland's hexagon.

#### 4. QUALITY AND USE OF INTEREST MEASURES

Vocational interests show remarkably robust psychometric characteristics, presumably because the distinctive likes and dislikes of people are deep and often enduring. By young adulthood, most people are quite clear about their passions and aversions. Moreover, people in different educational and occupational settings are often starkly different in what they like and dislike. The following are examples from the Strong inventory norm group: 74% of 200 women airline attendants liked "looking at things in a clothing store" versus 33% of 200 women physicists. For these same women, 65.5% of the flight attendants said they disliked "calculus" as a school subject versus none of the physicists.

##### 4.1. Reliability

Content scales are expected to have high item intercorrelations; this internal consistency reliability is usually expressed with Cronbach's alpha. Because interests are such robust phenomena, it is rather easy to develop interest scales with high reliability. Thus, the six Holland RIASEC scales in the 1994 Strong all have alpha reliabilities of 0.90 or more. Occupational scales are expected to be heterogeneous, so a different approach to reliability is appropriate. The typical approach is to examine test-retest reliability of occupational scales. Interests are usually quite stable over long periods of time. For example, people who liked mathematics as adolescents are likely to have the same feelings 30 years later.

##### 4.2. Validity

There is a large literature on the validity of vocational interests for differentiating occupational and educational groups and for predicting career outcomes. Most of the references in the Further Reading section speak to such issues. Holland's observation that aspirants for occupations often resemble those who are in occupations leads to the probability that evidence for concurrent validity will map into evidence for predictive validity. This means that if we can identify the distinctive interests of adults currently in an occupation, such a cluster of interests will also longitudinally predict entry to such an

occupation, or a similar occupation, for young people, or adult career changers, developing a career trajectory. This does not negate the special value of predictive studies, but it does convey the considerable value of concurrent validity studies.

##### 4.3. Counseling Utility

Especially in the hands of a skilled career counselor, interest inventories can be useful in several ways. The best use will help clients better understand themselves and their vocational options and stimulate directions for action. Best practices in vocational assessment blend subjective information from counseling with objective information such as interest inventories to reveal the client's career story.

Broad scales, such as Holland's RIASEC scales, and specific scales, such as basic content scales, have different roles to play in counseling and in self-understanding. The broad scales are readily understood and readily organize broad career options. They are the best place to start the process of formulating a career story and charting directions for exploration. The specific scales then supplement this broad information when clients start refining their career story and their more precise paths for exploration. A broad enterprising scale points to a broad venue of influencing, yet more precise scales, such as merchandising and public speaking, speak distinctively to the options of retail sales versus politics. The specific scales are especially useful in identifying client strengths because there are a larger number of them and therefore they are more likely to show some high scores indicating areas of interest.

#### 5. TRENDS IN INTEREST MEASUREMENT

The level of research about interests has been intense and lively in the past decade. Much work has been done to examine whether theories and measures of interests can be generalized across diverse groups. There is also a ferment of activity about how domains of interests relate to other domains of individuality such as personality and self-efficacy.

Recent research and theorizing have examined how interests are related to other domains of individuality, leading to speculation that there is a unified set of underlying traits linking interests, personality, and self-efficacy. This makes for exciting prospects for

both theory and applied assessment. For example, work in social cognitive career theory posits a developmental relationship whereby emerging self-efficacy leads to interest development. This causal proposition is stimulating much longitudinal research. For example, Nauta *et al.* found that the causal relationship was reciprocal, with interests leading to self-efficacy and conversely. Many interest inventories have a parallel set of measures of self-efficacy and there is burgeoning interest in their joint counseling use and theoretical implications.

Although Holland's position is that interests are expressions of personality, other writers have questioned the degree of overlap between the two domains. This stimulated a number of studies in the 1980s and 1990s to determine how interests and personality correlate with each other. A useful summary of this literature is provided by Larson *et al.*'s meta-analysis of 24 samples completing an interest inventory measuring Holland's Big Six dimensions and a personality inventory measuring the Big Five personality dimensions. From their results, it is evident that certain personality and interest dimensions are clearly related. Correlated 0.30 or more are personality openness and artistic interests, extroversion and enterprising interests, and extroversion and social interests. Also related are openness and investigative interests, and personality agreeableness and social interests. Such results demonstrate that there are clear links between certain broad dimensions of interests and broad dimensions of personality. The personality-interest links are even stronger and more meaningful when more specific dimensions are examined.

## 6. CONCLUSIONS

For more than 80 years, vocational interests have been studied and assessed with considerable success. Vocational interests are robust phenomena. The interests of people and occupations differ greatly and in theoretically meaningful ways. Several widely used interest inventories include high-quality measures with utility in counseling for building the client's career narrative. Vocational interests are the cornerstone of effective career counseling that generates options for exploration. Increasingly, it is apparent that other domains of individuality, such as personality and self-efficacy, are vitally linked to interests. A unified and integrated approach to all these domains will lead to enlightened theory and career counseling.

## See Also the Following Articles

Career Counseling ■ Holland's Theory (Vocational Personality Types) ■ Indecision, Vocational ■ Vocational Assessment in Schools ■ Vocational Psychology, Overview

## Further Reading

- Betz, N. E., Borgen, F. H., & Harmon, L. W. (1996). *Manual for the Skills Confidence Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Borgen, F. H., & Lindley, L. D. (2003). Optimal functioning in interests, self-efficacy, and personality. In W. B. Walsh (Ed.), *Counseling psychology and optimal human functioning* (pp. 55–91). Hillsdale, NJ: Erlbaum.
- Campbell, D. (2002). The history and development of the Campbell Interest and Skill Survey. *Journal of Career Assessment, 10*, 150–168.
- Day, S. X., & Rounds, J. (1998). Universality of vocational interest structure among racial and ethnic minorities. *American Psychologist, 53*, 728–736.
- Diamond, E. E., & Zytowski, D. G. (2000). Vocational assessment. In C. E. Watkins, & V. L. Campbell (Eds.), *Testing and assessment in counseling practice* (2nd ed.). Mahwah, NJ: Erlbaum.
- Donnay, D. A. C., & Borgen, F. H. (1996). Validity, structure, and content of the 1994 Strong Interest Inventory. *Journal of Counseling Psychology, 43*, 275–291.
- Donnay, D. A. C., & Borgen, F. H. (1999). The incremental validity of vocational self-efficacy: An examination of interest, self-efficacy, and occupation. *Journal of Counseling Psychology, 46*, 432–447.
- Fouad, N. A., & Zao, K. E. (2000). Meeting the revolution: Future trends in vocational assessment. *Journal of Career Assessment, 8*, 403–409.
- Harmon, L. W., Hansen, J. C., Borgen, F. H., & Hammer, A. L. (1994). *Applications and technical guide for the Strong Interest Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Holland, J. L. (1997). *Making vocational choices* (3rd ed.). Odessa, FL: Psychological Assessment Resources.
- Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of Big Six interests and Big Five personality variables. *Journal of Vocational Behavior, 61*, 217–239.
- Nauta, M. M., Kahn, J. H., Angell, J. W., & Cantarelli, E. A. (2002). Identifying the antecedent in the relation between career interests and self-efficacy: Is it one, the other, or both? *Journal of Counseling Psychology, 49*, 290–301.
- Prince, J. P., & Heiser, L. (2000). *Essentials of career assessment*. New York: Wiley.
- Savickas, M. L. (1999). The psychology of interests. In M. Savickas, & A. Spokane (Eds.), *Vocational interests: Their meaning, measurement, and counseling use* (pp. 19–56). Palo Alto, CA: Davies-Black.

- Savickas, M. L. (2002). Career construction: A developmental theory of vocational behavior. In D. Brown (Ed.), *Career choice and development* (4th ed., pp. 148–205). San Francisco: Jossey-Bass.
- Savickas, M. L., & Spokane, A. R. (1999). *Vocational interests: Their meaning, measurement, and counseling use*. Palo Alto, CA: Davies-Black.
- Savickas, M. L., Taber, B. J., & Spokane, A. R. (2002). Monograph: Convergent and discriminant validity of five interest inventories. *Journal of Vocational Behavior*, *61*, 139–184.
- Walsh, W. B. (2002). Introduction: Exploring interests, self-efficacy, and goals. *Journal of Career Assessment*, *10*, 147–149.
- Walsh, W. B., & Betz, N. E. (2001). *Tests and assessment* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Watkins, C. E., & Campbell, V. L. (Eds.). (2000). *Testing and assessment in counseling practice* (2nd ed.). Mahwah, NJ: Erlbaum.



# Vocational Psychology, Overview

Mark L. Savickas

Northeastern Ohio Universities College of Medicine, Rootstown, Ohio, USA

1. Person-Environment Fit Model
  2. Vocational Personalities and Work Environments
  3. Career Model
  4. Career Development
- Further Reading

## GLOSSARY

**career adaptability** The readiness and resources for making career choices and negotiating career transitions.

**planfulness** An attitude that favors planning and preparing for anticipated future events.

**psychological type** A constellation of traits such as particular abilities and specific interests that characterize a group of individuals.

**psychosocial** A relationship or interaction between an individual and a social group.

**skilling** Engagement in training and education that prepares one to enter a particular occupation.

**trait** An underlying dimension that structures and explains recurring uniformity in clusters of behavior emitted by an individual.

When asked what characterizes mental health, Freud reputedly answered, “love and work.” Although most psychologists prefer to focus on relationships, work also provides a fruitful focus for the study of personality and human development. This assertion is borne out in vocational psychology, the discipline in

applied psychology that concentrates on how individuals experience their work, occupations, and careers. The basic unit of study in vocational psychology is “vocational behavior,” that is, the responses an individual makes in choosing and adapting to an occupation. In addition to studying vocational responses to occupational stimuli, vocational psychologists also study the development of vocational behavior over the lifespan, from the growth of occupational aspirations during childhood to the disengagement from the world of work during old age. The overview of vocational psychology presented herein explains the structure and function of the discipline, particularly emphasizing its concern with person-environment fit and career development and introducing its major constructs.

## 1. PERSON-ENVIRONMENT FIT MODEL

The study of vocational behavior emerged as a scientific discipline at the beginning of the 20th century as industrialized countries across the globe saw increasing numbers of workers leave their farms to seek employment in the large factories that dominated modern cities. These companies assumed a bureaucratic form and a hierarchical organization. This organizational structure of ranked positions changed workers from self-employed individuals who performed a craft for their community to employees who performed just

some of the specialized tasks required to bring a product to market. Employers began to create jobs composed of a few circumscribed tasks. These factory jobs required only a small set of skills and interests, so employers sought to fill these jobs with employees who clearly displayed the required characteristics.

Upon arrival in the metropolis, with their immediate family in tow, individuals from the countryside and from different countries faced the task of finding employment. Unaccustomed to life in the city, with its factories and specialized occupations, many of these newcomers needed help in relating their talents to available occupations. Providing this vocational guidance became the mission of philanthropic organizations that helped newcomers to the city choose a job and adapt to urban life. Guidance workers in institutions such as the YMCA and settlement houses used a three-step model to match people to occupations. First stated by Parsons (1909), the person-environment (P-E) model consists of helping individuals increase self-knowledge and gather occupational information, and then rationally match the two to select a fitting occupation. Advocates of the model for P-E fit assert, with ample evidence, that a good match of a worker's abilities and interests to a job's requirements and rewards leads to important outcomes. The organization gets a satisfactory worker, one who is productive and reliable, and the successful worker earns economic rewards and feels job satisfaction.

The primary outcomes of fitting workers to work, that is, occupational success and job satisfaction, may result in an important secondary outcome—career stability. After leaving school, individuals usually progress through a sequence of trial positions in a process of zeroing in on a more-or-less permanent job. This sequence of trial jobs can entail horizontal movement to lateral positions in the hierarchy or vertical mobility up the organizational ladder. Ideally, each new job should improve correspondence between the individual's resources and the occupation's requirements, leading to more effective functioning. Once a good match has been made, individuals usually remain in that congruent position, enjoying a relatively stable work life. The length of this period of stable employment is the major variable in comprehending the wide range of work histories that individuals report. At one end of the continuum, some individuals find a fitting position, remain in it for 30 years, and then retire from it. At the other extreme, some individuals never stabilize for more than a year or two, having as many as 25 jobs in their career. A more typical work history finds workers remaining in a position for some significant

period, say 5 to 7 years, before moving to and re-stabilizing in another position, repeating this process four to eight times during their work lives. Various career patterns are thus characterized by the sequence and duration of work positions occupied by an individual. Using variations in tenure, researchers have identified a set of common career patterns that accommodate the work histories of most individuals. The career patterns experienced by large groups of workers have been examined to identify how occupations relate to each other and to describe the vocational personality types that populate particular occupations. These observations became systematized as researchers studied individual differences in vocational personalities and similarities in work environments.

## 2. VOCATIONAL PERSONALITIES AND WORK ENVIRONMENTS

Development of the psychotechnology for implementing the person-environment paradigm was spurred by World War I, as the methods devised for matching soldiers to military positions were adapted to civilian occupations. Two lines of research dominated that era, and remain dominant today. The first line concentrates on success and the second on satisfaction. Efforts to predict success led both to the development of tests that measure individual differences in abilities, aptitudes, and skills and to the assessment of the ability patterns that characterize workers in various occupations. Efforts to predict satisfaction led both to the development of inventories that measure individual differences in vocational interests and work values and to the assessment of the interests reported by satisfied workers employed in various occupations. Vocational guidance counselors match individuals to fitting occupations by comparing the abilities and interests that characterize different occupations to the profiles individuals produce on ability tests and interest inventories. Of course, personnel officers use this same approach in employee selection. So, the P-E theory and its associated techniques can be used to benefit the employee, the employer, and the community.

While industrial-organizational (I/O) and vocational psychologists both appreciate the roles that abilities and interests play in shaping careers, they differ in the relative emphasis they place on these critical variables. Industrial psychology, with its focus on an organization's productivity and employee success,

emphasizes ability tests and occupational ability patterns. In contrast, because it concentrates on the perspective of the individual worker and job satisfaction, vocational psychology emphasizes research on work motivation. Although some attention has been, and continues to be, paid to motivational variables such as personal needs and work values, the majority of vocational psychology's research on motivation has concentrated on vocational interests.

### 2.1. Vocational Interests

Interest denotes a state of consciousness characterized by a readiness to respond to particular environmental stimuli, such as objects, activities, and people. When activated, interests prompt selective attention that narrows the perceptual field to concentrate on the stimulus. This attention is accompanied by a pleasant feeling and an evaluation of liking that steers goal-directed behavior and maintains persistent effort to fulfill some personal need or value. As a personality trait, vocational interests denote a homogeneous group of specific interests that form a consistent, persistent, and stable dispositional response tendency which increases one's readiness to attend to and act upon a particular group of environmental stimuli. This orientation shows habit (or absolute) strength in how much stimulation is required to activate it. Habit strength can be assessed by behavioral analyses and interest autobiographies that reveal the ease and frequency with which an interest is initiated, as well as the duration for which it extends and the length of time for which it persists. A disposition shows relative strength in activity preferences, that is, competition with other interests for behavioral expression. Relative strength of interests can be measured with interest inventories.

In addition to being viewed as a personality trait, vocational interests may also be conceptualized as a psychosocial construct, psychosocial meaning the interrelation of individual and society in human development, especially connoting the influence of social factors on an individual's thoughts and actions. From this perspective, vocational interests denote a complex, adaptive effort to use one's environment to meet needs for survival and self-realization. Vocational interests expedite P-E interactions by uniting subject, object, and behavior into a vital relationship, a relation manifested in actions that satisfy needs, fulfill values, foster self-development, enhance adaptation, and substantiate identity.

The dominant figure in the study of vocational interests during the 20th century was E. K. Strong, Jr.,

who in the 1920s devised a method for measuring an individual's vocational interests and a scoring system for determining empirically how well those interests resemble the interests of workers employed in a number of different occupations. Strong's inventory, with continuing updates, remains one of the more popular tools used in vocational psychology research and in career counseling practice. The current version of Strong's inventory incorporates two additional types of interest scales that augment his empirically keyed occupational scales (with heterogeneous items) for measuring degree of resemblance to different occupational groups. The first set of scales was added to Strong's inventory in the 1960s; they were basic interest scales composed of homogeneous items that measure expressed vocational interests. Scores on these rationally constructed scales are interpreted as a direct indication of an individual's vocational interests, in contrast to Strong's empirically constructed scales of heterogeneous items which indicate occupational interests by indexing resemblance to workers in different occupations. In the 1970s, a second type of scale was added to Strong's inventory, one that measured resemblance to the six prototypes in Holland's theory of vocational personality types.

### 2.2. Vocational Personality Types

Based on factor-analytic studies of interests and research with Strong's inventory, Holland logically deduced six types of vocational personalities and devised a heuristic theory to describe vocational behavior and organize work histories. The six prototypes in Holland's theory are each characterized by a distinct syndrome of interests, competencies, and activities.

1. *Realistic* (R) types report outdoor and mechanical interests, prefer to work with animals and machines, enjoy the role of doer, display physical competencies, like leisure pursuits involving physical skill and challenges, and admire role models such as athletes and adventurers. They might be heard often saying, "just do it."

2. *Investigative* (I) types report scientific interests, prefer to work with ideas, enjoy the role of thinker, display intellectual competencies, like leisure pursuits involving reading and researching, and admire role models such as scientists, inventors, and detectives. They can often be heard saying, "let's explore it."

3. *Artistic* (A) types report artistic, literary, and musical interests, prefer to work with feelings, enjoy the role of creator, display aesthetic competencies, like leisure pursuits involving self-expression and appreciation of

concerts, theaters, and museums, and admire role models such as artists, composers, writers, and performers. Often, they can be heard saying, “let’s create it.”

4. *Social (S)* types report social interests, prefer to work with people, enjoy the role of helper, display communication competencies, like leisure pursuits involving conversation and social gatherings, and admire role models such as teachers and social workers. They can often be heard saying, “let’s talk about it.”

5. *Enterprising (E)* types report sales and managerial interests, prefer to work with opinions, enjoy the role of leader, display persuasive competencies, like leisure pursuits involving travel and politics, and admire role models such as public officials, military officers, and corporation presidents. Often they could be heard saying, “make it happen” or “make it so.”

6. *Conventional (C)* types report clerical and business interests, prefer to work with data and records, enjoy the role of member, display organizing competencies, like leisure pursuits involving collecting and organizing, and admire role models such as teams, organizers, and historians. They can often be heard saying, “keep it going.”

Holland subsequently elaborated six types of environments, or ecological niches, that correspond to each of the six vocational personality types. Along with his colleagues, Holland then used his typology of environments to construct an occupational classification system as well as to classify other behavioral settings such as college majors, universities, leisure activities, and hobbies. Researchers type a behavioral setting by identifying the modal vocational personality types that populate it. Alternatively, one can type an environment by observing its contents and activities. For example, an automobile parts store with mechanical parts lining its neat shelves is a realistic and conventional environment, whereas an elementary school with children discussing their feelings is a social and artistic environment. Holland’s hexagonal scheme for portraying the world of work has been compelling in vocational research and pragmatic in career counseling practice.

### 2.3. Consistency and Congruence

The six types of vocational personalities and work environments constitute the primary elements in Holland’s theory. In due course, he elaborated four secondary constructs, starting with consistency. Holland determined the relative frequency of type patterns and used this information to order the six types around a hexagon, starting with R in at the upper left

node, and followed clockwise by I, A, S, E, and C. Using the RIASEC ordering, researchers and practitioners can determine the degree of internal consistency for a vocational personality or work environment. Consistent types appear next to each other. For example, the most frequent secondary type for an individual with a primary type of artistic is investigative, which precedes A in the RIASEC ordering, or social, which follows I in the RIASEC ordering. Inconsistent types such as artistic and conventional are opposite each other; and, intermediate types such as artistic and enterprising are separated by one other type. The theoretical elegance and practical utility of Holland’s research on the structure of interests has been augmented by other researchers who have elaborated the hexagon first by embedding it into an interest circle and then converting the circle into an interest globe.

In addition to the construct of consistency, which deals with the intrapersonal organization of interests, Holland also elaborated his theory by conceptualizing how people behave in environments. In using the same construct language to type both people and environments, Holland made it clear to even casual observers that adaptive fitness is optimal when people work in congruent environments. Simply stated, the construct of congruence proposes the idea that corresponding matches of person to environment produce success, satisfaction, stability, and mental health. This idea enabled Holland to show, for example, that an individual who is ISA, say a psychologist, would be congruent in an ISA work environment, say working as a professor, but incongruent in a REC environment, say working as a roofer.

To further characterize each type of person and occupation, and more importantly the match between them, Holland assesses degree of differentiation or intensity of interests. Marked differentiation indicates a specialist with narrow interests, whereas limited differentiation indicates a generalist with broad interests. A final construct, vocational identity, refers to how clear a picture an individual has of his or her own type and how many distinct behavior situations occur in a work environment. Other things being equal, Holland proposes that it is easier to predict work adjustment for an individual who works in an organization with few distinct behavioral situations and who displays a clear identity with markedly differentiated interests. While still the dominant paradigm in vocational psychology, the individual differences view of occupations was augmented in the 1950s by a second grand model, namely, the individual development view of careers.

### 3. CAREER MODEL

Career is the development of vocational behavior over time or, according to the *Oxford English Dictionary*, “a person’s course or progress through life.” The career model, which concentrates on the development of vocational behavior, shares an interest with the P-E model in being concerned with how individuals fit into occupations at a particular point in their lives, yet prefers to emphasize how individuals fit work into their lives from childhood to retirement. Instead of the occupational stimulus-vocational response (S-R) paradigm for studying vocational behavior, career studies use a response-response (R-R) paradigm. Vocational behavior, or response, remains the basic unit of study, but instead of studying differences in vocational behavior between individuals, the career perspective concentrates on changes in vocational behavior by the same individual across time, often investigating the antecedents of career patterns. Thus, the career perspective takes a longitudinal view, in contrast to the cross-sectional view taken by the occupational perspective. Vocational development is inferred and career is denoted by the systematic changes that can be observed in vocational behavior across two or more points in time. Because P-E fit can be studied at a single point in time, vocational behavior and its outcomes may be directly observable. In contrast, careers synthesize past memories, present reflections, and future anticipations into a narrative with a unified plot. As one’s own story, a career is not directly observable. Accordingly, career denotes contemplation about the course of one’s vocational behavior, not vocational behavior itself. This deliberation can focus on actual events such as one’s occupations (objective occupational career) or on biographical themes (subjective vocational career) that give meaning to one’s occupations.

The objective view of careers, whether by external observers or the person him- or herself, looks at the evolution of P-E fit across the life course. It is a précis of the actual events in one’s work history, including all of the preparatory, participatory, and retirement positions an individual has occupied. This public definition of career can be objectively documented as facts and analyzed for patterns defined by the sequence and duration of positions occupied by an individual. The subjective view of careers as private meaning stands in contrast to the public pattern of occupations in a work history. Subjective career, while also a reflexive project, refers to how individuals understand their occupational roles and interpret their work experiences. A

subjective career concentrates on the expression of a self-concept in work roles; it tells one’s “own story,” usually by focusing on a sense of purpose that coherently explains the continuity and change in one’s self across time. It is this personal meaning that individuals use to orient themselves to their society’s occupational structure, impose meaning on their vocational behavior, and plan their future work life. Whether one takes the public or the private view of career, most researchers who use the term career study the development of vocational behavior, or career development.

### 4. CAREER DEVELOPMENT

Careers involve the psychosocial integration of self-concepts and social roles. They are constructed by recursive transactions between individuals and their environments. Individuals, as self-organizing systems with their own goals, must find ways to fulfill their needs through playing social roles predetermined by the community in which they live. Communities usually script these social roles according to gender, race, ethnicity, and social class. Members of the community then socialize children by inducting them into these roles and communicating expectations for role performance. These social expectations are called developmental tasks, and failure to master the tasks can result in delayed or deviant patterns of adaptation. When the social expectations pertain to the work role, they are called vocational development tasks. Success in adapting to each task of vocational development results in more effective functioning as a student, worker, or retiree and lays the groundwork for mastering the next task along the developmental continuum. Skipping a task in the normal sequence usually results in difficulties in coping with later tasks. Super charted the linear progression of vocational development tasks and demarcated career stages that correspond to life stages of childhood, adolescence, early adulthood, adulthood, and late adulthood. The name of each career stage communicates its central activity: growth, exploration, establishment, maintenance, and disengagement.

#### 4.1. Growth

The stage of career growth, usually ages 4 to 13, involves developing a vocational self-concept as well as the coping readiness and resources that will be needed to translate that self-concept into an occupational role. The vocational self-concept and

the translation attitudes and abilities constitute two interrelated yet distinct aspects of vocational development—its content and process. Career choice content centers on the vocational self-concept, that is, the totality of attributes that compose an individual's picture of self in the work role. Because vocational self-concepts develop primarily in social situations, an important determinant of their content is the social world that children inhabit. The social order, with its unequal arrangement of gender, race, and class, shapes children's occupational aspirations. Vocational self-concepts, and the careers they engender, are deeply grounded in an individual's internal representation of his or her location among unequal social positions. Using their "status identity," children circumscribe the range of occupational alternatives that they consider and later, when making choices, compromise their own interests to accommodate their perceptions of an occupation's status and sex type.

The process of circumscription and compromise in perceiving roles and aspiring to goals begins as children look to their parents as guides and identify with their preferences. This process of self-formation continues as children use imaginative play and school experiences to construct a bridge to take them from the family to the community. In what may be considered a very important career choice, children select role models who portray solutions to problems in growing up that they themselves face. As children imitate the desirable qualities of their role models, they rehearse relevant coping attitudes and actions, form values and interests in certain activities, and develop abilities and skills as they engage in these activities. Playing selective roles, with increasing attention to the results, enables a reality testing that strengthens or modifies vocational self-concepts.

While career choice content involves the subject matter of vocational self-concepts, career choice process deals with the attitudes and actions that construct a career. The construction process is initiated by the tasks of vocational development that individuals encounter during the growth stage of a career. Family, friends, and teachers expect students in middle school and junior high to develop a sense of concern about their vocational future, learn that they can control this future through self-determination and negotiation, become curious enough to explore possible selves and future scenarios, and develop confidence that they can implement plans and deal with problems that occur in actualizing choices. Attitudes of concern, control, curiosity, and confidence chart four developmental lines

that, during adolescence, will coalesce to form a teenager's career choice readiness and resources.

## 4.2. Exploration

The career stage of exploration, usually ages 14 to 24, involves seeking information about self and about the work world in preparation for making choices that match the two. Three vocational development tasks characterize the years of career exploration: crystallizing vocational preferences, specifying an occupational choice, and actualizing that choice by obtaining a job in the specified occupation.

### 4.2.1. Crystallization

The first vocational development task of the exploration stage is to crystallize a vocational self-concept and translate that picture into preferences for a group of occupations, usually in similar occupational fields and at the same ability level. Crystallizing vocational preferences requires in-depth exploration to learn information about the self, occupations, and the fit between the two. As information and self-knowledge accumulate, individuals conceptualize new distinctions about their work-role self-attributes. Vocational exploration is fostered by educational experiences and can be accelerated by considering the results of vocational testing that draws an objective portrait of one's vocational interests, occupational abilities, and work values. A schema for understanding the world of work, such as the RIASEC hexagon, can be used to organize an individual's occupational information into a coherent framework that makes the information meaningful and relates it to vocational self-concepts.

### 4.2.2. Specification

Specifying an occupational choice from among crystallized vocational preferences is the second task of career exploration stage. Specification emerges from exploration in depth that sifts through attractive preferences and possible selves. This search involves more than matching vocational self-concepts to occupational outlets; it involves the psychosocial process of identity construction, because in declaring an occupational choice, individuals identify themselves before some audience. Stating an occupational choice constitutes a very public presentation of the self; it displays who we are and announces what we want to become. Public declaration of an occupational choice specifies

meanings that both an actor and observers attribute to the self as a worker. It involves telling a story about the self, one with a plot that makes a life whole by unifying and assigning meaning to various elements of the self.

A key step in crystallizing preferences and specifying choices involves developing the attitudes, beliefs, and competencies (the ABCs of career construction) that bear upon career decision making. The adaptive response dispositions and cognitive competencies have been identified and organized into a framework for assessing career adaptability. The dispositions are charted by the four developmental lines that originated in the growth stage and continue to develop during the exploration stage. Ideally, adolescents show career concern by orienting themselves to vocational development tasks and actively participating in coping with current tasks and preparing for anticipated tasks. This forward-looking attitude fosters the development of planning competency. Concern prompts the issue of control, as teenagers answer the question "Who owns my future?" Hopefully their answers reflect beliefs in self-determination and feelings of agency. These beliefs foster the development of competency in making matching decisions. Individuals who feel some sense of control over their future become curious about the educational and vocational choices they might make. There are many styles and strategies for career decision making, but the most adaptive ones seek information about possible selves and alternative choices. Curiosity about how to match oneself to occupational positions fosters increasing competence in self-knowledge and occupational information. The final disposition involves confidence that one can perform the behaviors needed to make career choices and to solve the problems that arise in implementing them. These feelings of career self-efficacy foster development of competence in problem solving.

From this view, the career choice process evolves over time and is facilitated by strategic and decisive attitudes toward making choices, focused by accurate conceptions of the decision-making process, moved by confidence about executing choice behaviors, and supported by competencies that inform choice behaviors. Of course, the development of career choice attitudes and aptitudes shows individual differences. For each individual, career choice readiness and resources develop at different rates, with possible fixations and regressions. Delays within or disequilibria between the four developmental lines produce problems in coping with the tasks of crystallization and specification. Moderate disharmony among the developmental lines

produces variant patterns of indecision and unrealism, whereas strong disharmony produces deviant patterns of indecisiveness and indifference. Accordingly, the four developmental lines sketch a framework of career adaptability that can be used to assess career choice readiness and diagnose vocational decision-making difficulties, beginning in adolescence and continuing throughout adulthood.

Lack of readiness and resources for educational and vocational decision making lead to predictable career choice problems. These decisional difficulties have been studied since the 1930s. Originally, researchers investigated the construct of vocational indecision by assessing differences in personality between decided and undecided individuals. At first, it appeared that indecision was related to personality maladjustment; however, over time it became clear that delays in deciding might be developmentally appropriate and even adaptive. During the 1970s, researchers began to view "undecided versus decided" as a continuum rather than as two distinct statuses. Attention then turned to problems that delay or distort progressive movement along the continuum from undecided to decided.

Research has identified dozens of career choice problems and, through reflection and research, grouped these problems into homogeneous categories. Several alternative categorizations are available, yet each one depicts some combination of five types of career choice problems: informational, decisional, relational, motivational, and environmental. The most frequently experienced problems in making a career choice involve a lack of information about work-related self-attributes; about occupational requirements, routines and rewards; or about how to engage in the career choice process. The second most common group of career choice problems involves deficits in decision-making skill, including knowing and practicing the principles of decision making. Even when sufficient information is present, deficits in decisional competence may limit the ability to process information and make the decisions that produce a suitable occupational choice. The third group of career choice problems involves deficits in relational support, meaning that significant others such as parents or partners disapprove of a choice that attracts the individual or insist on a choice that repels the individual. The fourth category of career choice problems involves motivational problems that delay engagement in the choice process. These difficulties include dysfunctional career beliefs, anxiety about making choices, and adjustment

disorders. Whereas the fourth group of problems generally occurs before engaging in the choice process, the fifth group occurs after a choice has been made. Some individuals make choices that are difficult to convert into actuality because of internal or environmental barriers. The external barriers may involve inability to gain entry to or pay for training or apprenticeship programs as well as ageism, racism, and sexism. Internal barriers include lack of confidence and conflicting commitments.

It should be obvious that these career choice problems do not usually appear alone; typically, they occur in combinations and often form patterns. Therefore, researchers operationally define degree of indecision as the number of difficulties encountered, often distinguishing between undecided individuals who experience a few problems and indecisive individuals who experience many problems. Undecided individuals are generally considered to be experiencing a variant pattern of career development, with their inability to choose being viewed as developmentally appropriate and treated by information, decisional training, and social support. Indecisive individuals are generally considered to show a deviant pattern of vocational development in which chronic anxiety and lack of problem-solving skills characterize all their social interactions, not just those dealing with the work role. Indecisive individuals generally require counseling that concentrates on issues that cause their anxiety. Whether an individual finally chooses an occupation or lets time and circumstances dictate that choice, there comes a day when the individual must apply for a job.

#### **4.2.3. Actualization**

The third and final task of the exploration stage, actualizing an occupational choice, requires that individuals turn their verbal choice of an occupation into an actual fact. Actualizing a choice usually involves trying jobs in that occupation. Trial connotes employment in jobs that could become permanent, that is, adult employment as opposed to youth employment. The initial job allows individuals to try on the occupation for fit and then, through a process of elimination, subsequent trial jobs help the individual to zero in on a suitable job. The period of actualizing choices, which typically lasts about 4 to 7 years, is referred to as the school-to-work transition. During the years in which individuals seek to stabilize in a more-or-less permanent position, the critical vocational coping

behaviors consist of actions that move them to increasingly more congruent jobs. Any action aimed at vocational movement counts, including job search activities such as reading job advertisements and applying for positions.

The six main types of vocational movement vary in quality, with three being adaptive (skilling, experimenting, and stabilizing) and three being maladaptive (drifting, floundering, and stagnating). Drifting involves random movement to a position for which the individual is no better suited than the position being left or for which the individual lacks aptitude, interest, or preparation. Drifting implies being carried along by others or wandering aimlessly from position to position. Floundering means stumbling performance in a position or plunging ahead from one unsuitable position to another. Stagnating involves staying longer than appropriate or remaining in a blind alley job; it involves deterioration because it hurts future occupational possibilities.

The more adaptive coping behaviors start with skilling, that is, preparing to enter a regular adult occupation through further schooling, training, and apprenticeship. Experimenting involves moving from one related job to another in a process that eventually zeros in on a suitable and viable position. Stabilizing results in eventually settling into and making a place for oneself in a more-or-less permanent position. Compared to experimenting, stabilizing shows a lack of doubt. By age 25, about 80% of people have stabilized in a position. Over the long-term career, the quality of vocational coping behavior during the school-to-work transition is more important than sticking with the first job. Adaptive individuals may change jobs many times, moving through progressively more congruent positions, until they stabilize in a suitable position. In contrast, less mature individuals may flounder, drift, and stagnate rather than take the risks involved in advanced trial and in-depth exploration.

#### **4.3. Establishment**

The years of career establishment, usually ages 25 to 44, involve making one's job secure. Establishment involves three developmental tasks that coincide with the early, middle, and late periods of young adulthood: stabilizing, consolidating, and advancing. These vocational development tasks are discussed first in terms of their direct outcomes, then are reviewed in terms of their indirect outcomes, namely, job satisfaction and organizational commitment.

### 4.3.1. Stabilizing

The major developmental task of the early period in the establishment stage requires that the individual stabilize in an occupational position, primarily by doing two things. First, the individual must adapt to the organizational culture that surrounds the job. Adaptation to the organization deals with participating in the work environment, not with performing job tasks. Organizational adaptation involves transactions between the worker and the job environment, with the person engaged in efforts to learn about the company of workers and the company engaged in efforts to socialize the newcomer. There can be some reciprocity in the newcomer changing the company, but this is infrequent, and minimal when it does it occur. Adaptive efforts can begin before the employee is hired. Organizational adaptation is smoother when one's expectations are met, so providing a job candidate with a "realistic job preview" is a useful practice that reduces surprises and confusion in newcomers.

As they begin their employment, greenhorns must learn how things are done in their new company. This includes learning about the people, politics, values, language, and history of the organization in general and of their work group in particular. It is important that novices learn how their job fits into the big picture, that is, why their jobs are important and what they contribute to the organization's goals. Organizational adaptation is furthered by learning about and complying with the company's formal policies and procedures as well as its informal practices and unofficial rules. Observing others do the job is important in "learning the ropes" and figuring out unstated rules. In addition to observational learning, prior work experience is helpful, as is taking advantage of orientation programs, buddy systems, and mentoring.

Position performance is the second task of the early period in establishment. In addition to adapting to the organizational culture, individuals must demonstrate competence at work. They must clearly understand their job tasks, take these responsibilities seriously, and exhibit efficient and effective work performance. This is accomplished by concentrating on the task at hand and doing it right the first time. One must also be flexible in learning new ways of doing the tasks. In addition to adhering to production standards and quotas, new employees must follow safety procedures that reduce accidents in different occupations. During idle time, workers in the early years of establishment must learn how to keep busy. If workers cannot perform the

job tasks adequately, or cannot adapt to the company's culture, then instead of stabilizing they may leave voluntarily or be terminated. If they are comfortable doing the job, and doing it in that organization, then they can turn their attention to increasing their job security.

### 4.3.2. Consolidating

The tasks of the middle period of the establishment stage involve consolidating one's position for the long run. In contrast to position performance, which involves what a person does on the job, consolidation involves how one goes about doing that job. How one works has two components, one interpersonal and the other personal. The interpersonal component involves the vocational development task of forming congenial relationships with coworkers and supervisors. Coping behaviors that facilitate cooperation at work include communicating appropriately, supporting colleagues, encouraging coworkers, demonstrating goodwill to supervisors, tolerating individual differences, and dealing with diversity. Skill at organizational politics helps with this, as does finding safe ways to blow off steam. Of course, personality plays a significant role in coworker relationships, and lately attention has focused on how agreeableness and extroversion, two factors in the five-factor personality model (FFPM), may predict the enthusiasm and friendliness with which one engages coworkers. Inability to get along with coworkers is the most frequent reason for leaving or losing one's job during the middle years of establishment.

While the interpersonal component of how one works involves relationships on the job, the personal component of how one works involves work habits and attitudes. It is relatively easy to be industrious and dependable during a probationary period, or even during the first few years in a new position. A task of the middle years of establishment is to maintain positive attitudes and disciplined habits. Over time, some workers become complacent or lackadaisical in approaching their job tasks; they may even experience boredom. Of course, this posture diminishes their productivity and their value to the organization. Conversely, workers who meet the task usually increase their productivity as they expand their job knowledge and master the "tricks of the trade." Conscientiousness, as its very name indicates, is the personality factor in the FFPM that predicts success in maintaining adaptive work habits and realistic attitudes.

### 4.3.3. *Advancing*

After consolidating their positions, some workers, but not all, think about advancing to another position in their organization. To earn promotions, they must envision the career paths mapped by their organization and perform the behaviors needed to get ahead, which include showing initiative, taking on more responsibility, engaging in on-the-job training, and enrolling in continuing education. Workers who wish to advance, rather than maintain what they have established, yet see no way forward in their current organization may consider advancing their career by moving to a new organization that offers better opportunities. An even greater transition, or turn in the career path, involves advancement by preparing for and entering a new occupation. Changing occupations is best done by recycling through the tasks of exploration and establishment. The direct outcomes of mastering the vocational development tasks of stabilization, consolidation, and advancement are organizational adaptation, position performance, coworker relationships, work habits and attitudes, and promotions. The indirect outcomes are job satisfaction and organization commitment.

### 4.3.4. *Job Satisfaction*

Job satisfaction is the most frequently studied variable in organizational research because employers want to know how to avoid labor turnover and reward their best workers. In contrast to the organization's perspective on the relation of job satisfaction to productivity, vocational psychology views job satisfaction from the perspective of the worker, defining job satisfaction as the attitude or the general feeling that a worker has toward her or his job. The cognitive component of this attitude is an evaluative belief about how much one likes or dislikes a job. General job satisfaction can be explained by satisfaction with parts of a job that the worker likes or dislikes. When general satisfaction is high, little attention is paid to satisfaction with job facets. However, when job satisfaction in general is lower, examining the facets may explain the low morale by identifying the disliked job components. There are many facets, and many ways of conceptualizing and measuring them.

Vocational psychologists conceptualize job satisfaction as an indirect outcome of coping with the vocational development tasks of the establishment stage, so they link the facets of job satisfaction to the direct outcomes of coping with those tasks. When examining

job dissatisfaction, they look first at the facet of P-E fit as played out in organizational adaptation and position performance. This includes variables such as satisfaction with the organization, including its policies, compensation and other benefits, as well as satisfaction with job duties, including the job's routines and requirements, work load, and empowerment. Next, they consider satisfaction with coworkers, communication patterns, supervisors, and then the employee's conscientiousness and work ethic. Finally, they consider satisfaction with opportunities for advancement. Satisfaction with advancement opportunities can produce career satisfaction, even when job satisfaction is low, if the worker views that job as a temporary rung on a meaningful career ladder. For example, a university instructor may be quite dissatisfied with her position and pay, yet be satisfied with her career progress, knowing that with her next promotion she will be an assistant professor, a goal she has held for 10 years.

In addition to person-position fit and adaptation to the developmental tasks of establishment, job satisfaction relates to two more variables. The first variable is psychological, meaning that it resides inside the person rather than in the person-position interaction. This psychological variable is negative affect. People who generally experience negative affect score lower on job satisfaction, as they do on measures of satisfaction with other life roles. The second variable, one from sociology, is life structure or the arrangement of an individual's social roles. For example, part-time work interacts differently than full-time employment with an individual's other social roles. Job dissatisfaction can be caused by, or be coincidental to, role conflicts and role overload.

Relationships between family and work have been frequently studied, with a particular concentration on the interaction between the work role and the family roles of spouse, parents, and child. Some workers can be dissatisfied with their work because the job does not permit flexible scheduling or the company is not family friendly. And, of course, there is spillover between domains, with a divorce at home deteriorating work performance or a demotion at work causing arguments at home. Relationships between leisure and work have also drawn significant study. For example, type A personalities may experience high job satisfaction yet low life satisfaction if time and behavior at work reduce their engagement in other core roles such as recreation and friendship. Conversely, job dissatisfaction can be ameliorated by leisure, say by enjoying on a fishing line that which cannot be enjoyed on an assembly line.

The effects of job satisfaction are addressed by two relatively distinct research literatures. One deals with leaving dissatisfying jobs, while the other deals with being forced to stay in unsatisfying positions. Workers who must remain in unsatisfying positions may experience job stress that reduces their general emotional well-being. When people are chronically dissatisfied with a job they must perform 40 hours a week, they may eventually suffer from job burnout, a condition associated with a wide range of physical and psychological symptoms—headache, upset stomach, anxiety, and depression. Job dissatisfaction has even been linked to life expectancy, with those who hate their work predicted to have a shorter life expectancy.

Of course, an alternative to suffering the emotional and physical consequences of chronic job dissatisfaction is job withdrawal, including absenteeism and quitting. Although job dissatisfaction is widely believed to relate to absence, the empirical support is weak because there are many other reasons for being absent, including illness, personal business, and family responsibilities. In fact, having primary responsibility for the care of a child predicts absence better than does job dissatisfaction. Compared to absenteeism, there is stronger evidence of a causal link between job dissatisfaction and turnover. Dissatisfaction may lead to an intention to leave, followed by engagement in job search activities that include writing a resume, enrolling with job-finding agencies, and interviewing for positions. If another job is found, then the dissatisfied employee can leave. However, without a job offer in hand, employees are unlikely to quit. Also, it must be noted that in deciding whether to resign a position, individuals think about more than just how they feel about their jobs. They also think about the interpersonal relationships that they may have established at work.

#### **4.3.5. Organizational Commitment**

When work relationships are characterized by social support and fair treatment, individuals may feel an attachment to the organization and prefer not to leave their colleagues. This attachment between a worker and others in the work environment is called commitment. Typically researchers concentrate on organizational commitment in general, yet commitment can be connected to other foci such as coworkers, supervisors, the job, and the occupation. In addition to multiple foci, commitment itself is multidimensional, typically characterized by three dimensions. The affective dimension of commitment denotes emotional

attachment and identification with the organization, job, or coworkers. These bonds make one want to remain in the organization. The continuance dimension of commitment denotes the need to stay because no other job is available or, more often, because quitting would cost too much or require too many sacrifices. The normative dimension of commitment denotes a perceived obligation to stay because leaving would be wrong or let down too many people. Affective and normative commitment correlate positively to attendance, productivity, organizational citizenship behaviors, and pleasant affective experiences at work. Continuance commitment is typically unrelated, or even negatively correlated, to positive work behaviors. Of course, the affective, continuance, and normative dimensions of commitment can interact. For example, low affective commitment with high continuance commitment can cause problems such as making an individual feel locked into a situation with which they do not identify.

The normative dimension of commitment brings to mind the “psychological contract” between the employee and the organization. This construct refers to beliefs about reciprocal obligations held by people in an exchange relationship, or simply stated, issues of loyalty and payback. A psychological contract involves an employee’s perception of explicit and implicit promises regarding the exchange of the worker’s effort, abilities, and loyalty for the organization’s pay, promotions, and security. Individual workers prefer clear, unambiguous promises because clarity makes it easier to evaluate promises made and promises kept. Fulfilled contracts increase job satisfaction and commitment, whereas breached contracts lead to dissatisfaction and withdrawal behaviors. Breaches in the psychological contract require justification; these appraisals of workplace justice concentrate on fairness and inclusiveness in the process and outcome of reward allocation.

#### **4.4. Maintenance**

At some point in a career, whether it be successful or not, the worker seriously evaluates issues of satisfaction and commitment. During this period of renewal, individuals engage in self-questioning. They ask themselves, sometimes prompted by a partner, the midlife question: “Do I want to do this for the next 25 years?” Essentially, they engage in a process of re-evaluating their identity as they ask themselves, and sometimes their family and friends, if they should hold on or let go. If they decide to change organizations or occupations, then they recycle through

the tasks of exploration and establishment by crystallizing and specifying a different choice and then actualizing and stabilizing a new position. What we have been considering as a maxicycle of career stages can, in terms of a position change, be considered a minicycle of job transition. Having stabilized and maybe even advanced in a position, an individual may experience new growth that raises identity questions and leads to advanced exploration and then establishment in new position. Career patterns are defined by the repetition, frequency, and duration of these minicycles.

In the large, bureaucratic organizations of the 20th century, individuals might go through the maxicycle and maintain themselves in the stabilized position for 30 years before retiring. At the other extreme are individuals who frequently traverse the transition minicycle as they repeatedly grow, explore, stabilize, and then move to positions of new growth. Between these two extremes of a job for life and a job for now lie most of today's workers, who every 5 or 10 years find their career destabilized by illness and injury, layoffs and downsizing, or social change and personal upheaval. Developing a career today can be more about management of one's skills and opportunities than maintaining oneself in an established position.

When people do maintain themselves in a job, they enter the maintenance stage of a career, generally defined as ages 45 to 64, or from midlife to retirement. Maintenance addresses the requirement to remain in place, take care of what one has produced, and continue to produce. Society expects mature adults to hold steady in the positions they occupy, actively strive to preserve their status, remain interested in their work, and stay committed to their organizations. To remain in place, one must cope with novel, non-maturational problems that arise from environmental change or inner experience and preserve one's self-concept through modifying and adjusting one's views and beliefs. In contrast to the prior career stages, the maintenance stage is not characterized by tasks that progress in a predictable order related to age-related social expectations; the focus is on preservation, not progress. Consequently, research on this career stage has concentrated on the styles with which individuals retain their positions and preserve their identities.

Three basic styles of maintenance have been identified: upholding, updating, and innovating. The first style, upholding, involves continuing to do what one has done before, maintaining the same standards, and keeping things going in good condition. Workers who uphold their positions can hold their head up in pride

about what they continue to do. They take care of what they have accomplished and continue to be industrious. In contrast to this active upholding, there is passive holding on, in which one is more careless at work and shifts attention to waiting for retirement or looking for an early exit from the workforce. Rather than being careful in what they do, workers who hold on care more about how they are doing than what they are doing. They are careful how they behave when supervisors are watching. Instead of holding up their heads with pride, they hold up the organization by not doing their tasks as they should be done and in extracting rewards they do not deserve. The second positive style of maintenance, after upholding, is doing tasks better by updating. Workers using this style of maintenance adapt efficiently and respond to changing technology and tasks by keeping up with new developments. They stay fresh by updating their skills and knowledge, and in the process they reaffirm their self-concepts. While updating means doing tasks better, the third positive style of maintenance is to do tasks differently. Innovating involves actively seeking new ways of doing routine tasks and discovering new challenges. Of course, as they break new ground, some innovators may eventually decide to recycle to new positions and re-establish themselves. Each style, whether it be upholding, updating, or innovating, involves different trade-offs, yet workers using each style can maintain job success, satisfaction, and stability as well as provide security for their families.

#### **4.5. Disengagement**

At some point in a career, energy wanes and retirement looks good. It is time to decelerate by turning tasks over to others and planning for retirement. The task of actually disengaging from the work role goes better if the worker can decelerate first. Similar to other major transitions, small steps smooth the process. Deceleration can be accomplished by phase-outs, in which the individual works fewer and fewer hours over a period of time, occasionally lasting as long as several years. Another form of deceleration is called "bridge employment," in which an individual retires from a permanent position and then, before complete retirement, works at a part-time job, becomes self-employed, or works full-time at a temporary job. Bridge employment is becoming increasingly common as four out of five retirements are now early retirements and older workers remain in excellent health

well into their 70s, a trend that is compounding age discrimination in employment.

Deceleration, whether with one's permanent employer or through bridge employment, enables individuals to avoid the swift social isolation, reduced income, and bereavement that can accompany retirement from the work role. Nevertheless, at some point, the worker who lives long enough finally must retire and adjust to retirement living. Often, retirees engage in hobbies they once enjoyed during late childhood, bringing the career cycle full circle. If they live near grandchildren, they can once again play the games they enjoyed as children, thereby helping the next generation of workers begin to grow their careers and prepare to sustain our world. For, as Freud intimated, the world we know would perish without work.

### See Also the Following Articles

Career Counseling ■ Holland's Theory (Vocational Personality Types) ■ Indecision, Vocational ■ Job Analysis, Design, and Evaluation ■ Organizations, Careers in ■ Person-Environment Fit ■ Traits ■ Vocational Assessment in Schools ■ Vocational Interests ■ Work Adjustment ■ Work Environments

### Further Reading

- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.
- Kelly, K. R., & Lee, W. (2002). Mapping the domain of career decision problems. *Journal of Vocational Behavior*, 61, 302–326.
- Parsons, F. (1909). *Choosing a vocation*. New York: Agathon Press.
- Savickas, M. L. (2002). Career construction: A developmental theory of vocational behavior. In D. Brown and Associates (Eds.), *Career choice and development* (4th ed., pp. 149–205). San Francisco, CA: Jossey-Bass.
- Savickas, M. L., & Spokane, A. R. (Eds.) (1999). *Vocational interests: Their meaning, measurement and counseling use*. Palo Alto, CA: Davies-Black.
- Strong, E. K., Jr. (1943). *The vocational interests of men and women*. Palo Alto, CA: Stanford University Press.
- Sterns, H. L., & Huyck, M. H. (2001). The role of work in midlife. In M. Lachman (Ed.), *Handbook of midlife development* (pp. 447–486). New York: Wiley.
- Super, D. E. (1984). Career and life development. In D. Brown, & L. Brooks (Eds.), *Career choice and development* (pp. 192–234). San Francisco, CA: Jossey-Bass.
- Walsh, W. B., & Savickas, M. L. (Eds.) (in press). *Handbook of vocational psychology: Theory, research, and practice* (3rd ed.). Hillsdale, NJ: Erlbaum.



# Well-Being

**Richard E. Lucas**

Michigan State University, East Lansing, Michigan,  
USA

**Ed Diener**

University of Illinois at Urbana-Champaign, Champaign,  
Illinois, USA

1. The Hierarchical Structure of Subjective Well-Being
  2. Measurement Issues
  3. Theoretical Issues
  4. Effects of High Well-Being
  5. Summary
- Further Reading

## GLOSSARY

*affective dimension of well-being* Emotions and moods.

*emotions* Distinct evaluative (i.e., positive or negative) reactions to specific events.

*experience sampling method* A tool for assessing a person's subjective well-being that requires participants to track their well-being over time.

*life domains* Separable aspects of a life, including health, family life, occupation, social life, etc.

*life satisfaction* An individual's judgment that his or her life is going well.

*moods* Diffuse, affective feelings that are not necessarily tied to specific events.

*negative affect* Emotions and moods such as fear, anger, and sadness.

*positive affect* Emotions and moods such as happiness, joy, and excitement.

*subjective well-being* Well-being from an individual's own perspective.

Subjective well-being (SWB) is a broad construct that comprises multiple, separable components. These components are linked by a few key characteristics. First, as its name suggests, subjective well-being refers to well-being from an individual's own perspective. Although external observers may agree that the conditions in a person's life are less than ideal, that person may still feel a sense of satisfaction and well-being. SWB researchers recognize that the effects of objective circumstances are filtered through individuals' unique interpretations and goals, as well as their current life stage. Thus, the field of SWB places much weight on subjective evaluations. Second, SWB includes both positive and negative aspects of an individual's life. Although SWB researchers are interested in the factors that might prevent or improve negative states such as anxiety and depression, they are also interested in moving people from neutral to more positive states. Thus, positive constructs such as life satisfaction and positive affect play an important role in SWB research. Finally, SWB researchers often focus on evaluations of an individual's life as a whole. Although this can be accomplished either by asking for global judgments or by aggregating across numerous more specific judgments (including specific domain satisfactions or momentary experiences), SWB researchers often strive to understand how various circumstances and characteristics combine to create a life that is considered to be positive. Thus, SWB researchers value the utilitarian goal of creating the

greatest happiness for the greatest number of people, and they believe that by studying the causes, correlates, and outcomes of high levels of well-being, this goal can be attained. SWB constructs can be useful in a number of applied settings. First, because high well-being is a valued goal for most people, well-being can often serve as a useful outcome measure. Psychologists interested in improving work or community environments could measure SWB to determine whether interventions are successful. In addition, modern SWB research focuses on the effects that SWB can have on other outcome variables. It is possible that by measuring and even changing levels of SWB, psychologists will be better able to predict and affect important practical outcomes. This article reviews the topics that will be most useful to applied psychologists interested in understanding well-being. First, an overview of the components that are typically linked under the overarching construct of SWB are discussed. Then, the techniques that can be used to measure these components are presented. Finally, theoretical issues regarding the causes and consequences of SWB are discussed.

## 1. THE HIERARCHICAL STRUCTURE OF SUBJECTIVE WELL-BEING

Diener *et al.* noted that the construct of subjective well-being (SWB) has a hierarchical structure. At the highest and most abstract level is an overall evaluation of a person's life. Although the goal of SWB researchers is to understand this high-level evaluation, such an evaluation is difficult to measure directly. Instead, psychologists usually assess a variety of lower level constructs. For instance, researchers often distinguish between the affective and cognitive dimensions of well-being. On the affective side, researchers focus on emotions and moods. Emotions are thought of as distinct evaluative (i.e., positive or negative) reactions to specific events, whereas moods are often described as diffuse, affective feelings that are not necessarily tied to specific events. Presumably, if one's life is going well, he or she will experience high levels of positive emotions and moods and low levels of negative emotions and moods.

There are both theoretical and empirical reasons to distinguish positive affect (which includes such emotions and moods as happiness, joy, and excitement) from negative affect (which includes such emotions and moods as fear, anger, and sadness). Distinct

emotions often correlate more strongly with other emotions of the same valence than they do with emotions of the opposite valence. In addition, positive and negative affect are usually not so strongly correlated as to suggest that they reflect opposite poles of the same underlying dimension (although there is still debate about the bipolarity of affect). Furthermore, these separable aspects of affective experience are often differentially related to distinct predictor and outcome variables. For instance, positive affect is more strongly related to the personality trait of extraversion than is negative affect, whereas negative affect is more strongly related to the personality trait of neuroticism than is positive affect. Therefore, it is important to assess these variables separately.

Both positive and negative affect can also be distinguished from the more cognitive component of life satisfaction. Life satisfaction reflects an individual's judgment that his or her life is going well. It may be possible, for instance, for an individual to experience high levels of negative affect and still be satisfied with the conditions in his or her life. He or she may not value emotional experiences or may be pursuing some goal that is inherently satisfying yet which often leads to frequent experiences of negative affect. In 1996, Lucas *et al.* used multitrait-multimethod techniques to show that life satisfaction, positive affect, and negative affect were related but discriminable constructs.

The fourth component of SWB is domain satisfactions. Presumably, when constructing a life satisfaction judgment, an individual examines the various domains in his or her life (including health, family life, occupation, and social life), weighs the importance of these domains, and then aggregates the various judgments to get an overall evaluation of satisfaction with life. SWB researchers know, however, that this process does not occur perfectly. People do not have the motivation or ability to compile and aggregate across an exhaustive list of life domains. Therefore, when trying to understand the ways that people make life satisfaction judgments, or simply to ensure that certain important life domains are included in an overall evaluation, psychologists may want to assess satisfaction with specific life domains.

The hierarchical structure of SWB does not end with the four components of positive affect, negative affect, life satisfaction, and domain satisfactions. Each can be further divided. For instance, some researchers may be interested in specific emotions or reactions to specific emotion-eliciting experiences. Similarly, researchers may be interested in both satisfaction with a broad domain of life and satisfaction with narrower features

of that domain. Thus, for certain purposes, it may be useful to develop and use more precise measures of constructs that fall at a lower level. However, for most psychologists, the four components will provide a broad overview of an individual's subjective evaluation of his or her life.

## 2. MEASUREMENT ISSUES

Although the fact that SWB is a subjective phenomenon necessitates the use of some self-report methods, these are not the only measurement techniques that SWB researchers use. Instead, most SWB researchers share the view that multimethod assessment will provide the most informative perspective on the subjective quality of a person's life. In fact, by comparing information attained through different methods, SWB researchers hope to understand how evaluations are made, which in turn affects understanding of the factors that influence the construct. Next, the various methods that have been used to assess SWB are reviewed.

### 2.1. Self-Reports

The simplest, most cost-effective, and most intuitive way to assess SWB is simply to ask individuals how they would evaluate their lives. There are many different self-report measures from which to choose. For instance, self-report measures can include single items, such as "All things considered, how satisfied are you with your life?" or "In general, how happy are you?" Although single-item measures can be problematic (they may have poor reliability and random error cannot be canceled by aggregating across multiple items), existing research using these measures suggests that they have some degree of reliability and validity. For instance, in 2002 Diener *et al.* showed that a single-item measure of cheerfulness correlated very strongly with an aggregated multiple-item positive affect scale that had been administered multiple times during a 4-month period. More impressively, this single-item cheerfulness measure was able to predict respondents' income 20 years later. Lucas *et al.* showed that a single-item life satisfaction scale was sensitive to life changes such as marriage, widowhood, and unemployment. In addition, this measure also prospectively predicted the experience of life events. Thus, single-item measures can be useful when longer multiple-item scales cannot be used.

Although single-item scales show promise, multiple-item scales are desirable when feasible. Multiple-item scales are likely to be more reliable, they allow reliability to be assessed, and they often capture a broader construct than a single-item is able to assess. There are multiple-item measures that tap each of the components of SWB described previously. For instance, in 1985 Diener *et al.* created a five-item scale to measure life satisfaction. This scale correlates strongly with other SWB measures and can distinguish among various groups that are expected to differ in levels of well-being. Lucas, Diener, and Larsen and Larsen, Diener, and Lucas review a number of measures that can be used to assess positive and negative affect. Most domain satisfaction scales simply ask respondents to indicate how satisfied they are with a variety of areas in their lives.

### 2.2. The Experience Sampling Method

An important variation on standard self-report assessments is the experience sampling method (ESM). Unlike global questionnaire measures, ESM assessments require participants to track their well-being over time. For instance, participants may carry hand-held computers that signal alarms multiple times per day. Each time an alarm sounds, the participant completes a brief questionnaire. This technique allows researchers to avoid certain memory and aggregation problems that occur when respondents are asked to make global judgments, and it also allows researchers to investigate more dynamic aspects of well-being. For instance, psychologists can investigate the extent to which well-being fluctuates over time or correlates with various situational features.

### 2.3. Problems with Self-Reports

As with any self-report measure, self-reports of well-being are not without their problems. These reports have been shown to be influenced by a variety of irrelevant situational factors, including current mood, current weather, question wording, and availability of response options (although some studies have suggested that the effect of these factors is not strong). Thus, additional methods of assessment are useful when it is feasible to obtain them. One common alternative method is the use of observer reports.

### 2.4. Observer Reports

Although individuals are thought to have privileged information about their own subjective feelings of well-being, signs of these subjective feelings may be exhibited to others. For instance, individuals may share their thoughts about subjective well-being with close friends and family members, or these close acquaintances may be able to interpret nonverbal signs of well-being, such as smiling and crying. Thus, researchers interested in using multimethod assessment often obtain observer reports of well-being constructs. At their simplest, observer reports may consist of standard self-report instruments in which the wording has been changed to reflect the fact that respondents are rating someone other than themselves. In general, such informant reports correlate moderately with self-reports of well-being. It is also possible to obtain observer reports from individuals who do not know the target well. For instance, expert raters may be trained to recognize specific indicators of emotional experience, including nonverbal behaviors or even specific facial muscle movements.

### 2.5. Problems with Observer Reports

Although observer reports may prevent certain biases associated with self-report from occurring, they have their own set of problems. Observers cannot see subjective feelings directly, and they must infer subjective states from what the target person says and the behavior that the target person exhibits. Thus, both self-reports and informant reports can be supplemented with additional measurement techniques.

### 2.6. Physiological Indicators

Recent advances in psychophysiology and neuroscience have allowed researchers to identify specific indicators that tend to be linked with emotional experience. For instance, heart rate, blood pressure, bodily temperature, and skin conductance have been shown to be related to certain types of emotional experience. In addition, certain patterns of brain activity can be assessed using electroencephalograms, positron emission tomography scans, and magnetic resonance imagery. These patterns have been shown to correlate with individual differences and within-person changes in emotions.

### 2.7. Nonphysiological Indicators

Other researchers look to nonphysiological indicators of emotions and well-being. For instance, in 1998 Rusting reviewed evidence that trait and state levels of affect are associated with distinct cognitive processes that can be reliably assessed. For example, happy people are often quicker to recognize positive words than are unhappy people, and happy people are more likely to come up with positive words than negative words in word-completion tasks. Both cognitive and physiological measures are likely to correlate only weakly with self-reports of well-being. Thus, at this point, they probably cannot be entirely substituted for self-reports. However, non-self-report techniques can be a useful addition to a multimethod study, particularly when there are concerns about social desirability or other response problems.

## 3. THEORETICAL ISSUES

Psychologists who are interested in developing theories of well-being generally focus on two questions: What are the factors that lead to high levels of well-being? and What effects do high levels of well-being have in people's lives? Whereas the first question guided early research in the field, modern research increasingly focuses on the latter. Thousands of studies have been conducted, and researchers now know a great deal about these questions.

### 3.1. Demographic and Situational Characteristics

Early research on subjective well-being tended to focus on the demographic and situational characteristics that correlate with well-being. Perhaps the most surprising finding from these studies was that external circumstances seem to account for very little variance in reports of well-being. For instance, when researchers assess a wide variety of demographic factors (sex, age, income, health, etc.) and life circumstances, these variables only account for 15–20% of the variance in well-being measures. Even when extreme circumstances are examined, the effects are not large. In perhaps the most famous of these studies, Brickman *et al.* in 1978 showed that individuals who had won large sums of money in lotteries were not much happier than a matched control group. Similarly, individuals with spinal cord injuries were not much less happy than a

matched control group. One possible explanation for this finding is that individuals initially react to life events but then gradually adapt back to their previous levels. Support for this hypothesis was found in longitudinal studies that traced SWB after the experience of these life events. For instance, in 1996 Suh *et al.* found that only recent life events (those occurring within the past 3 months) affected current levels of well-being.

### 3.2. Adaptation Processes

However, recent large-sample longitudinal studies have suggested that adaptation processes are complicated and that life events can have important long-term consequences for people's well-being. Lucas *et al.*, for instance, showed that the experience of unemployment causes long-term changes in people's levels of life satisfaction. Although people who had been unemployed in the past started to rebound toward their initial levels of satisfaction following a bout of unemployment, they seemed to be permanently affected by this experience and tended not to return all the way back to initial levels. Similarly, although individuals who experienced the death of a spouse did return very close to their prewidowhood levels of satisfaction, this process was very slow, taking an average of 8 years. Furthermore, Lucas *et al.* showed that there is much variability in people's reactions to many different types of life events. Although, on average, people may adapt to events such as marriage, there are considerable individual differences in reaction and adaptation, and some people experience long-term changes in satisfaction following these events.

### 3.3. Traits and Personality Characteristics

These individual differences in reactions to life circumstance suggest that certain personality characteristics play an important role in determining levels of SWB. In contrast to the weak effects of demographic characteristics, personality traits are often moderately to strongly correlated with well-being. In particular, the personality traits of extraversion and neuroticism have been identified as moderate to strong correlates of SWB. Although it is unclear exactly why these constructs are related, a number of researchers have linked personality and affect through underlying physiological systems. For instance, in 1981 Gray suggested that there are two physiological systems that underlie the

personality traits of extraversion and neuroticism: a behavioral activation system (BAS) that governs reactions to signals of reward and nonpunishment and a behavioral inhibition system (BIS) that governs reactions to punishment and nonreward. Tellegen and others have suggested that the BAS is closely linked to extraversion and positive emotionality, whereas the BIS is closely linked to neuroticism and unpleasant affect. The suggestion that extraversion and neuroticism share common physiological underpinnings with positive and negative affect is supported by the fact that affective traits are moderately stable over time and have moderate to high heritability.

### 3.4. Nontrait Personality Characteristics

It is also likely that additional nontrait personality characteristics can directly affect well-being or can serve to moderate the effects of external circumstances on well-being. For instance, in 1986 Emmons suggested that by understanding individuals' goals, SWB researchers can better predict well-being outcomes. Emmons showed that simply having valued goals is associated with higher levels of well-being. In addition, various characteristics of one's goals (including whether goals conflicted with one another) are related to SWB outcomes. Other researchers have shown that the effect of specific external circumstances and demographic characteristics such as one's income may differentially affect well-being depending on one's goals and values.

### 3.5. Circumstances in Context

Finally, it is important to consider life circumstances in the context of the lives in which they are being experienced. In 1976, Campbell *et al.* argued that these life circumstances only have meaning when they are compared to various standards that an individual might have. For instance, a certain income level may be seen as desirable or undesirable depending on the income that an individual had in the past, his or her expectations regarding that income, or the income that salient and proximal individuals are earning. Thus, it may be important to assess various contextual factors and standards of comparison before a complete understanding of the factors that influence well-being can be attained. However, in 1997 Diener and Fujita reviewed evidence showing that even these comparison

processes are flexible and function differently across individuals. For instance, it is not the case that surrounding oneself with individuals who are worse off than oneself will inevitably lead to higher well-being. Although it is possible that individuals will recognize that they are better off than those with whom they can compare, this comparison information can be used in different ways. On the one hand, downward comparisons can make one value one's current status, which could lead to high well-being. On the other hand, downward comparisons can remind individuals that things can be much worse. This pessimistic interpretation might lead to lower well-being. The choice of comparison targets and the way this comparison information is used vary across individuals.

#### 4. EFFECTS OF HIGH WELL-BEING

Although fostering high levels of SWB is itself an important and valued utilitarian goal, other researchers have wondered about the effects of high well-being. Is it the case that happy people become apathetic and lazy because they feel like they have achieved all their goals? Or do happiness, positive affect, and life satisfaction serve to motivate individuals to reach even higher goals? Much recent research seeks to answer these questions. If SWB is functional, then this will have important implications for applied psychologists.

##### 4.1. Happiness and Productivity

There is a long line of research in the organizational literature that is dedicated to determining whether happy workers are more productive than their less happy coworkers. Much of the early research on the happy/productive worker hypothesis was based on the very simple idea that well-being (particularly the job satisfaction component on which most early research focused) was an attitude. If attitudes affect behavior, then positive attitudes should elicit positive behaviors. Recently, well-being researchers have suggested that affect and satisfaction may function in many more complex ways, and these functions may have implications for outcomes such as worker productivity.

##### 4.2. Positive Social Outcomes

Based on a 2003 review by Lyubomirsky *et al.*, Lucas and Diener specified six types of outcomes that may be relevant for the happy/productive worker hypothesis.

First, there is a great deal of evidence that SWB, particularly the positive affect component, is linked with positive social outcomes. Correlational research shows that feelings of sociability and social activity are correlated with positive affect. In addition, both correlational and experimental studies show that induced positive mood can cause changes in helpfulness, cooperation, and friendliness and self-disclosure. Happy people like other people more than do unhappy people, and people like happy people more than unhappy people. Thus, higher levels of positive affect are likely to lead to greater social outcomes and increases in one's social resources.

Happiness may also lead to benefits in motivation. For instance, researchers interested in the structure of mood have shown that feelings of energy and activity are strong markers of the positive affect dimension. Correlational experience sampling studies have also shown that people feel and act more energetic and active when they are feeling happy. In 1988, Cunningham experimentally manipulated positive affect and showed that when in positive moods people were more likely to prefer active and exciting activities. Thus, happy people may simply have more energy to complete important tasks.

Perhaps more important, however, there appear to be effects beyond just boosts in energy. Happy people are also more likely to set higher goals and approach those goals with greater confidence. Correlational evidence shows that happy individuals are more confident and assertive, and experimental evidence shows that manipulating happiness leads to higher levels of self-confidence and more positive self-evaluations. Furthermore, manipulating levels of affect can lead to changes in the goals that individuals set for themselves. Thus, if happy individuals set higher goals and have greater confidence about their potential for success, they may, in fact, achieve more than their less happy counterparts.

A fourth potential benefit of subjective well-being is that a variety of emotional variables are empirically linked with health and physical well-being. Specifically, individuals who experience high levels of stress and negative affect experience more health problems than do individuals with low levels. In addition, there is increasing evidence from both correlational and experimental studies that positive mood can also have a protective effect, although it is unclear whether these effects are direct or whether they simply moderate the effects of negative affect on health. These protective effects may even translate into greater longevity for happy individuals.

### 4.3. Effects of Happiness and Well-Being in the Cognitive Realm

Finally, there are two potential effects of happiness and well-being in the cognitive realm. First, there is evidence from experimental studies of induced mood that higher levels of happiness and positive affect lead to greater creativity. Individuals who have undergone positive mood inductions tend to make more unusual responses to word association tasks, to see more creative connections among concepts, and to list more unusual exemplars for categories. In situations in which novel and creative responses are required, individuals who are in a positive mood may have an advantage.

In addition, mood seems to affect the specific cognitive strategies that individuals are likely to use. On the one hand, unhappy individuals have often been shown to think more carefully about arguments and to rely on stereotypes less heavily than do happy people. This may lead to an advantage for less happy people when attention to detail is required. Happy people, on the other hand, may process information more efficiently, using effective heuristics. Each of these strategies may be more or less useful depending on the specific situation. At this point, it is also unclear whether distinct emotional states lead directly to the differential use of cognitive strategies. Martin *et al.*, for instance, in 1993 argued that moods simply provide information about the state of the world. This information is then used as one factor in the decision to use a specific strategy. Martin *et al.* suggested that mood is not inevitably linked with a particular cognitive outcome. In support of this viewpoint, in 1994 Bodenhausen *et al.* found that although individuals in positive moods were more likely than less happy individuals to use stereotypes when making judgments, the happy individuals could overcome these stereotypes if they were told they would be held accountable for their decisions.

## 5. SUMMARY

Applied psychologists will find SWB constructs to be useful in a variety of settings. On the one hand, SWB can be seen as the ultimate utilitarian outcome measure. Presumably, any intervention that is designed to effect change in an individual, an organization, or a community could be evaluated in terms of its effect on SWB. In addition, SWB may be functional. Various components of well-being have been shown to have a

causal impact on such factors as sociability, motivation, health, and cognition. Therefore, intervention programs aimed at increasing well-being may, in turn, lead to changes in many other areas. To successfully incorporate SWB measures into their programs, applied psychologists must first determine which aspect of SWB they wish to study and then develop measures that can adequately tap these dimensions.

### See Also the Following Articles

Emotion ■ Social Comparison and Subjective Well-Being ■ Traits

### Further Reading

- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616.
- Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173–191). New York: Guilford.
- Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality Psychology*, 80, 804–813.
- Diener, E., Nickerson, C., Lucas, R. E., & Sandvik, E. (2002). Dispositional affect and job outcomes. *Social Indicators Research*, 59(3), 229–259.
- Diener, E., & Oishi, S. (2000). Money and happiness: Income and subjective well being across nations. In E. Diener, & E. M. Suh (Eds.), *Culture and subjective well being* (pp. 185–218). Cambridge, MA: MIT Press.
- Diener, E., Scollon, C. N., & Lucas, R. E. (2004). The evolving concept of subjective well being: The multifaceted nature of happiness. *Advances in Cell Aging and Gerontology*, 15, 187–219.
- Eid, M., & Diener, E. (in press). Global judgments of subjective well being: Situational variability and long term stability. *Social Indicators Research*.
- Harker, L., & Keltner, D. (2001). Expressions of positive emotions in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of Personality and Social Psychology*, 80, 112–124.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job-satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127, 376–407.
- Larsen, R. J., Diener, E., & Lucas, R. E. (2002). Emotion: Models, measures, and individual differences. In

- R. G. Lord, R. J. Klimoski, & R. Kanfer (Eds.), *Emotions in the workplace* (pp. 64–106). San Francisco: Jossey-Bass.
- Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology*, 84(3), 527–539.
- Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the setpoint for life satisfaction. *Psychological Science*, 15, 8–13.
- Lucas, R. E., & Diener, E. (2003). The happy worker: Hypotheses about the role of positive affect in worker productivity. In M. R. Barrick, & A. M. Ryan (Eds.), *Personality and work* (pp. 30–59). San Francisco: Jossey-Bass.
- Lucas, R. E., Diener, E., & Larsen, R. J. (2003). Measuring positive emotions. In S. J. Lopez (Ed.), *Positive psychological assessment: A handbook of models and measures* (pp. 201–218). Washington, DC: American Psychological Association.
- Meloy, M. G. (2000). Mood-driven distortion of produce information. *Journal of Consumer Research*, 27, 345–359.
- Salovey, P., Rothman, A. J., Detweiler, J. B., & Steward, W. T. (2000). Emotional states and physical health. *American Psychologist*, 55, 110–121.
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76(5), 820–838.
- Wright, T. A., & Staw, B. M. (1999). Affect and favorable work outcomes: Two longitudinal tests of the happy-productive worker thesis. *Journal of Organizational Behavior*, 20, 1–23.
- Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from Experience Sampling data. *Journal of Research in Personality*, 34(2), 178–197.



# Women's Health

Pilar Sánchez-López

Complutense University, Madrid, Spain

1. Introduction
  2. Gender Bias in Health Research and Care
  3. Life Cycle: Differential Health Characteristics of Elderly Women
  4. Some Cultural Differences: Interactions among Sex/Gender, Social Class, Ethnic Groups, and Culture
  5. Mental and Physical Health
  6. Women's Health on the Internet
- Further Reading

This has given rise to consideration of the existence of possible gender bias in matters related to health, a bias that may have facilitated the so-called “invisibility” of women in areas of health research, professional training, and care. However, there is a differential profile of women with regard to health (both physical and mental).

## GLOSSARY

**bias** Trend in the recording, analysis, interpretation, and/or revision of data leading to the conclusions drawn from them being false in some way or another.

**health** State of physical, social, and mental well-being.

**morbidity** Lack of physical or psychological well-being.

**mortality** Frequency of deaths in a given population during a specific time period.

**prevalence** Number of registered cases of a given illness divided by the number of persons in a particular population during a specific time period.

**risk** Any variable of personal, environmental, or biological lifestyle that might be associated with a greater likelihood of suffering a specific health impairment.

During the past 20 years or so, significant data have appeared on the differences between men and women regarding the way in which they become ill and die.

## 1. INTRODUCTION

The worst consequences of the possible gender bias in health probably are that less effective provisions are made in community health programs, the conclusions of empirical studies are less valid, and estimates of the prevalence rates for health problems are less effective given that all of these are frequently based on what is known of illnesses and problems from the more studied sex, that is, the male one.

Gender bias in health questions is already apparent when statistical data on health are collected and published. In many countries, even developed ones, there were no differential statistics by sex on morbidity and mortality until recently, so conclusions could not be reached regarding the possible existence of differential rates or their causes.

Moreover, it is true that there is an increasingly greater awareness of this gender bias in works related to health, so there are an increasing number of studies of subjects related to gender bias and how it is exerted

on the medical/psychological act, on diagnoses, on treatments, on community action plans, and on how health-related matters are researched.

What all of this implies is that women are beginning to become "visible" to health professionals who increasingly recognize that there may be both differences and similarities between men and women in the ways in which they succumb to illness, in how they are cured, in how they metabolize pharmaceutical products, and in how they react to therapy in the etiology of illness.

This recognition of the differential fact has two main causes: the stubbornness of the facts (which serve as a constant indicator) and the growing feminization of health-related professions (which has led to an increase in the sensitivity of health professionals themselves to matters related to women's health).

Be that as it may, the fact is that since the late 20th century, there has been an explicit recognition that women's health is a specific area of study. Scientists, health professionals, politicians, and the like are beginning to realize that there are health-related aspects that affect men and women in differential ways (both quantitatively and qualitatively).

Because it is recognized that women do not necessarily react to preventive, diagnostic, and treatment processes for illness (both physical and mental) in a similar manner to that of men, health-related research is beginning to include, in systematic fashion, feminine subjects. Until quite recently, this was still quite infrequent.

Naturally, "women" itself is not a compact uniform unit. The health issues may vary, to a large extent, from one woman to another. This depends on variables such as a woman's stage in the life cycle, her social class, her ethnicity, her working conditions; her social and family environment, recent or distant life events, and her personality style. Many of these factors have similar effects on the health of both men and women, whereas others do not. Under the heading "women's health," one can find factors that are known or suspected to have differential effects on the health of women, to a greater or lesser extent, and the illnesses that, for one reason or another, are more related to women.

## **2. GENDER BIAS IN HEALTH RESEARCH AND CARE**

Gender bias in health research and care can be displayed in three different ways: false assumptions of equality

between the sexes, false assumptions of differences between the sexes, and reproduction-oriented limits.

### **2.1. False Assumptions of Equality Between the Sexes**

It is assumed that problems, risks, and protecting factors in the health of men and women are the same when in fact they are not the same. A typical example is that of the efficacy of aspirin in stroke prevention. Clinical tests usually do not include a sufficient number of women to produce any conclusion (instead, an extrapolation is made from studies carried out on men), but those studies that do include sufficient women find that they do not always benefit from aspirin in preventing primary strokes. This difference between the sexes seems to be due to the existence of different stroke mechanisms in men and women.

### **2.2. False Assumptions of Differences between the Sexes**

Biological and psychological differences also are assumed to exist when in fact what exist are similarities. Until only a few years ago, cardiovascular disease was not believed to be a "female illness." Unfortunately, many women still believe this; in a 1995 Gallup poll, only 19% of American women expected to die from cardiovascular disease, whereas two-thirds of the women interviewed expected to die from cancer. Nevertheless, in industrialized countries, cardiovascular disease is the main cause of death not just among men but also among women, and not just among elderly women but also among younger women. As a result of this gender bias, differences have been produced—and on occasion are still being produced—between men and women with respect to access to medical care. Women wait longer before seeking medical help, and even today they are less likely to receive aggressive intervention once their illness has been diagnosed. Consequently, they are more likely to die from their cardiovascular disease than are men.

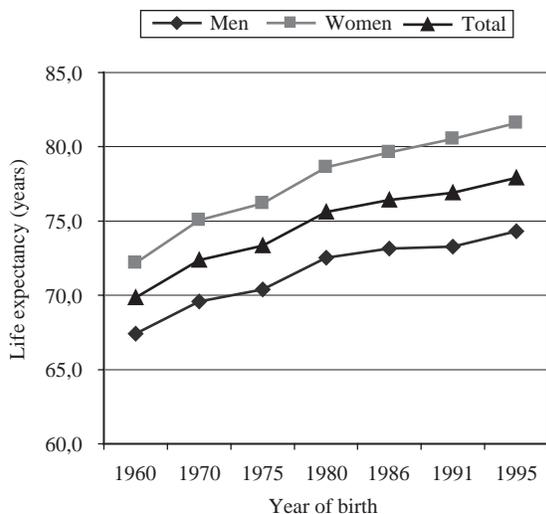
### **2.3. Reproduction-Oriented Limits**

Differential problems of women's health are assumed to be restricted to just one area: reproductive health. Although these problems are important, they are not the only ones. Thus, differential attention to women's health cannot be restricted to the areas of pregnancy, contraception, and family planning.

### 3. LIFE CYCLE: DIFFERENTIAL HEALTH CHARACTERISTICS OF ELDERLY WOMEN

Each stage of woman's life cycle has specific health-related aspects: pregnancy, menstruation (e.g., menstrual disorders, premenstrual syndrome), contraception, abortion, maternal morbidity, and infertility. But it may be the final stage of the life cycle that is receiving the most attention currently given the sociological changes taking place worldwide. These changes are leading to a constant increase in the number of elderly people, particularly the number of elderly women given that women have a longer life expectancy than do men (Fig. 1).

Thus, the study of health in the final stage of the human life cycle is becoming particularly important with regard to women and health. However, emphasis must be placed only on studying the fatal illnesses at this stage but also on studying the conditions and disorders that, although not fatal, have important functional consequences and affect people's quality of life. In the case of women, these conditions include matters such as menopause-related changes, functional and cognitive deterioration, osteoporosis, and urinary incontinence. The prevalence of these disorders increases considerably with advancing age, making them very important for the population of elderly women, who have a greater likelihood of surviving to an advanced age than do men.



**FIGURE 1** Life expectancy at birth of men and women in a developed country.

Furthermore, older people frequently suffer from chronic disorders, so it becomes essential not to assess illnesses in isolation but rather to study the global impact of these illnesses as a whole on any person. To evaluate an elderly woman's state of health, one must focus on her functional state, her incapacity, and her level of independence—and not just on a list of her illnesses—regardless of the interaction among them.

But neither should it be forgotten that recent research into healthy aging has helped to cast doubt on the stereotypes that regard aging as a period of inevitable frailty and decline. Individual differences also play an important role here. Epidemiological studies have shown that many elderly women still function at a high physical level, still maintain good cognitive abilities, and still retain a state of well-being for many years of old age. This has meant that research into the determining factors of a healthy old age is ever more frequent.

Subjects that must be tackled when dealing with the health of elderly women include menopause (which is being studied more and more), risks and benefits of hormone replacement therapy, musculoskeletal diseases such as osteoporosis (with important implications due to the risk of fractures, for its functional consequences, and for the use it requires of medical services and long-term care), arthritis (which is clearly the most frequent chronic condition in the elderly population, at least in developed countries), and urinary incontinence (with its possibly important psychological, social, and physical impacts). Sensory impairment must also be considered given that sight has a great impact on global functioning and quality of life and that impairment ranges from easily correctable cases (e.g., cataracts) to the most serious cases (e.g., macular degeneration, diabetic retinopathy, glaucoma). Also, hearing impairment heavily influences quality of life for the elderly, and its prevalence rapidly increases with age. Finally, there is loss of taste and smell, and although this has less impact on the quality of life, it is nonetheless important. For example, in relation to food, the loss of taste and smell may, on the one hand, lead to a reduction in calories ingested (because food is less appetizing) and may, on the other, hamper the ability to distinguish food that has gone "bad" (with an increasing likelihood of food poisoning). Last is senile dementia and especially Alzheimer's disease. Women themselves are not more susceptible, but because they live longer than do men, they have a greater chance of suffering from dementia and of living a long time with its symptoms.

#### **4. SOME CULTURAL DIFFERENCES: INTERACTIONS AMONG SEX/ GENDER, SOCIAL CLASS, ETHNIC GROUPS, AND CULTURE**

Existing data show, logically, that the risk of suffering illnesses varies among different subgroups of women, characterized by ethnic group and/or socioeconomic level. Most studies are still descriptive, and the few that center on the explanatory level in general opt for the interaction between environmental factors (e.g., differences in the likelihood of exposure to risk factors) and biological factors (e.g., genetic differences affecting the metabolism of these risk factors),

Coherent relationships have been found between women's state of health and their ethnic and class groups. In general, the financially poorest women and those belonging to minorities have poorer health. In any case, there are exceptions to this general rule, meaning that nonsimplistic explanations must be sought for this association and attempts must be made to develop causal theories that would enable an improvement to be seen in the capacity for intervening to improve health levels.

The case of obesity may serve as an example. The frequency of this disorder has shown a steady increase during recent years, especially in countries such as the United States, where one-third of White women and half of Hispanic and African American women are considered to be obese. Moreover, there is an important inverse correlation between socioeconomic status and obesity among women, although it has not been possible to establish the causal direction of this relationship. Causality almost certainly works both ways insofar as poverty increases the risk of women becoming obese and obesity increases the risk of women's socioeconomic level declining.

Cultural differences also appear when comparisons are made between industrialized countries and developing countries. In this case, there are two typical examples: maternal mortality and violence inflicted on women.

##### **4.1. Maternal Mortality**

Although maternal mortality is a practically unknown phenomenon in the developed world, it still has tragic dimensions in developing countries. The fundamental cause of these drastic differences is related to the adequate distribution of health resources that enable factors such as uterine bleeding and other birth-related crises to be controlled.

##### **4.2. Violence Inflicted on Women**

Violence inflicted on women ranges from what is known as female circumcision (or, less euphemistically, as female mutilation) to physical and psychological ill treatment, either in the home or in areas of war and social conflict. These latter two types of violent situations are produced both in developed and poor countries, as shown by recent history and in daily newspaper headlines. But it is true that the frequency of both types may vary, depending on the social characteristics, and that the consequences may be different for both the victim and the aggressor, depending on social sensitivity to this type of aggression.

#### **5. MENTAL AND PHYSICAL HEALTH**

An example of representative data regarding predominantly physical illnesses is reported in [Table I](#). With regard to illnesses related to the immunological system, the basis for the differential prevalence between the sexes is related to the greater response by the immunological system of men, probably due to both the differences in sexual hormones and the influence of environmental factors. Specifically, the incidence of asthma would be related, on the one hand, to the greater exposure of children to factors increasing the incidence of asthma (which would explain the lower prevalence in girls under 12 years of age) and, on the other, to the greater likelihood of women being diagnosed as asthmatics and the influence of sexual hormones on predisposition to allergy. These latter two factors would explain the greater incidence in females over 12 years of age.

With regard to mental illnesses, there is no longer any doubt that, in general, men differ from women in the prevalence and severity of mental illnesses. [Table II](#) provides a list of the major disorders, including reasons for such a differential prevalence.

#### **6. WOMEN'S HEALTH ON THE INTERNET**

Information regarding women and health on the Internet is plentiful. Three types of Web addresses on women and health can be differentiated. First are those that refer to women's studies and that consist of pages or Web sites on women and including lists (in alphabetical order) of centers, departments, and programs of

TABLE I

Differences between the Sexes in Illnesses Related to the Immunological System, Cardiovascular Illnesses, and Cancer

	Differences between sexes	Sex-age interaction
Illnesses related to immunological system	Hypothyroidism: between 3 and 20% for women and between 1 and 7% for men	
Thyroids	Hyperthyroidism: between 2 and 19% for women and 1% for men Thyroid cancer: three times greater in women than in men	
Rheumatic arthritis	Better prognosis for women than for men	
Multiple sclerosis	Affects mostly women (65% of total population affected)	Women most affected at beginning of 20s and men at end of the 20s
Lupus	Predominance in women	From ages 15 to 45 years, the ratio between women and men is 12 to 1; in infancy and old age, the ratio is 2 to 1
Asthma	Severity greater in women; differences not too great	Before years of age, 12 lower prevalence in girls; after 12 years of age, greater incidence in females
Cardiovascular	Prime cause of death among males and females; differences in prevalence according to type of illness	Principal cause of death among younger women; accounts for half of deaths in older women
Cancer	Similar prevalences except for those related to reproduction	
AIDS	Less than half (40%) of cases found among women	

women's studies. As an extra section, there are subjects related to women's health (e.g., <http://creativefolk.com> and [www.research.umbc.edu](http://www.research.umbc.edu) in the United States). In Europe, the Women's Studies EuroMap Web site (<http://women-www.uia.ac.be>), which offers detailed information on WISE (the European professional association for women's studies), is particularly noteworthy.

Second, are the pages of official bodies, associations, and organizations that provide information on health-related services throughout the life cycle of women. Examples include those for Canada ([www.swc-cfc.gc.ca](http://www.swc-cfc.gc.ca)) and the United Nations ([www.un.org](http://www.un.org)). Also included are the pages for the Society for Women's Health Research ([www.womens-health.org](http://www.womens-health.org)) and for the International Women's Health Coalition ([www.iwhc.org](http://www.iwhc.org)). In Europe, one page ([www.aviva.org](http://www.aviva.org)) collates information on women in all aspects, including those that are health related. There are also examples of national pages, including those of the Women's Health Council ([www.nwci.ie](http://www.nwci.ie)), which represents organizations of women throughout Ireland; the Conseil National des Femmes Luxembourgeoises ([www.cnfl.lu](http://www.cnfl.lu)), which

collates official documents from the various international conventions on women, statements, and resolutions from bodies such as the United Nations and the European Community; and the Instituto de la Mujer in Spain ([www.mtas.es/mujer/](http://www.mtas.es/mujer/)) and the Sociedad Iberoamericana de Información Científica in Latin America, ([www.siicsalud.com](http://www.siicsalud.com)), which provides information on health in general, and on women's health in particular, with an excellent database.

Third are the Web gateways "women and health" and the online magazines. There are both general and specific examples, and both kinds are geared to women interested in a wide variety of topics about their state of health who can find what they need there. For example, the "webofwomen" (Spanish) provides information on lifestyles, social assessment, health, key concepts in health, Web addresses, topical information, meetings of interest, and the like ([www.webdelamujer.com](http://www.webdelamujer.com)). Some are commercial in character and include advertisements for commercial products and services for women. They also offer the opportunity to consult online with experts on various topics as well as online questionnaires on cardiovascular

**TABLE II**  
**Differences between the Sexes in Mental Disorders**

	<i>Basis of differences between sexes</i>	<i>Differences between sexes</i>	<i>Sex-age interaction</i>
<i>Mood disorders</i>	Women, all things being equal, tend to seek more professional help. There are also differences in the response to treatment	Higher rates in women	Differences increase during puberty
<i>Anxiety disorders</i>	Women show higher levels of arousal, greater psychophysiological response to stress, higher frequency of somatic symptoms, and greater sensitivity to anxiety symptoms	High rates in women; one-third have some type of attack throughout their lives	
<i>Posttraumatic stress</i>	Differences between sexes appear when this syndrome is related to violence and a particular moment in the life cycle.	Frequency two times higher in women than in men; duration of symptoms four times higher in women than in men	At the end of adolescence and the beginning of adult life, greater risk of suffering this type of violence
<i>Eating disorders</i>	Biological vulnerability exists in integrating and tackling the following factors: <ul style="list-style-type: none"> <li>• Sociological factors: role of woman in cultural expectations regarding weight and physical appearance</li> <li>• Psychological factors: characteristics of personality, family functioning, and stressful life events</li> </ul>	<ul style="list-style-type: none"> <li>• Anorexia: 0.28% average incidence among young women</li> <li>• Bulimia: 1% average incidence among young women</li> </ul>	Increase of incidence among young women between 15 and 24 years of age and from 50 years of age onward
<i>Addictive disorders</i>		Higher in men than women.	

symptoms, depression, anxiety, and the like. Among them are pages devoted exclusively to certain of these topics, for example, one devoted to pregnancy and its circumstances (<http://womenshealth.about.com>).

### See Also the Following Articles

Aggression and Culture ■ Gender and Culture ■ Gender Role Development ■ Prejudice and Discrimination

### Further Reading

- Goldman, M. B., & Hatch, M. C. (2000). *Women and health*. San Diego: Academic Press.
- Pregler, J. P., & DeCherney, A. H. (2002). *Women's health principles and clinical practice*. San Diego: Academic Press.
- Shimp, L., & Smith, M. (2000). *Twenty common problems in women's health care*. New York: McGraw-Hill.
- Veet, L. L. (2000). *Women's health care case studies*. San Diego: Academic Press.
- Zerbe, K. J. (1999). *Women's mental health in primary care*. San Diego: Academic Press.



# Work Adjustment

Beryl Hesketh

University of Sydney, Sydney, New South Wales, Australia

1. Structural Aspects of Work Adjustment
  2. Work Adjustment Process
  3. Work Adjustment Assessment
  4. Work Adjustment Counseling
- Further Reading

## GLOSSARY

**adjustment style** Modes of adjustment that individuals use to achieve a better fit with their work.

**correspondence** The match between the skills, abilities, interests, and needs of the individual and the skill and ability requirements of the job as well as the rewards offered by it.

**person–environment fit** A summary statement of the match between a person and the environment, typically used in reference to a work environment.

**Minnesota Theory of Work Adjustment (TWA)** A comprehensive structural and dynamic framework used to explain how individuals maintain and improve their job satisfaction and satisfactoriness.

**work adjustment** Ongoing process by which employees and employers seek to achieve and maintain a fit between their respective needs and supplies.

Work adjustment is a term that is used to describe a desired state of correspondence or fit between an individual's capabilities and wants and the commensurate requirements and rewards or opportunities provided

within an organization. The term can also refer to the processes used by individuals and their employers to achieve work adjustment. Work adjustment has been used widely within the rehabilitation and vocational literature. The Minnesota Theory of Work Adjustment originated from research undertaken for the Rehabilitation Services Administration but has gained status as a major general fit theory in vocational and industrial/organizational psychology. In the rehabilitation context, work adjustment is part of the recovery process following illness, accidents, or other trauma that severely affect performance capabilities and limitations, requiring an adjustment to a different set of work expectations. In a more general context, work adjustment becomes important whenever change occurs within the individual or in the organizational requirements. Technological change, globalization, mergers, takeovers, and other external events have increased the requirement for ongoing work adjustment and adaptive performance. Change may also occur as a part of life span development, such as in transitions from school to tertiary study and work, transitions from work to unemployment, and retirement transition. Although this article focuses on a narrower meaning of work adjustment, the term in its broader sense incorporates a range of concepts, such as adaptive performance, coping behavior, change self-efficacy, role-breadth self-efficacy, ergonomic interventions, selection, and training.

## 1. STRUCTURAL ASPECTS OF WORK ADJUSTMENT

The Minnesota Theory of Work Adjustment (TWA) provides a useful framework for examining work adjustment and relating the concept to other theories and components of industrial/organizational and vocational psychology. In its original form, TWA outlines a structural component that describes people and work environments in terms of the correspondence between the knowledge, skills, and abilities (KSAs) required in a job and those provided by the individual, giving rise to satisfactory performance. This is the performance component of work adjustment, which is valued from an employer perspective. However, it also contains a motivational component that is dependent on the correspondence between the needs, values, and interests of the individual and the corresponding rewards and reinforcers provided by the job, giving rise to satisfaction. Continued interaction (tenure) between an individual and the job/organization is a function of how satisfied an individual is and how satisfied the employer is with the individual's performance. Higher levels of satisfaction and performance contribute to an increased period of interaction between the parties (tenure). Dissatisfied individuals do not necessarily perform less well, but they are more likely to leave a job. In most organizations, if an individual is not performing as required, the appointment is terminated.

There are numerous studies directly testing the main propositions of fit in TWA, but studies on employee selection also represent a large body of research relevant to the relationship between the required KSAs and those supplied by the individual, a major component of TWA. Validity studies of ability or personality dimensions that take into account "fit" with job requirements when choosing appropriate predictor and criterion dimensions are more compatible with TWA than are those that focus only on the relationship between predictors and performance without the moderating influence of job requirements.

In terms of the motivational component of the TWA, much vocational research has emphasized the match between individual needs, interests, and values and outlets for these as a predictor of job satisfaction and eventually turnover. In addition to the work value and ability (KSA) dimensions used in the TWA, other constructs can be used as a basis for commensurate assessment of people and their environments in relation to both the performance component and the satisfaction and motivation component. Holland's hexagonal theory of interests

(realistic, investigative, artistic, social, enterprising, and conventional) provides one such example, as do multi-attribute test batteries and the Five Factor model of personality. Tracey has outlined a spherical model of interests, the Personal Globe Inventory, and Gustaffson and Mumford and also De Fruyt argue for fit on the basis of a configurational or person-centered approach.

## 2. WORK ADJUSTMENT PROCESS

Adjustment at an individual level can be likened to adaptation at the species level. Dawis and Lofquist outline style variables that describe the extent to which discorrespondence is tolerated by both individuals and organizations (flexibility), the type of action initiated to improve correspondence (active or reactive modes of adjustment), and the length of time during which the parties continue to make an effort to achieve adjustment before terminating the relationship (perseverance). An active mode of adjustment relates to the activities undertaken by the individual to change the job (e.g., negotiating to eliminate aspects of the work that are too difficult or not enjoyable) or to activities the employer initiates to change the individual (e.g., training courses to provide the employee with the desired skills). Reactive modes of adjustment relate to changes made to the self. For example, in the case of individuals, they may recognize that their expectations are unrealistic and hence modify their work-related goals and expectations. From the perspective of the environment, an employer may redesign the job to take better account of the performance capabilities and limitations of the individual. Active and reactive processes of adjustment are continuous, and together aim to achieve better work adjustment.

In the recent literature in industrial and organizational psychology, terms used that are similar to the style variables in the TWA include flexibility, adaptability, and coping behavior. Perseverance in achieving adjustment is related to the extent to which an individual believes that he or she has the skills and capacities to make the necessary active and reactive adjustments.

## 3. WORK ADJUSTMENT ASSESSMENT

Determining the extent to which an individual is adjusted to a work situation involves discerning the fit between the individual's performance capabilities and those required

by the job as well as the fit between the individual's needs and the rewards of the work. Measures used draw on the individual differences literature for assessing cognitive and other skills, personality, interests, needs, and values. Job analysis techniques are also needed to analyze jobs to determine job requirements. This level of assessment provides a diagnostic basis for helping individuals achieve a better fit. Outcome measures used to assess work adjustment include job satisfaction, performance, and tenure. Satisfaction can be considered as part of the broader family of constructs related to mental health and well-being, whereas recent models of performance highlight the need to consider task, contextual, and adaptive performance. Tenure, the ultimate assessment of work adjustment, is best considered in the broader context of withdrawal behavior—a concept that includes absenteeism, tardiness, and early retirement as well as turnover.

#### 4. WORK ADJUSTMENT COUNSELING

The processes used to help individuals achieve better work adjustment derive from the core constructs discussed previously. These constructs are discussed explicitly by Lofquist and Dawis, who address additional areas such as relationship counseling in which the concepts of correspondence and adjustment are also important. In work adjustment counseling, placement in a job that offers a good fit from both the individual and the employer perspective is the primary aim. However, it is seldom possible to achieve adjustment and fit purely on the basis of selection and placement. Counselors can work with individuals and their employers to try to encourage realistic expectations of each other and to help ensure that the required training and ergonomic or job redesign interventions are undertaken to facilitate better fit. All career counseling includes elements of work adjustment counseling, but the term tends to be reserved for use in the rehabilitation of employees following injury or other trauma.

#### See Also the Following Articles

Career Counseling ■ Coping ■ Holland's Theory (Vocational Personality Types) ■ Person–Environment Fit ■ Structure of Interests ■ Vocational Interests

#### Further Reading

- Allworth, E. A., & Hesketh, B. (1999). Construct-oriented biodata: Capturing change related and contextually relevant future performance. *International Journal of Selection and Assessment*, 7, 97–111.
- Bretz, R. D., & Judge, T. A. (1994). Person–organization fit and the theory of work adjustment: Implications for satisfaction, tenure, and career success. *Journal of Vocational Behavior*, 44, 32–54.
- Dawis, R. V. (1994). The theory of work adjustment as convergent theory. In M. L. Savickas, & R. W. Lent (Eds.), *Convergence in career development theories: Implications for science and practice*. Palo Alto, CA: CPP Books.
- Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis, MN: University of Minnesota Press.
- De Fruyt, F. (2002). A person-centred approach to P–E fit questions: Using multiple-trait model. *Journal of Vocational Behavior*, 60, 73–90.
- George, J. M., & Jones, G. R. (1996). The experience of work and turnover intentions: Interactive effects of value attainment, job satisfaction, and positive mood. *Journal of Applied Psychology*, 81, 318–325.
- Gustafson, S. B., & Mumford, M. D. (1995). Personal style and person–environment fit: A pattern approach. *Journal of Vocational Behavior*, 46, 163–188.
- Hanisch, K. A., & Hulin, C. L. (1991). General attitudes and organizational withdrawal: An evaluation of a causal model. *Journal of Vocational Behavior*, 39, 110–128.
- Hesketh, B., & Gardiner, D. (1993). Person–environment fit: A reconceptualization and empirical test. *Journal of Vocational Behavior*, 42, 315–332.
- Holland, J. H. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Englewood Cliffs, NJ: Prentice Hall.
- Lofquist, L. H., & Dawis, R. V. (1991). *Essentials of person–environment correspondence counselling*. Minneapolis, MN: University of Minnesota Press.
- Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.
- Tracey, T. J. G. (2002). Personal Globe Inventory: Measurement of the Spherical Model of Interests and Competence Beliefs. *Journal of Vocational Behavior*, 60, 113–172.
- Warr, P. (1987). *Work, unemployment, and work adjustment counselling mental health*. Oxford, UK: Oxford University Press.



# Work and Family, Relationship between

Jeffrey H. Greenhaus

Drexel University, Philadelphia, Pennsylvania, USA

Romila Singh

University of Wisconsin, Milwaukee, Wisconsin, USA

1. Introduction
  2. Definitions of Work and Family
  3. Mechanisms Linking Work and Family
  4. The Effect of Work on Family Life
  5. The Effect of Family on Work Life
  6. Gender Differences in Work–Family Relationships
  7. Organizational Actions to Help Individuals Manage Work–Family Relationships
  8. Individual Actions to Manage Work–Family Relationships
  9. Conclusion
- Further Reading

## GLOSSARY

*family* An emotional unit based on love and affection that provides psychological security and nurturance to its members.

*work* An instrumental activity that is intended to provide goods and services to support life or produce something of value for others.

*work–family accommodation* The process by which individuals reduce their involvement in one role to accommodate the demands of the other role. It also refers to a strategy that can be used in response to actual or anticipated work–family conflict.

*work–family balance* The extent to which individuals are equally involved in—and equally satisfied with—their work role and their family role.

*work–family compensation* Efforts to offset dissatisfaction in one role by seeking satisfaction in another role. Compensation may be either reactive or supplemental.

*work–family conflict* Simultaneous pressures from both work and family roles that are mutually incompatible in some respect such that meeting the demands of one role makes it difficult to meet the demands of the other role.

*work–family enrichment* The process by which one role strengthens or enriches the quality of the other role.

*work–family segmentation* The intentional separation of work and family roles such that the thoughts, feelings, and behaviors of one role are actively suppressed from affecting the individual's performance in the other role.

*work–family spillover* The transfer or application of skills, values, emotions, and behavior from one role to the other role. Spillover may be either positive (helpful) or negative (harmful).

Work and family lives are interconnected in many ways. Work experiences can influence family life, and family experiences can affect work life. Some of these effects are negative (work–family conflict), and other effects are positive (work–family enrichment). In this article, different mechanisms by which work and family lives influence each other, gender differences in work–family relationships, and organizational and individual actions to reduce work–family conflict and to promote work–family enrichment are discussed.

## 1. INTRODUCTION

Scholars and journalists write extensively about the difficulties of juggling work and family commitments in contemporary society. Individuals clamor for information about how to cope with extensive work and family responsibilities, and employers increasingly offer work–family programs to help their employees relieve work–family stress and, at the same time, meet their obligations in the workplace. Previously thought to be separate worlds that do not impinge on one another, work and family lives are currently recognized as being inextricably intertwined, sometimes for the better and sometimes for the worse.

What lies behind this interest in work–family relationships? First, the increasing representation of women in the workforce—especially mothers of young children—has rendered the dual-earner family a dominant family structure in the United States. With the blurring of traditionally rigid gender roles—women as full-time homemakers and men as sole providers—both partners in a dual-earner relationship are confronted with the daily challenge of handling their work and family responsibilities in a way that meets the needs of the family and the employer alike.

Moreover, a divorce rate that has stabilized at approximately 50% in the United States has produced increasing numbers of single parents in the workforce. These single parents, most of them women, often struggle to meet extensive work and family responsibilities, often without a great deal of support from their families or their employers.

In addition to, or perhaps because of, these structural changes in family life and workforce participation, employees' values have changed in recent years. Women and men are seeking balance in their lives in deciding what jobs to accept and what jobs to leave. Job applicants are increasingly raising the issue of work–family balance in their employment interviews, and a sizeable number of employees seem willing to make sacrifices in their careers to achieve a high quality of life. Some individuals have chosen to leave their profession when it continually interferes with their family lives.

All these forces have combined to produce an intense interest in the relationship between work and family lives. Researchers in psychology, sociology, and management have risen to the challenge. The past three decades have witnessed an explosion of research on the intersection of work and family roles. Understanding how work can affect family—and how family can affect work—has

been the aim of most of this research. Gaining insight into the relationships between work and family lives is important for a number of reasons.

First, the more that is learned about the intricate relationships between work and family commitments, the more likely employers can provide relevant programs to help their workers manage their work and family lives. Second, understanding the interplay between work and family roles can help women and men learn how to obtain employment that is compatible with their desired lifestyle, how to cope with the stresses of daily life, and how to gain the support of family, friends, and coworkers in relieving work–family stress.

In this article, a wide range of issues regarding the ways in which work and family lives affect one another are discussed. First, the two key roles—work and family—are defined. Then, a variety of mechanisms that have been proposed to explain the relationship between work and family roles are discussed. More detailed discussions of the effects of work on family life and the effects of family on work life are then presented. Moreover, because gender is so central to the interplay between work and family roles, gender issues in work–family relationships are discussed. Finally, some steps that employers and individuals can take to manage the work–family interface effectively are presented.

## 2. DEFINITIONS OF WORK AND FAMILY

In order to explain the interdependence between work and family roles, it is first important to define the terms work and family. Work has most commonly been defined as an instrumental activity that is intended to provide goods and services to support life or produce something of value for others. Most of the research in this area has focused on paid work or employment.

Researchers have defined family most often in terms of the presence of a spouse, children, or kin in the household. A more inclusive definition expands the notion of family beyond marital, adoptive, or biological ties to include people related by affection, obligation, dependence, or cooperation. Thus, a family is considered an emotional unit based on love and affection and one that provides psychological security and nurturance to its members. Moreover, the United States has witnessed an increasing multiplicity of family structures, such as two-parent families, one-parent families, cohabiting couples, gay and lesbian families, and

extended family households. Nevertheless, much of the research on the work–family interface has been conducted on samples of two-parent families.

### 3. MECHANISMS LINKING WORK AND FAMILY

Researchers have identified several mechanisms that help explain the relationship between work and family roles. Early research focused on the ways in which work and family lives conflicted or interfered with one another. Over the years, other explanations emerged that explained a more diverse array of linkages between work and family lives. The most prominent mechanisms that are believed to explain the dynamics of the work–family interface are conflict or interference, accommodation, enrichment, spillover, balance, compensation, and segmentation.

An examination of the work–family literature reveals the consistent popularity of the work–family conflict perspective as an explanation of the manner in which work and family roles influence one another. Work–family conflict refers to simultaneous pressures from both work and family roles that are mutually incompatible in some respect such that meeting the demands of one role makes it difficult to meet the demands of the other role. Work–family conflict has been further distinguished into three specific types of conflict: time-based conflict, strain-based conflict, and behavior-based conflict. Time-based conflict occurs when devoting substantial time to one role makes it difficult to meet the demands associated with the other role. Strain-based conflict occurs when stress arising in one role is carried into the other role and affects individuals' ability to meet the demands of that role. Behavior-based conflict occurs when behavior that is developed in one role is incompatible with the expectations regarding behavior in another role.

Because work–family conflict represents interference between the two roles, researchers have also examined the direction of this interference. Specifically, work-to-family conflict occurs when work demands interfere with the quality of family life. On the other hand, family-to-work conflict occurs when family demands and pressures interfere with the fulfillment of work responsibilities. The consequences of work–family conflict on individuals' work and family lives are examined later.

Another mechanism that has been used to explain the relationship between work and family roles is that of work–family accommodation. Work–family

accommodation is a strategy that can be used in response to actual or anticipated work–family conflict. Accommodation refers to the process by which individuals reduce their involvement in one role to accommodate the demands of the other role. Researchers believe that the reduction in involvement usually occurs in the role that is less important to the individuals and can take two forms. Reduction in behavioral involvement in a role occurs when individuals curtail the amount of time devoted to a role, such as cutting back on the time devoted to work in order to meet their family responsibilities more effectively. Reduction in psychological involvement occurs when individuals restrict their level of ego attachment to a particular role or reduce the importance attached to that role, such as when individuals display less involvement in their family lives in order to fulfill their job responsibilities more effectively.

Whereas work–family conflict represents the negative effects of one role on the other role, work–family enrichment refers to the process by which one role strengthens or enriches the quality of the other role. Work and family roles each provide a variety of resources—tangible and intangible—that can enhance the quality of the other role. Subsequent sections discuss how specific resources from work and family roles have the capacity to provide positive experiences in the other role.

Work–family spillover most often refers to the transfer or application of skills, values, emotions, and behavior from one role to the other role. Work–family spillover can originate in either of the two roles (i.e., spillover can be either from work to family or from family to work). Furthermore, researchers have distinguished two forms of spillover—positive spillover and negative spillover. With positive spillover, the transfer or application of skills, values, emotions, and behavior from one role promotes an individual's performance or satisfaction in the other role, thereby producing work–family enrichment. Negative spillover reduces individuals' performance or satisfaction in the other role, producing work–family conflict. For example, individuals' satisfaction at work is believed to enhance their family life (positive spillover) and, conversely, job stress is likely to create stress in their family interactions (negative spillover).

Presumably, individuals who lead a balanced life can reduce work–family conflict and increase work–family enrichment. Although there is no shortage of advice on how to achieve work–family balance, relatively little research has been conducted on balance or its consequences. We view work–family balance as the extent to which individuals are equally involved in—and equally satisfied with—their work roles and family roles. It has

been suggested that individuals who are balanced in their orientation gain a great deal from work and family activities, experience less stress in carrying out their roles, and derive high self-esteem from the competence they achieve in their work and family pursuits.

Work–family compensation represents efforts to offset dissatisfaction in one role by seeking satisfaction in another role and as such suggests that work and family experiences tend to be antithetical. Researchers have distinguished two forms of compensation. First, individuals may choose to decrease involvement in a dissatisfying role and increase involvement in a more satisfying role. Alternately, individuals may respond to dissatisfaction in one role by pursuing rewarding or fulfilling experiences in the other role. The latter form of compensation has been further distinguished into supplemental and reactive compensation. Supplemental compensation occurs when individuals shift their pursuits for rewarding experiences from the dissatisfying role to a potentially more satisfying one. For example, individuals with little autonomy at work may seek autonomy outside of the work role. Reactive compensation occurs when individuals seek to redress negative experiences in one role by pursuing contrasting experiences in the other role. Pursuing leisure activities after work or resting at home due to fatiguing work are examples of reactive compensation.

Although the mechanisms discussed in this section describe how work and family can affect one another, there is another perspective that focuses on the independence of the two roles. Work–family segmentation originally referred to the notion that work and family spheres exist side by side such that individuals can participate in one sphere without any influence on the other sphere. This notion of separateness (or independence) between work and family lives in time, space, and function implies the compartmentalization of one's life into distinct and separate worlds. Recently, work–family segmentation has referred to the intentional separation of work and family roles such that the thoughts, feelings, and behaviors of one role are actively suppressed from affecting the individual's performance in the other role. Segmentation is now believed to be a process of coping with the stress from either the work role or the family role. As such, it is viewed as a deliberate strategy to inhibit the occurrence of interference between domains and also as a way to maintain a boundary between one's work and family roles.

The previous brief description of mechanisms that explain the nature of the relationship between work and family roles provides the foundation for discussing

in more detail the different ways in which work and family lives affect one another.

#### **4. THE EFFECT OF WORK ON FAMILY LIFE**

Work experiences can affect family life in many different ways. Some of these effects are negative, in which work may be considered an “enemy” of the family. In other respects, however, work can be an “ally” of the family, enriching the quality of family experiences.

##### **4.1. Negative Effects of Work on Family Life**

Much of the early research on the work–family interface was based on a scarcity hypothesis that assumes that time and energy are fixed commodities, and that people who participate in a multitude of social roles (such as work and family) inevitably experience conflict between the roles. Many individuals' job responsibilities are so time-consuming or so inflexible that they interfere with a full and satisfying family life. Moreover, as a result of this time-based conflict, individuals can become dissatisfied with their marriage and their family and may experience substantial depression and dissatisfaction with their lives.

Although the shortage of time to devote to both work and family activities is part of the problem, another issue, perhaps as significant as time, is the psychological involvement that some employees invest in their work and career. There is a segment of the population for whom work is the central part of their lives. These individuals identify so strongly with their work and are so absorbed in their work that they have little left to give to their family. They may be so preoccupied with work that even when they are physically present at home, their thoughts lie at their workplace. An imbalanced, overinvolvement in work, sometimes referred to as workaholism, may have negative effects on individuals' relationships with their family and can detract from the quality of these individuals' lives.

Beyond the time devoted to work and the psychological absorption in work that can interfere with family, job stress can have negative consequences for employees' family life. Work-related stress can have many underlying causes—a boring job, too much responsibility, interpersonal conflict with a colleague, undue pressure for productivity, or the threat of job loss. Regardless of its

source, the tension, irritability, fatigue, or anxiety produced by job stressors may be “taken home” and interfere with one’s performance or enjoyment within the family domain. This negative emotional spillover can render individuals in stressful jobs vulnerable to problems at home. Such individuals have trouble “turning off” the stress when they cross the boundary from work to family.

One important by-product of job stress that plays out in the family domain is the crossover of stress that may occur from one spouse or partner to the other. When individuals’ job-related stress spills over into the home, the stress may be transmitted to their partners. For example, individuals who feel burned out at work may become anxious, tense, and irritable at home. These individuals’ partners may begin to perceive the individuals’ tension and, out of empathy toward the partners, become stressed themselves. Burned-out employees may, out of frustration, behave harshly toward their partners, which triggers stressful reactions in the partners. It is reasonable to expect that the increased stress experienced by the partner will arouse even more stress in the already anxious employee, thereby creating a vicious cycle.

Stressful work may also have negative effects on children. Individuals who bring home extensive stress from their work may withdraw from interactions with their children, act coldly toward them, and experience conflict dealing with their adolescents. One source of parental stress—job insecurity—has been found to be associated with children’s negative work attitudes, low self-confidence, cognitive difficulties, and poor grades in school.

In addition to the family problems produced by time, overinvolvement, or stress within the work role, there is another way in which work lives can interfere with family. Behavior that is learned or reinforced at work may be inappropriately applied in the family domain, thereby producing behavior-based conflict. For example, individuals may believe that the pursuit of career success requires them to be detached and distant from subordinates so they can objectively appraise their subordinates’ performance and make difficult decisions about their future careers. They may also believe that being highly directive toward their subordinates—telling them exactly how to perform their jobs—is an effective leadership style. However, individuals who apply the same detached, distant, and directive style at home may find that their spouse and children are understandably displeased. Therefore, the inability to “switch gears” when crossing the border from work to home can produce tension and anger in family members that detract from the quality of family life.

In summary, the negative effects of work on family life reflect one form or another of work-to-family

conflict due to extensive and inflexible work hours, overinvolvement in work, job stress that carries over into the home environment, or work-related behaviors that are inappropriately applied within the family domain. Such conflict can produce distress within the family domain, withdrawal from family responsibilities, and negative parenting, and it can adversely affect one’s quality of life.

#### **4.2. Positive Effects of Work on Family Life**

Although work experiences can interfere with home and family responsibilities in a variety of ways, they can also enrich the quality of family life. Therefore, in the right circumstances, work can be an ally of the family. Employment potentially provides a number of resources that can be applied fruitfully to the family domain.

One of the most powerful resources provided by a job is income. A high income can enhance a family’s standard of living, provide a stimulating and healthful home environment for children, and enable parents to spend more time with children through the purchase of outside help to perform some household chores. As observed frequently in everyday life, money does not guarantee happiness, but it can serve as a resource to enhance the quality of a family’s life.

However, resources derived from work go beyond money to include more intangible assets that are associated with employment. For example, individuals who are satisfied with their job tend also to be satisfied with their family life. Just as extensive job stress can spill over to the family in a negative way, good feelings about work can be transferred to the home environment through positive spillover, so an individual can be more relaxed and participate more fully in family life.

One especially important resource at work is the amount of autonomy or freedom provided on the job. Employees who experience substantial job autonomy provide a stimulating home environment for their children, tend to have more positive interactions with their children, and report that their children have relatively few health and behavior problems. Moreover, individuals with considerable job autonomy experience low levels of work–family conflict, probably because they have control over the timing and location of their work. Similarly, flexible work schedules enable individuals to gain more control over their work responsibilities, reducing their level of work–family conflict and increasing their satisfaction with family life.

Another important resource that can be garnered at work is the acquisition of skills, information, and behavior that can be applied effectively at home—the opposite of behavior-based conflict. For example, employees may learn that listening attentively to coworkers or clients is an effective work style that, when used at home, improves the quality of their relationship with other family members. Managers may realize that letting subordinates learn from their own mistakes is a technique that can be used at home to promote their children's self-esteem and independence. Moreover, employees may find that their self-esteem is bolstered when they perform well on the job. This heightened sense of self-esteem can serve as a resource because it may enable employees to interact more confidently with others in the family domain.

Finally, the social support provided by others in the workplace is a significant resource to many individuals. Employees who receive substantial support from their coworkers or supervisors tend to experience relatively little work–family conflict and enjoy greater well-being in the family domain. Sometimes, this support is based on informal relationships and friendships with other people who can provide understanding, information, and advice on juggling work and family responsibilities. In other cases, support is provided through an employer's institutionalized “family-responsive” practices and a culture that respects an employee's life outside work. Individuals who work for supportive organizations have been found to experience relatively little conflict between their work and family responsibilities and are more satisfied with their personal growth and development.

## **5. THE EFFECT OF FAMILY ON WORK LIFE**

The previous description illustrated the numerous ways in which work experiences affect family life. Because work and family roles are mutually interdependent, family life and experiences can also affect an individual's workplace behavior, attitudes, performance, and career decisions. Just as work can either enhance or detract from the family, there are parallel positive and negative influences of family experiences on work life.

### **5.1. Negative Effects of Family on Work Life**

Much of the early research on the effects of family on work was based on gendered work–family models that

examined the influence of women's family roles and responsibilities on their work choices and performance. With the growing recognition of men's involvement in family life, scholars have begun to offer a more balanced (or gender-neutral) explanation of the effects of family on work.

A large body of the work–family literature has focused on the effects of time spent in family activities on a variety of work attitudes and decisions. The underlying premise is that both work and family roles make myriad demands on individuals' time, and in trying to meet those demands conflict inevitably results. Certain facets of one's family life are likely to make more demands on individuals' time than others. One dimension of family life—marriage—may impose only limited time demands and hence exercise a limited influence on work behaviors and performance. However, an increase in time demands may be introduced in the form of spousal job commitments and attitudes that influence a variety of career decisions, such as work hours, involvement, and withdrawal. For example, individuals whose spouses are unavailable or reluctant to share in the household chores or other family responsibilities may end up shouldering more of these responsibilities with the result that they have little time or energy left to give to work.

One particular facet of family life that has received a great deal of attention in terms of its effect on work is parental responsibilities. Parenting undeniably introduces new time and energy demands on individuals and has been shown to affect certain employment decisions, such as seeking part-time work, self-employment, and employment in family supportive organizations. Once employed, parents with young children are likely to spend relatively fewer hours on the job than non-parents and are likely to make adjustments to their work schedules that may often result in reduced time at work. Furthermore, parents may often turn down career development opportunities that involve travel in order to spend more time with their families.

Another characteristic of caregiving that taps into employees' already stretched time and energy resources is care for the elderly. With the progressive aging of the population, it has been estimated that approximately 20–30% of all workers are likely to be involved in providing some form of assistance to an older person. Employees who have extensive elder care responsibilities may limit their behavioral involvement at work by cutting back on their work hours, rearranging their work schedules, turning down relocation opportunities, and taking unpaid leave. When these responsibilities

become very demanding, it is not uncommon for some employees to withdraw completely from the workforce.

Different caregiving responsibilities demand not just physical and behavioral involvement but also require a great deal of psychological involvement in the family role with the result that individuals have little left to give to their work role. Some employees who are behaviorally and psychologically focused on their families have considerably lower aspirations for senior management positions than those who are either career focused or both career and family focused. Other individuals who are deeply involved with their family lives may choose to scale back their involvement in work, which may eventually affect their potential for getting jobs with greater autonomy or other career-enhancing assignments.

In all these instances, individuals seek to accommodate the increase in caregiving responsibilities and their associated time and energy demands by limiting their participation and investment in the work role. Although some individuals voluntarily scale back their involvement in the work role, it is also possible that some organizations are less willing to invest in employees who are extensively involved in their family lives. A full discussion of organizational responsiveness to work–family issues is presented later.

Extensive obligations to fulfill parental and caregiving roles may make individuals irritable, anxious, or tense and create disruptive family dynamics that lead to family distress. Such negative emotions and distress generated within the family role may spill over into the work domain and reduce individuals' work productivity and work attitudes.

In summary, different facets of individuals' family life, such as spousal behavior and attitudes, extensive caregiving responsibilities, and intensive involvement with family activities, can conflict with their work lives. Such conflict can limit individuals' career choices and involvement and negatively affect their work involvement, job satisfaction, aspirations for senior management positions, and their intentions to continue employment.

## **5.2. Positive Effects of Family on Work Life**

Home and family life can also strengthen, expand, or enrich the quality of one's work life, and in this sense family becomes an ally of work. Individuals living in a family can draw on a variety of supportive resources that directly or indirectly affect their work lives.

Family members—one's spouse, children, and other relatives—are perhaps the major source of support that

can either prevent family stress or buffer the individuals from the pressures of home life and enable them to participate more actively and effectively in work life. This support can take various forms, such as providing direct physical help with various household tasks or child care activities and/or providing emotional support in the form of affection, advice, affirmation, and information that helps individuals cope more effectively with various work-related problems. By acting as a sounding board and offering suggestions and feedback to their spouses, a supportive environment can be created that enables individuals to devote more time to the pursuit of important career goals.

Supportive family members, especially one's spouse, who participate more actively in family life and provide emotional support allow the individuals to spend more time and energy at work, make fewer adjustments to their work schedules, avail themselves of career-enhancing opportunities, and experience greater satisfaction with their careers than those who receive little support. Support from one's spouse not only enables individuals to cope more effectively with their work lives but also can reduce the extent to which their family interferes with work. This, in turn, can free up individuals mentally, emotionally, and physically to channel greater effort and energy into their work.

Working spouses can also provide support in the form of financial resources that can augment their personal efforts and thus create a more satisfying family environment. In theory, the security provided by spousal income can enable individuals to leave a dissatisfying job and pursue a career that is more in line with their values and aspirations.

It is not just spouses who provide significant sources of support to individuals. Children, although demanding on working parents, can also be a source of help to them. It is not uncommon for children, particularly older ones, to help out with household work and look after their younger siblings. For parents who are already stretched for time and energy, such support is valuable in enabling them to increase their time commitment to work.

Supportive family relationships can foster emotional closeness, bring peace of mind, and produce general satisfaction with one's family life. Positive feelings and fulfillment generated in one's home environment can spill over into one's work life, producing satisfying feelings toward one's work and allowing individuals to apply themselves fully at work. Positive spillover can also occur when individuals transfer the skills, knowledge, and perspectives that they acquired from their home experiences to their work lives. For

example, parenting skills such as active listening, expressing concern, and providing feedback may enhance an individual's effectiveness at work.

In summary, a supportive family environment may offset some of the negative effects of family on work and also strengthen, enhance, and enrich an individual's work life in a variety of ways. However, to fully appreciate the intricacies of work–family relationships, one must take into consideration whether these influences are similar for both men and women.

## **6. GENDER DIFFERENCES IN WORK–FAMILY RELATIONSHIPS**

Despite a weakening gender-based segregation of roles, men and women continue to spend their time differently, especially when there are children in the family. Fathers generally devote more time to paid employment than do mothers, and mothers spend more time on household chores and caring for the children than do fathers. Moreover, women generally feel more responsible for the physical and emotional well-being of their children than do men. Given these persistent gender differences in work and family role orientations, it is reasonable to ask whether the interdependence of work and family roles is different for women and men.

Considering first the negative relationships between work and family roles, some studies have found that women experience more work–family conflict than do men. The results of these studies are consistent with the reality that employed women typically spend more total time on work, home, and family activities than do men. The number of hours that women work is rarely accompanied by a commensurate reduction in the time they devote to home and family. For this reason, it has been suggested that women work a “second shift” when they come home after a day's employment. Therefore, when women experience more work-to-family conflict than men, it is likely because they have more family responsibilities with which work can interfere.

There is also a ready explanation for those situations in which women experience more family-to-work conflict than men. Women with children, especially young children, are more likely to reduce their involvement in work in order to accommodate their family's needs than are men. This accommodation may take the form of cutting back on work hours, restricting the amount of weekend or evening work, or curtailing extensive travel. Any of these forms of accommodation may cause a

woman to perceive that her family responsibilities have limited or interfered with her work-related pursuits.

However, many other studies have found that women and men experience a similar level of work–family conflict. Again, there are several explanations for this finding. First, it is possible that gender differences in work and family responsibilities have eroded to the point where women and men basically face the same degree of conflict in juggling their life roles. More plausibly, however, women may be more likely than men to choose jobs that are less demanding and therefore less likely to produce extensive work–family conflict. Historically, women have been more likely than men to base their selection of a particular occupation on its compatibility with extensive family responsibilities. Moreover, as noted previously, many women with young children have already reduced their work involvement to alleviate the daily conflicts between family and work. It is also possible that women are more effective than men at coping with work and family pressures, thereby keeping work–family conflict at a manageable level. All of these factors may explain why many studies have observed no gender difference in work–family conflict.

There has not been a great deal of research on whether positive relationships between work and family roles are different for men and women. It seems that both women and men can benefit from the support provided by their spouses, supervisors, or coworkers. However, some research suggests that the positive effect of work experiences on family life may be somewhat different for men and women. For example, women may be more likely than men to use their work experiences (especially autonomy and social relationships) to benefit their families. In addition, women respond more positively than men to the presence of a flexible work schedule and place more importance than men on child care support and alternative work hours. Despite these differences, it is clear that women and men are capable of using resources from a particular role to improve life in the other role.

## **7. ORGANIZATIONAL ACTIONS TO HELP INDIVIDUALS MANAGE WORK–FAMILY RELATIONSHIPS**

As indicated previously, employees can experience conflict between their work and family lives. Although women continue to face extensive pressures in juggling their home and work demands, it is evident that men are increasingly involved in these same struggles. As these

challenges intensify in the future, employees will continue to demand opportunities that help them attain a balance between their work and family roles and lead more fulfilling lives. In response to changing demographic trends and employee preferences, organizations are offering a variety of programs that help employees address their work–family concerns. These programs can be broadly classified to include practices that focus on dependent care issues and others that focus on flexible work arrangements.

### 7.1. Dependent Care Arrangements

As previously discussed, a large number of working parents often scale back their time at work, make significant adjustments to their work schedules, miss work, and curtail their involvement at work in order to fulfill some of their child care responsibilities. Organizations have responded to these concerns by providing a wide range of child care programs that better assist their employees to meet their parenting needs with the belief that this will permit their employees to channel their time and energies more fully to work. Not only can employees benefit from such practices but also organizations can profit from these actions by retaining a motivated and productive workforce.

Many large companies provide on-site or near-site child care centers for their employees' children. Employees who avail themselves of these facilities feel more satisfied with their jobs, express greater loyalty toward their companies, and miss fewer days of work. People who are seeking employment are attracted to organizations that offer some type of child care assistance. Organizations, in turn, reap benefits in the form of improved employee productivity and morale. Another way that organizations gain from having these facilities is by being able to form a progressive public image that they can use to attract and recruit better talent.

Some companies are not able to provide on-site child care centers and instead offer alternative arrangements to meet their employees' needs. Such arrangements include, but are not limited to, contracting with individual child care providers and getting discounts for their employees, providing cash vouchers that employees can use at the child care centers of their choice, and offering referral services that help parents with information and advice regarding different child care facilities. Some employers get involved with the communities in which they do business and initiate and/or expand a variety of child care services, such as after-school programs or "sick child care centers." A strategy adopted by

many companies involves a revision of their sick leave policies so employees can take sick leave when their family members are ill.

These are some of the more popular and widely used programs that provide direct assistance to individuals with child care needs. However, as previously mentioned, a rapidly growing number of employees are also responsible for the care of their elderly relatives. Employers are increasingly responding to their employees' demands for elder care assistance by providing them with referral services, offering them support groups through their employee assistance programs, and providing social workers who visit their employees' elderly relatives.

Despite the wide range of dependent care programs offered by organizations, a key issue for employees with any kind of dependent care responsibilities involves the freedom to have some degree of control over their time. Organizations offer a variety of flexible work options that allow employees to choose a work schedule that best fits their needs. Some of the widely offered flexible work arrangements are discussed next.

### 7.2. Flexible Work Arrangements

Flexible work arrangements are designed to help individuals balance their family and work responsibilities by giving them the option of choosing when to work, how much to work, and where to work. One of the earliest and most enduring types of flexible work arrangements—flextime—involves working a core set of hours with flexible starting and leaving times. By giving people more control over their work times, employees are better able to meet their various family obligations. Employees who use flextime report being able to effectively perform their jobs, feeling satisfied with their work, having better relations with their supervisors and team members, and expressing relatively little intention to leave the company. Companies may not only gain in terms of improved employee productivity but also can incur savings through reduced use of sick days, absenteeism, and lateness.

Another option that is increasingly being offered by many companies is that of part-time employment or a reduced-load work arrangement. Companies can allow part-timers to jointly take responsibility for a full-time assignment so that the total job is "shared" by two or more part-timers, an option known as job sharing. Part-time employment was initially proposed as an option for women managers and professionals with young children who often chose to quit their jobs rather than give up time with their children and

continue to work full-time. Often referred to as the “mommy track,” the option of part-time employment has been surrounded by controversy with regard to its effectiveness. Some people believe that individuals who opt for part-time work—whether mothers or fathers—perform just as effectively in reduced-load work arrangements as they did when they worked full-time. Moreover, they also believe that employees who participate in such work arrangements feel happier and more satisfied with their work–home balance than they did when they worked full-time.

However, there are others who believe that part-time employment, no matter how brief, can signal to the organization that the employees may not be committed enough to their jobs and organizations. Such stigmatization is more common for women than for men, perhaps because women are much more likely than men to avail themselves of this option or because men are more likely to negotiate informal reduced work involvement with their bosses rather than openly declare their desire to spend more time with their families.

To ensure the success of part-time arrangements, companies have to overcome their stereotypes of people who use these programs and treat them equitably. More fundamentally, companies have to rethink their models of career success. They can offer alternatives to fast-track linear career paths and assignments that can have negative effects on employees’ personal and family lives.

Another form of workplace flexibility, telecommuting, focuses not so much on reducing the amount of time employees spend at work but rather on offering employees the option of working from home or from a satellite office for a given portion of the workweek. This allows employees greater discretion in structuring their work and family time. However, working from home may not necessarily solve the employees’ child care needs or eliminate work–family conflict. Instead, the blurring of the boundaries between their work and family roles can make individuals increasingly more vulnerable to continuing stressors from both roles. They may also find themselves increasingly isolated from their colleagues and the social interchanges at work that can provide a welcome reprieve from the daily pressures. Nevertheless, telecommuting has much to offer employees who wish to balance their work and family demands.

Another widely used initiative to help employees meet their family obligations is that of family leave. Since the passage of the Family and Medical Leave Act in 1993, a number of companies have instituted various programs that allow their employees to take leave to care for their children, spouse, or parents. Many of these leave policies

offer job protection to employees and provide them with medical benefits for a certain period during the leave. By directly addressing employees’ needs for caregiving time, family leave programs can improve recruitment and retention of talented employees.

Organizations can offer a patchwork of solutions and programs to enable their employees to fulfill their work and family obligations. However, to be truly effective in the long term, companies have to undergo a fundamental shift and alter their cultures in a way that recognizes and legitimizes work–family issues. In addition, the success of most of these programs depends in large part on the support and commitment of the top management to work–family issues.

In summary, many of these programs may not be able to eliminate work–family interference, but they can provide employees with sufficient flexibility to achieve a more fulfilling balance between their work and family lives.

## **8. INDIVIDUAL ACTIONS TO MANAGE WORK–FAMILY RELATIONSHIPS**

Individuals with extensive work and family responsibilities need to find ways to manage their work–family interface effectively. This involves not only reducing their level of work–family conflict to a reasonable level but also increasing the likelihood that their work and family roles will enrich one another. Given the advantages of working in a family responsive organization discussed previously, individuals seeking employment should assess this critical aspect of a prospective employer’s culture and use this information in deciding whether to accept a job offer.

Three proposed principles for creating allies of work and family are worth serious consideration. First, individuals should clarify what is important in their lives. Individuals vary in the relative importance they attach to different parts of their self-identity (e.g., family, work, leisure, and community service), and some people are more aware of their underlying values than others. Self-awareness is crucial for managing the relationship between work and family roles because it helps individuals make effective decisions about the allocation of time and emotion to different roles and the application of resources from one role to enhance the quality of life in the other.

Second, individuals should recognize and support the whole person. This requires a realization of how

participation in different life roles can enrich one's life. It also involves building supportive relationships with individuals in different roles, such as work, family, and community. The support can be tangible assistance (e.g., help around the home) or it can be intangible (e.g., listening to problems and offering advice on solving them). The stress literature indicates that social support from others can enhance well-being by reducing the amount of stress an individual experiences and by helping the individual cope with the stress that is experienced. Seeking support from others requires a willingness to ask for—and accept—others' help. Because of norms of reciprocity and fairness governing human behavior, it also requires a readiness to provide support to others in need of assistance.

Third, individuals should continually experiment with how goals are achieved. This experimentation requires periodically reexamining goals, developing new strategies to accomplish these goals, and adopting new perspectives on one's work and personal life. It involves an openness to change—a willingness to question what is important in life and what approaches are necessary to achieve desired outcomes. Like the problem-focused coping technique lauded in the stress literature, continual experimentation often means extensive communication and problem solving with other people (e.g., spouse and boss) to reduce the conflicts in one's life and to strengthen the quality of one's significant life roles.

## 9. CONCLUSION

Work and family lives are intertwined in many respects. Experiences at work can influence the quality of family life, and family experiences can affect the quality of work life. Some of these consequences are negative, such as when one role interferes with full and enjoyable participation in the other role. In other cases, however, resources from one role can be used to strengthen the quality of life in the other role. Organizations and individuals can take actions to relieve the stress arising from the intersection of employees' work and family roles and to strengthen the positive effect of each role on the other role.

## See Also the Following Articles

Elder Caregiving ■ Family and Culture ■ Leisure and Work, Relationship between ■ Part-Time Work

## Further Reading

- Barnett, R. C. (1998). Toward a review and reconceptualization of the work/family literature. *Genetic, Social, and General Psychology Monographs*, 124(2), 125–182.
- Barnett, R. C., & Hyde, J. S. (2001). Women, men, work, and family: An expansionist view. *American Psychologist*, 56, 781–796.
- Edwards, J. R., & Rothbard, N. P. (2000). Mechanisms linking work and family: Clarifying the relationship between work and family constructs. *Academy of Management Review*, 25, 178–199.
- Friedman, S. D., & Greenhaus, J. H. (2000). *Work and family—Allies or enemies? What happens when business professionals confront life choices*. New York: Oxford University Press.
- Greenhaus, J. H., & Beutell, N. J. (1985). Sources of conflict between work and family roles. *Academy of Management Review*, 10, 76–88.
- Greenhaus, J. H., & Parasuraman, S. (1999). Research on work, family, and gender: Current status and future directions. In G. N. Powell (Ed.), *Handbook of gender and work* (pp. 391–412). Thousand Oaks, CA: Sage.
- Haas, L. (1999). Families and work. In M. Sussman, S. K. Steinmetz, & G. W. Peterson (Eds.), *Handbook of marriage and the family* (2nd ed., pp. 571–612). New York: Plenum Press.
- Hall, D. T. (1990). Promoting work/family balance: An organization-change approach. *Organizational Dynamics*, 18(3), 5–18.
- Hochschild, A. (1989). *The second shift: Working parents and the revolution in the home*. New York: Viking.
- Lambert, S. J. (1990). Processes linking work and family: A critical review and research agenda. *Human Relations*, 43, 239–257.
- Lobel, S. A. (1991). Allocation of investment in work and family roles: Alternative theories and implications for research. *Academy of Management Review*, 16, 507–521.
- Marks, S. R., & MacDermid, S. M. (1996). Multiple roles and the self: A theory of role balance. *Journal of Marriage and the Family*, 58, 417–432.



# Work Environments

Rotraut Walden

University of Koblenz, Koblenz, Germany

1. Introduction
  2. Using Buildings for Increasing Productivity
  3. A Historical Perspective on Working Environments
  4. Some Recent Studies on Office Environments
  5. General Models Describing Environment–Behavior Relations in the Workplace
  6. Future Developments
- Further Reading

## GLOSSARY

**Hawthorne effect** Effect where if workers notice that they are being observed and that conditions in the workplace change because of them, efficiency increases independently of actual improvements.

**personalization** The opportunity to make personal changes in the environment through decorations (e.g., photographs, artwork) and in the form of functionality.

**postoccupancy evaluation (POE)** Evaluation that investigates to what extent a constructed environment provides for current users' needs and demands made by experts.

**productivity** A synonym for output, efficiency, motivation, individual performance, organizational effectiveness, production, profitability, cost-effectiveness, competitiveness, and/or quality of work of a company or an individual.

**sick building syndrome** Syndrome where the users of a building become ill as a result of faulty construction and/or through the use of hazardous materials.

**user-needs analysis** An assessment of the needs, desires, and preferences of future users of a building; it is usually employed during the early phase of the construction

program so that potential recommendations can still be incorporated for the improvement of the future building.

**Yerkes–Dodson Law of 1908** States that the link between arousal and performance results in an inverse U-shaped curve; performance is better during medium arousal than during low or high arousal.

Performance and well-being are the most common criteria for the improvement of work environments. A glance at history shows the focal points of psychological research over the ages, including advancements in evaluatory methods for improving the working environment. This article presents and discusses some classic studies in this field. Of current interest are building performance evaluation and facility management. The article presents psychological frameworks and models summarizing the current knowledge about environment–behavior relationships in workplaces, envisaging possible amendments to them. Because of the growing importance of information technology, customer service, and globalization, the article concludes by dealing with the foreseeable development of offices and the consequences of internationalization.

## 1. INTRODUCTION

The aim of environmental psychology of workplaces and organizational psychology is to improve workers' performance and well-being. Performance refers to individual,

team, or organizational efficiency. Well-being complements the question of satisfaction and is a more holistic approach of recent research that includes health. Satisfaction is generated when, among other things, a worker or an employee can achieve goals and satisfy needs in the workplace. It is an individual's general affective and cognitive evaluation of the job and the physical environment that can support or contrast these goals and needs. The effects of increased performance at reduced costs are of particular interest to companies. It is also important for a business to have creative and innovative individuals who work to meet the constant demands made on the company and who are willing to commit themselves to the company. For users, there is a pivotal interest in the continuity of the organization so as to ensure the stability of their own jobs. Consequently, they concern themselves with the company's performance and make individual contributions to this success. If one considers the proportion of time spent at work over a lifetime, there are compelling reasons for increasing well-being during this period. In the long run, well-being in the workplace is, like salary, an important incentive for employees.

## **2. USING BUILDINGS FOR INCREASING PRODUCTIVITY**

After a review of many different studies, Gifford concluded in 2002 that the majority of companies still do not take into account the fact that even a minimal investment in the workplace can result in an improvement of 10 to 50% in productivity. Companies see their employees as their biggest potential growth factor, but at the same time, employees create the highest costs. A reduction in construction costs usually results in deteriorating working conditions. The strategy for how an investment in new construction or renovation can be implemented is directly connected to the psychological factors of the work environment itself. Therefore, a major question explored in this article is the following: How can performance and well-being in workplace buildings be measured and improved?

## **3. A HISTORICAL PERSPECTIVE ON WORKING ENVIRONMENTS**

Writing about work environments in 2001, Bell and colleagues differentiated among factories, large firms, and offices. This historical retrospective leans closely

toward Becker's 1991 article. During the early 1900s, Frederick Taylor assumed that incentives in terms of salary, production techniques, and the work environment should enhance workers' productivity. In fact, it had more to do with the assembly line in the factory. Industrial growth was explosive. Social processes, well-being, and satisfaction—the points of view of the employees—were hardly considered. Scientific management led to routine, standardized jobs and to studies about heating and lighting.

Avoiding Taylorism by using human factors, and an increase in automation of factories to increase efficiency and individual performance between 1910 and 1940, was followed by an emphasis on task performance and human relations between 1940 and 1950. Many results that had been considered valid until that point were questioned in the light of the famous experiments conducted by Elton Mayo and colleagues between 1927 and 1939 at the Western Electric Company plant in Hawthorne, Illinois (near Chicago and Cicero), on the effects of physical work environment on workers' performance. As Roethlisberger and Dickson reported in 1939, work performance increased in both experimental and control groups, although lighting was improved only for the former. In follow-up observations, researchers established that performance improved again despite the fact that illumination had been reduced by 70%. When they replaced a set of lights with bulbs of the same wattage, the employees were even more satisfied. These results gave rise to the following interpretation: If employees notice that the conditions of their immediate environment change because they are being observed, performance increases independently of any actual improvements. This reaction was named the Hawthorne effect. As a result, the naive determinism that dictated that simply improving environmental conditions has a direct effect on desired behavior was questioned. Instead, the role played by the employees' own perceptions, beliefs, preferences, experiences, and personalities was considered more carefully.

Studies on group dynamics, communication, and conflict then followed between 1950 and 1960. These studies were motivated by the need for a decrease in interpersonal conflicts and the development of democratic and cooperative group processes at work. This was often provoked by the preoccupation with World War II. In 1967, Robert Sommer developed a "small group ecology." Arguments for building planning and equipping furnishings, as well as for seating arrangements, formed the origins of environmental psychology. Architects and planners were viewed as managers of human resources.

Subsequently, between 1960 and 1970, a dominant theme was the focus on nonpaying users. Participation of users in the planning of work environments was also encouraged. This should result in a building layout that better mirrors the needs, values, and preferences of its users. Communication, comfort at work, and satisfaction emerged as central concerns in numerous studies carried out between 1970 and 1980.

Studies on office landscape revealed that employees rejected office layouts designed for the ease of communication within the company. There were often complaints about noise and lack of privacy. Workplace research increasingly busied itself with office space, and this corresponded with the increasing value of customer service. For example, in 1984, Brill and colleagues reported the results of an extensive 5-year investigation of 70 firms conducted from the Buffalo Organization for Social and Technological Innovation (BOSTI). The study addressed three basic questions. First, does the overall design of the office have a fundamental influence on employees' productivity and their daily quality of life at work? Second, which specific office characteristics breed these sorts of influences, and in what ways does this happen? Third, how great is the value of these influences in financial terms? Between 1977 and 1983, 6000 people were interviewed over two time periods: before and approximately 8 to 12 months after a generally positive remodeling of their workplace. A wide set of "objective" and "subjective" criteria were used for assessing productivity and office characteristics. After the first investigation, BOSTI recommended targeted changes whose success would be measured after 5 years in terms of monetary value. The results of the BOSTI study suggested that the changes that the firms aimed to make contributed to an increase in efficiency of 15 to 17% in the annual salary of three career groups over the 5-year period: managers, professional or technical workers, and clerical workers. In doing this, investments and the costs of the workplace were calculated.

With regard to work satisfaction, relocation frequency and noise had the highest monetary value. Approximately 25 to 30% of the changes favoring job satisfaction by way of targeted investments contributed to an increase in productivity. Aspects of job satisfaction were more important to clerical workers than to managers and professional/technical workers. Personal storage space for office materials and coats was deemed to be very important by clerical workers. The remaining 70 to 75% of the monetary value was determined by enclosure and layout in job performance.

During the 1980s, Preiser and colleagues introduced the concepts of postoccupancy evaluation (POE) and performance profiles as methods of evaluation for these studies. Today, there are several methods of data collection, such as program development, user needs analysis (UNA), postconstruction evaluation, and POE, that adhere to the comments made by users and experts alike. The POE investigates to what extent a built environment is set into use in accordance with users' needs and the demands made by experts. Therefore, feedback for planners and building owners on the effects of the built environment should be identified, as should the prerequisites for the construction of a better environment. The POE assesses environments according to what extent they encourage their designated function, or at least do not disturb it, and to what extent these environments correspond to the needs, interests, and desires of the users. In comparison with the assessments made by competing architectural firms, the users' view plays a special role.

A broad range of techniques may be used for collecting data, including interviews, questionnaires, checklists, collections of critical events, targeted time-budget studies in the form of log-book entries, objective data taken from thermometers and hygrometers, and counts made from the analysis of physical traces, photographs, and video recordings.

Diversity of methods and the consultation of people in central groups with different visions are a means of choice. The results of the POE are short, medium, and long term. The POE contains feedback from users about problems with the building and proposed solutions. Also, lessons about how to improve the design cycle of a building can be implemented, and planning and design criteria for specific types of buildings can be developed. The immediate goals of these evaluations are to improve building performance with regard to health, security, functionality, and psychological satisfaction. Furthermore, they aim to save costs on maintenance with regard to the whole life spans of buildings. Finally, benchmarks, concepts, and guidelines can be developed for future projects.

During the 1990s, various concepts were also introduced, including facility management, building delivery, and life cycle management. These concepts are also related to the principle of improving building performance and adopting different evaluation methods to combine architects' and users' perspectives on how the improved performance of buildings is perceived. According to a 1997 chapter by Preiser and Schramm, all phases and evaluation methods can be integrated into the building performance evaluation.

#### 4. SOME RECENT STUDIES ON OFFICE ENVIRONMENTS

Currently, more than 50% of people in the United States work in offices. However, throughout history, offices have changed significantly—from the cubicles of the *Ufficis* in Florence, Italy, around 1559 (from which the term “office” was derived), to offices alongside corridors, to open-plan offices using “landscape offices.” Group offices have once again become smaller, whereas so-called “combi-offices” are a collection of individual offices grouped around a service and communication core area used by all. The nonterritorial office is equipped with workstations that are not dedicated to any one individual employee but rather are used by lots of different people who share space at different times. This makes sense if employees often have long periods of work outside the office, as in the insurance industry. Hotelling is one form of nonterritorial office for those who do not require a dedicated, personally assigned office. Like a hotel reservation, available offices and appropriate equipment can be reserved. This kind of office lacks opportunities for personalization. It hinders familiarity between colleagues because office neighbors are always changing. In 1997, Schneider and Gentz observed that “intelligent offices” have a floor space of more than 5000 square meters and that their users are various businesses with complex and innovative types of work. These offices are equipped with personal computers, modern conference technology, and access to the Internet, all of which are connected to the individual workstations. The central computers regulate the building, regulate in-house technology, and support facility management. This also provides round-the-clock support for employees, thereby increasing productivity. “Intelligence” is related to the building in its urban context, the total energy needs, and the real estate market. As for the users, intelligence controls changes through the duration of use, avoiding the sick building syndrome.

In 1995, Becker and Steele reported on their study of a new building, compatible with new demands made on a company, that was planned in the state of Michigan. Steelcase is the world’s largest manufacturer of contract office furniture. To remain competitive, employees must continually develop new and innovative products. Steelcase’s Corporate Development Center (CDC) achieved success through user participation of 450 employees when planning and designing new workplaces (completed in 1989). Top management decided on the site location, building forms, general design

concepts, and flow of communication. User focus groups governed the design of individual workstations, shared commons areas, and dedicated project rooms. In so doing, work patterns and environmental requirements were incorporated. The total workplace concept of the CDC includes the idea of integrating decisions such as human resources, information technology, design, construction, building operations, and management. It should be developed as a concept for the entire process in buildings. Informal communication should stimulate creativity. Therefore, teamwork and face-to-face interaction should be encouraged. Different settings were created for different activities. Spatial mobility during the course of daily activities was expected to increase the probability that people would talk about their work informally. Movement through a building should be used to encourage personal contact; therefore, construction did not adhere to the idea of making people go the shortest way while at work.

A second study conducted by the BOSTI was reported in 2001 by Brill and colleagues. Whereas the earlier BOSTI study lacked a specific theoretical model, rudiments of it can be found in this second one. At the very least, in this study a large number of aspects were named and their effects were described. It is hoped, particularly because the results are not yet final, that interconnected conclusions can also be examined. One of the study’s authors, Sue Weidemann, observed the following in a personal communication in November 2003: “The BOSTI study reported the final results of the analyses which were based on only the direct predictors of the outcomes. If both direct and indirect predictors were considered, it would be seen that the work environment has an even greater impact than that reported.”

During the 15 years between the two BOSTI studies, some long-term and stable business trends have emerged. These trends are called “new officing.” The goal of new offices is to use workplaces, technologies, and work processes as an integrated system of enablers to work smarter at every work location. New officing means a radical redesign of workplaces, work from anywhere, and hotelling.

The second BOSTI study examined the effects of new officing through a 6-year investigation involving 13,000 people in 40 business units that was undertaken between 1994 and 2000. Various kinds of evaluative information concerning the effects of the qualities of work settings on employees’ work were collected together with ratings of job performance, team performance, and job satisfaction as well as information on employees’ learning.

The results suggest that workplaces can most strongly affect job satisfaction, then team performance, and then individual performance. Included among the environmental features with a strong effect on satisfaction and performance are settings supporting distraction-free work, settings for solo work, settings supporting interactions with coworkers, and settings supporting meetings and undistracted group work. To promote these, enclosure is particularly important. Communication is also seen as an important component in satisfaction and performance, and the spatial organization of team workspace can enhance face-to-face communication. The proximity of workstations and shared service supplies can foster informal interaction. Altogether, these results would advocate the use of combi-offices for various purposes. Results also suggest that a carefully designed workplace supporting employees' and teams' activities is an investment that pays off in business terms as well as in positive changes in corporate culture.

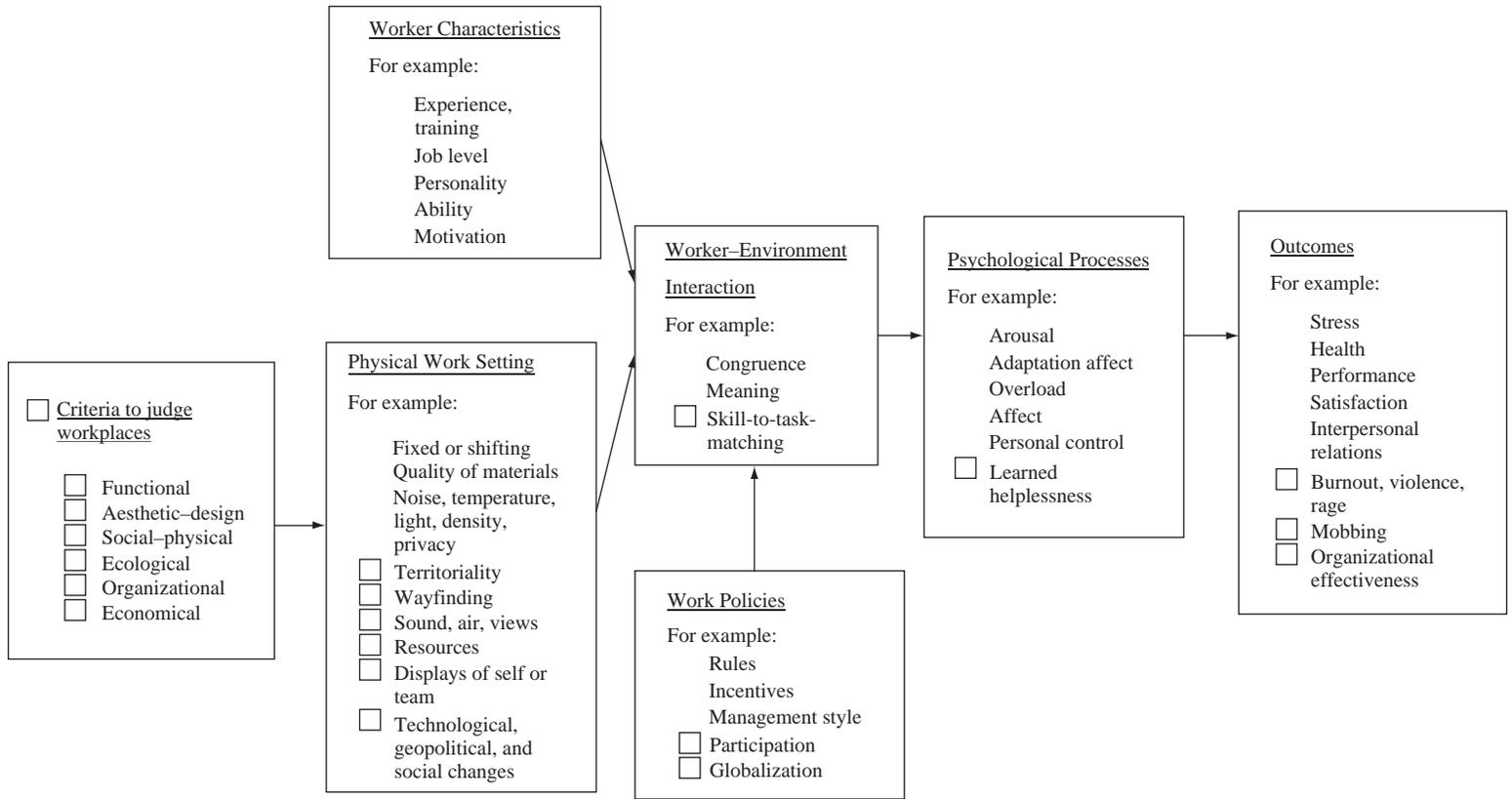
## 5. GENERAL MODELS DESCRIBING ENVIRONMENT-BEHAVIOR RELATIONS IN THE WORKPLACE

Most of the studies described so far lack a general framework or a model that would shed light on the interaction between the physical working environment and the employees' characteristics, behavior, and experiences. In 1986, Eric and Mary Sundstrom developed a framework of the physical environment of offices and factories. In a 1987 chapter, Eric Sundstrom also reported empirical evidence supporting this framework. The analytic framework developed by the Sundstroms treats people and their work environments as interdependent elements of a system and sees environmental relationships as being time dependent. It differentiates among three levels of analysis: individuals, interpersonal relationships, and organizations. At each level of analysis, dynamic processes and different outcomes are involved. The physical work environment may influence various psychological processes such as adaptation, arousal, overload, stress, fatigue, and attitudes. Outcomes for the individual workers can be satisfaction and performance. Changes involve coping as an active attempt to alter any unsatisfactory environment or its effects and other forms of adaptation. Outcomes for interpersonal relationships include adequacy of communication, group formation, and cohesion. An outcome for organizations can be effectiveness. First, the individual worker

must deal with ambient conditions such as illumination, color, temperature, air, noise, and music. In addition, the individual's environment includes aspects of the building, its furnishings, its equipment, the individual workstations, and the supporting areas. Second, design aspects such as room layout, building layout, enclosure, and gathering places may influence interpersonal relationships and communication. The regulation of immediacy, choices in communication, and privacy is a key process in these interactions. The office's physical environment, serving a symbolic, self-presentation, or status expression function, can affect job satisfaction. Third, in an organization, people are collected to act in concert toward common goals. In so doing, each person has a specific role in the hierarchy of the company. The facets of physical environment of the organization related to separation and differentiation of work units play a role in this domain. Key processes include congruence of organizational structure and physical environment.

In 2002, Gifford described a model that also tried to cover environmental psychological processes in the workplace in their entirety (Fig. 1). Based on this model, six categories relevant for describing environment-behavior relations at work can be identified. A first category comprises employees' characteristics, including personality, intelligence, cognitive styles, self-concept, emotions, motivation, satisfaction, creativity, skills, experience, training, and job level.

A second category groups the characteristics of the physical environment affecting workers' behavior. The physical environment can be broken down with regard to fixed or shifting, quality of materials, noise, temperature, light, scent, the haptic environment, density of people per space, privacy, territoriality, wayfinding, status, and personalization in displays of self or team. Five fundamental aspects can be filtered out from this: sound (e.g., noise, music), temperature (e.g., heat, cold), air (e.g., pollution, freshness, scent), light and color (e.g., sunlight, incandescent, fluorescent, windows, views), and space (e.g., amount of it, arrangement of work stations). Additions include equipment (e.g., furniture, tools, technology) and displays of self by the users (e.g., improvements in functionality, aesthetic appearance). (According to Walden in 2004, work environments can be assessed using six criteria: functional, aesthetic, social, ecological, organizational, and financial. These criteria are developed by applying the basic central themes of architectural trends, such as "form follows function," to environmental psychology. This means that functional aspects regarding, e.g., layout, wayfinding, and quality of materials, can save time and energy.



**FIGURE 1** A model for workplace environmental psychology. Items designated with empty boxes (□) were added by the author after completion of the literature review. Adapted from Gifford (2002) with the permission of Robert Gifford.

Aesthetic design results in feelings of beauty or newness. Social–physical aspects can result in conflicts that arise from simultaneous use of one setting by multiple parties or in opportunities through communication. Ecological aspects mean that the consequences of a building's existence must be taken into account—from breaking ground, to recycling, to health concerns. Organizational aspects comprise the space–time breakdown of resources, provision of information, materials and storage facilities for logistics, and sharing methods for task cycling. Financial aspects cater to the possibility of the entire organization's purchases and sales as well as its production run.)

A third category refers to work policies.

A fourth aspect is worker–environment interaction in terms of the “fit” between individuals' characteristics and the environment (e.g., skill-to-task matching, the ergonomic fit between humans and machines) as well as the meaning attributed to (and the image of) the working environment. It is important to check whether or not the physical environment and the typical behavioral patterns shown by the user groups (e.g., employees, visitors) match up. The extent to which a fit between a task (especially its level of difficulty) and the abilities of the person involved will be facilitated by the environment should also be considered. Typical examples illustrating the importance of an ergonomic fit between individuals' abilities and perceptions and the use of machines or controllers can be found in control installations for atomic power plants, for space travels, and in the cockpits of airplanes.

A fifth category involves specific individual psychological processes such as arousal, adaptation, overload, personal control (or, oppositely, learned helplessness), and affect. According to the Yerkes–Dodson Law of 1908, performance is better during medium arousal. Higher arousal blocks concentration, and at low arousal a person sleeps. Adaptation occurs when arousal subsides over time due to a process of adjustment. This can also result in a marginal degree of arousal, so that performance is then reduced. Overload arises if too many stimuli emerge over a long period of time, especially when an organism is weakened. (According to Walden, the desire for personal control may result in an appropriation of the environment, that is, in self-regulation of ambient environments and improvements in functionality and aesthetic appearance. Control and user involvement relate to noise, thermal comfort, crowding, space planning, furniture layout, environmental change, and the like. Destructive effects

include vandalism and neglect. The latter can be consequences of learned helplessness.)

A sixth category is outcome, defined as the significant changes made to people and to environments through the processes just described. Included among these are stress, health, performance, satisfaction, and interpersonal relations. Morale changes (positive or negative) in the body or the mind are linked to stress and health. (Burnout, violence, and rage can also result from stress. Mobbing is facilitated through loss of privacy.) Performance affects productivity itself but also affects indicators such as resignation rate, time spent on work, and attendance. Feelings relate to satisfaction, evaluation, attitude, emotion, and perception. Interpersonal interaction, work climate, territoriality, privacy, and status are related to social behaviors regarding the way in which space is occupied (i.e., appropriation). The layouts of space, telephones, videoconferencing facilities, and/or e-mail involve components of working conditions and, as such, serve for the transfer of work-related communication. Status demarcation underlines social roles within the hierarchy of a company through objects in the environment. Privacy is optimal at the point between isolation and too many people per room (i.e., crowding). Territoriality marks the possession of space by a single person cut off from others.

Other authors, such as Becker in 1991, have mentioned the improvement of control, communication, environmental change processes, performance, and international influences as the central psychological challenges facing workplaces. Control, communication and performance were already present in the previously described model by Gifford. Environmental change processes, such as office relocation and residential mobility, require adjustment and adaptation. International influences demand increased understanding of the significance of individual organizations and adaptation to labor in other countries.

In 2002, McCoy referred to the multifaceted individual aspects of work environments and emphasized recent empirical findings on the relationship between the physical environment and the health, safety, and behavior of occupants. In addition, she considered aspects related to office development during the past two decades or so. According to her review, more than 50% of working people in the United States work in offices. McCoy mentioned the importance of architectural details such as decorative styles, personalization, treatment of boundaries, signs, colors, and artwork that encourage a sense of identity and purpose among the individuals in an office. Personalization

includes photographs of family members, hobbies, and/or vacations. These displays of self or team can express the personal commitment to the job and can balance between work and personal activities. The importance of a view for satisfaction and psychological restoration is also emphasized. Access to the necessary resources, the building's furniture and equipment, facilities, and people are part of a system that enables the handling of information for innovations and ordinary office work. Supporting the accessibility of resources can save time for the users and can save money for the organization.

## 6. FUTURE DEVELOPMENTS

The future developments of workplaces include, on the one hand, the creation of superlative office spaces in the highest buildings in the world such as the Petronas Towers in Kuala Lumpur (Malaysia), a high-rise in earthquake-hit Taipei (Taiwan), and the planning of the Bionic Tower in Shanghai (China). On the other hand, there is an increasing concern for organizational ecology with an emphasis on sustainability. Offices in large complexes are often leased, and conference offices, especially for worldwide teleconferences among members of the same company, are being built at an increasing rate. Linked to this will be hotels with office equipment. Residential offices will have access to neighborhood work centers with telecommuting offices or satellite offices in which large numbers of people can rent well-equipped workstations in their local area or close to other places of activities. Telecommuting includes the virtual office, the home office, and satellite offices. Virtual offices can happen anywhere, including at the client's work site. In these offices, the use of a portable computer and a mobile phone is all that is needed. Trade and factories offer people the chance to work at home. This should help to alleviate the disadvantages of poor public transportation and commuter traffic.

The studies described in this article used multiple methods of investigation to fully understand the complex relationships involved in human behavior and the changing environments of workplaces. In the future, models describing environment-behavior relations at work should be expanded to include the variables that reflect the fast-paced changing conditions of workplaces. More research is needed to understand the value of those physical places and their relationship to their occupants. Even with increasing preconditions

for complete freedom of a connection between work and location, there will still be physical workplaces.

## See Also the Following Articles

Environmental Design and Planning, Public Participation in ■ Environmental Stress ■ Job Stress ■ Person-Environment Fit ■ Privacy ■ Territoriality

## Further Reading

- Becker, F. (1991). Workplace planning, design, and management. In E. H. Zube, & G. T. Moore (Eds.), *Advances in environment, behavior, and design* (Vol. 3, pp. 115-151). New York: Plenum.
- Becker, F., & Steele, F. (1995). *Workplace by design: Mapping the high-performance workscape*. San Francisco: Jossey-Bass.
- Bell, P. A., Greene, T. C., Fisher, J. D., & Baum, A. (2001). *Environmental psychology* (5th ed.). Fort Worth, TX: Harcourt Brace.
- Bonnes, M., & Bonaiuto, M. (2002). Environmental psychology: From spatial-physical environment to "sustainable development." In R. B. Bechtel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 28-54). New York: John Wiley.
- Brill, M., Margulis, S., & Konar, E., for Buffalo Organization for Social and Technological Innovation. (1984). *Using office design to increase productivity* (Vols. 1-2). Buffalo, NY: Workplace Design and Productivity.
- Brill, M., Weidemann, S., Allard, L., Olson, J., & Keable, E. B., for Buffalo Organization for Social and Technological Innovation. (2001). *Disproving widespread myths about workplace design*. Jasper, IN: Kimball International.
- Clements-Croome, D. (Ed.). (2000). *Creating the productive workplace*. London: E & FN SPON.
- Gifford, R. (2002). *Environmental psychology: Principles and practice* (3rd ed.). Colville, WA: Optimal Books.
- McCoy, J. M. (2002). Work environments: The changing workplace. In R. B. Bechtel, & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 443-460). New York: John Wiley.
- McCoy, J. M., & Evans, G. (2004). The physical environment. In J. Barling, E. K. Kelloway, & M. Frone (Eds.), *Handbook of work stress*. Thousand Oaks, CA: Sage.
- Preiser, W. F. E., Rabinowitz, H. Z., & White, E. T. (1988). *Post-occupancy evaluation*. New York: Van Nostrand Reinhold.
- Preiser, W. F. E., & Schramm, U. (1997). Building performance evaluation. In D. Watson, M. J. Crosbie, & J. H. Callendar (Eds.), *Time saver standards* (7th ed.). New York: McGraw-Hill.
- Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.

- Sundstrom, E. (1987). Work environments: Offices and factories. In I. Altman, & D. Stokols (Eds.), *Handbook of environmental psychology* (Vol. 1, pp. 733–782). New York: John Wiley.
- Sundstrom, E., & Sundstrom, M. G. (1986). *Work places: The psychology of the physical environment in offices and factories*. New York: Cambridge University Press.
- Walden, R. (2004). Assessing the performance of “offices of the future.” In W. F. E. Preiser, & J. C. Vischer (Eds.), *Assessing building performance: Methods and case studies*. Oxford, UK: Butterworth–Heinemann.
- Wineman, J. D. (1986). *Behavioral issues in office design*. New York: Van Nostrand Reinhold.



# Work Motivation

Edwin A. Locke

University of Maryland (Emeritus), College Park, Maryland, USA

1. Needs
  2. Values
  3. Goals
  4. Self-Efficacy
  5. Emotions
  6. Job Satisfaction
- Further Reading

form of psychological measurement, with one's values serving as the measuring instrument. For example, cognition might state, Here are three possible courses of action and their possible outcomes. Motivation would state, This one is most important or beneficial to me. Both motivation and cognition are explicitly or implicitly involved in every action.

## GLOSSARY

*emotion* The form in which one experiences subconscious value appraisals.

*goal* The object or aim of an action.

*job satisfaction* Liking for one's job.

*needs* The objective requirements of survival and well being.

*self-efficacy* Task-specific self-confidence.

*value* That which one considers good or beneficial.

The concept of motivation refers to factors within the individual that impel him or her to action. Motivation is not the same as arousal. Being motivated means being motivated to do something specific. Motivation affects people's choice of action (direction), the amount of effort expended on a given action (intensity), and the persistence of action (duration). In reality, motivation is never separate from cognition. Cognition answers the question, What is? It involves perception and conceptual identification. Motivation pertains to the question, How important is it? It involves evaluation. Evaluation is a

## I. NEEDS

There are five fundamental concepts in the realm of motivation: needs, values, goals, self-efficacy, and emotions. Needs are the objective requirements of an organism's survival and well-being. The existence of needs stems from the fact that life is a conditional process. For a living organism to survive, it must take the actions its nature requires. If an organism's needs are not satisfied, it experiences discomfort, suffering, unhappiness, pain, illness, and/or death, depending on the need in question and the severity and duration of deprivation. Needs are discovered largely through the effects of deprivation. People without food feel hungry and weak; people without self-esteem feel anxious and depressed.

Human needs are most logically divided into two broad categories: physical and psychological. Physical needs are the requirements of a healthy body. They fall within the province of medicine. These needs include air, water, food (including all the nutritional requirements that this implies), rest, sleep, and temperature maintenance. These needs differ in terms of the

symptoms of deprivation and also in the time lag between total deprivation and illness or death. For example, 2 minutes without air (unless the brain is cooled to a low temperature) leads to irreversible brain damage. A few minutes in very cold water can also be fatal. However, people can survive several days without water and sometimes weeks without food (if they have water). Specific dietary deficiencies (e.g., the B vitamins) may take years to manifest themselves. People may live, and even thrive, for years and even decades with partial deprivation of some physical needs.

Psychological needs are the requirements of a healthy consciousness. The core psychological need is that of self-esteem, the conviction that one is fundamentally worthy and efficacious. The opposite of self-esteem is self-contempt or self-doubt; low self-esteem is often associated with depression. Another psychological need is that of pleasure. There are many possible sources of pleasure in life, including food, drink, work, hobbies, sports, leisure activities, art (e.g., music, literature, and film), friendship, and romantic love (including sex). Art, friendship, and love are also needs in themselves. Art provides humans with spiritual fuel by presenting fundamental value judgments about humans and the world in directly perceivable form.

Contrary to Maslow's well-known need theory (with the following hierarchical order, from low to high: physiological, safety, love, self-esteem, and self-actualization), there is no evidence that needs exist in any built-in hierarchy. Deprivation of some needs causes death faster than deprivation of other needs (e.g., air vs food), but people have the power to prioritize needs (e.g., dive into water and risk drowning in order to save a child). Nor is it the case that all physical needs take precedence over all psychological needs. People may commit suicide in very short order if they conclude that they are totally worthless. Furthermore, people may undergo drastic deprivations or take incredible risks for the sake of preserving a certain view of themselves. In contrast, the deprivation of other psychological needs (e.g., art) may have more gradual effects (e.g., loss of joy in life). It also must be noted that needs are never permanently satisfied; needs operate cyclically. Life is not a one-time, static event but "a process of self-sustaining and self-generated action" (Rand, 1964, p. 16). Some needs may be habitually satisfied (e.g., food), thus leaving the person free from discomfort in relation to that need, but this does not mean that everyone who is satiated with respect to a specific need will automatically proceed to give priority to the same not-yet-fulfilled need.

The concept of need explains why living organisms act at all and must act continually if they are to sustain their lives. The most fundamental reason why people work to earn a living is to fulfill their needs and thus sustain their existence. However, the concept of need alone is not sufficient to explain human action. Need deprivation does not automatically guide people toward the correct action. First, they may not know what to do; pain does not automatically endow one with knowledge of how to relieve it. Feeling hungry does not tell people where to get food or which food to get. Second, people may act against their own needs because they want things that are not good for them, such as marijuana or high-cholesterol foods. Third, there are often options regarding to what need-fulfilling actions people can take (e.g., what career to pursue or which job to apply for). Thus, another motivational concept is required—that of value.

## 2. VALUES

A value is "that which one acts to gain and/or keep" (Rand, 1964, p. 16). It is what one wants or considers beneficial to one's welfare. Whereas needs are inborn, values are acquired. One needs food but acquires the desire for pizza and ice cream. One needs art but acquires the love of Chopin or Vermeer. Values are a link between needs and action.

Unlike needs, values do exist in a hierarchy, although not an inborn one. If people did not have value hierarchies, they would be paralyzed in the realm of action (not to mention thought). They would not be able to choose among the many action alternatives with which they are constantly faced (e.g., "Should I go to work or stay home and watch TV?") Values prioritize action and therefore prioritize needs.

People are not necessarily fully aware of their value hierarchies. Their subconscious hierarchies may differ from their conscious ones, and this discrepancy may be a source of conflict (e.g., "I want to do well in my work but somehow I feel held back").

Value hierarchies do not remain static. There are many context factors involved, including knowledge, life and career stage, family situation, health, and finances, that can change over time. The following is an important question in the realm of work: How important is one's career in comparison to other values? Highly ambitious people make career achievement a major priority and thus tend to have more career success than those who are less ambitious.

Values also account for individual differences in relation to what career and job one chooses. People differ widely in what they want from their jobs, especially the specific type of work they prefer. Preferences for different types of work are often called interests, but an interest is a form of value (what, specifically, one enjoys doing).

Values are also the basis for the effects of incentives. An incentive is the offer of a value in return for specified actions or results (e.g., a sales increase of 10%). Money is widely used as an incentive because it can be used to purchase many other values and thus is valued itself. People are more motivated at work if they believe that their efforts will lead to a valued reward. This is the basic premise of valence–instrumentality–expectancy theory.

### 3. GOALS

Values have implications for action but values are general (e.g., “I value education”), whereas action is specific. The last motivational link leading to action is the concept of goal. A goal is the specific object or aim of an action. The word intention is sometimes used as a synonym for goal, although the word intention has more of a short-term focus.

Goals and intentions are the most direct conscious determinants of action. Performance goals are significantly related to actual job performance. The highest level of performance is attained when goals are both difficult and specific, feedback is provided showing goal progress, there is commitment to the goals, and there is knowledge of what strategies are needed to attain the goals.

The motivational sequence, then, goes from needs to values to goals. Goals mediate value effects, and values mediate need effects.

### 4. SELF-EFFICACY

Although it is discussed elsewhere in this encyclopedia, the concept of self-efficacy must be mentioned here. Self-efficacy is task-specific self-confidence. (Bandura does not use the concept to refer to global efficacy, which, as noted previously, is an aspect of self-esteem.) It is similar in meaning to effort–performance expectancy in expectancy theory, except that it is more inclusive in that efficacy judgments can be based on many factors in addition to effort effects. Although self-efficacy is a belief, it has powerful motivational effects. People with high

efficacy set higher goals, are more committed to difficult goals, are more resilient in the face of failure and negative feedback, and choose better task strategies than people with low self-efficacy. There is evidence that goals and self-efficacy mediate the effects of both personality and other incentives, such as assigned goals, money, participation in goal setting, and job enrichment.

## 5. EMOTIONS

Emotions are the form in which people experience subconscious value judgments. The sequence is as follows: perception (including all stored information about the entity or situation being appraised) → subconscious value appraisal → emotional response. Emotions involve implicit psychological measurements, viz. what is the significance of this in relation to my values? People react differently to the same object when they have different beliefs about it and/or different value standards. Every emotion is the result of a specific form of appraisal. Fear is the response to the appraisal of physical threat, anxiety is the response to a self-esteem threat, anger stems from a perceived injustice, guilt results from the perceived violation of a moral value, etc.

## 6. JOB SATISFACTION

Job satisfaction stems most directly from the appraisal of one’s job as providing important job values or attaining one’s goals. (Need-value conflicts can occur when values are irrational—that is, contradictory to one’s objective needs.) The most powerful determinant of job satisfaction is mental challenge. This includes being given responsibility for decision making, the chance to use one’s abilities, the chance to achieve goals, the chance to learn new things, and feedback regarding one’s performance. However, it is important to feel challenged at work one personally likes or considers significant; thus, the work, as noted previously, must be consonant with one’s personal interests. Other commonly held job values include competent and considerate leadership and supervision; fair treatment in terms of both content (what rewards one gets, i.e., distributive justice) and process (i.e., procedural justice); the opportunity to advance; competent and compatible coworkers; safe, comfortable, and convenient working conditions; and competitive fringe benefits.

Although perceived job attributes in relation to one’s job values are critical determinants of job satisfaction,

other factors are also involved. The individual's core evaluations also play a role. Core evaluations (CEs) are tied to one's needs; they are fundamental estimates that people make of themselves. The four CE components are self-esteem, general self-efficacy, (internal) locus of control, and (low) neuroticism. Core evaluations have been found to affect job satisfaction directly and indirectly through their effects on how people perceive their jobs, the difficulty of the tasks they choose, and the complexity levels of the jobs they hold.

Core evaluations are associated with several of the Big Five personality traits, especially neuroticism (by definition), extraversion, and conscientiousness. Conscientiousness is the best single Big Five predictor of job performance, although its effects are partly mediated by goals. Passion for the work is also important.

The traits of positive and negative affectivity have also been found to be strongly associated with job satisfaction and dissatisfaction. However, it is not clear whether this relationship is causal. Asking people to rate how they feel and then to rate how they feel about their jobs or their lives may simply be asking the same question in slightly different forms.

The issue of the relationship between job satisfaction and job performance has been debated within industrial/organizational psychology for approximately half a century. The correlations between the two tend to be low, although they are higher when corrected for reliability. The correlations are higher for more complex jobs. The real question, however, is the following: What is the causal relationship between them, if any? Certainly, high performance could be a cause of high satisfaction if high performance yielded more inner satisfaction and practical rewards than low performance.

The other side of the coin is more complex. Emotions do contain felt action tendencies; the tendency is to avoid or harm negatively appraised objects and to approach, protect, or retain positively appraised objects. However, a felt action tendency does not compel action and does not identify the many alternative courses of action that one can identify through reason. Fisher and Locke have shown that there are many possible reactions to job dissatisfaction, including job avoidance, aggression (e.g., lawsuits), defiance (of company rules), constructive protest, and inaction, in addition to reducing work effort and productivity. Thus, the question is, What form of action will people choose when they do or do not like their jobs? Given that avoidance is the natural action tendency built into dissatisfaction, one would expect that turnover would be the most likely response to dislike of the job, and, in fact, it is—mediated through the intention to quit and

moderated by various factors. It remains to be determined in what circumstances people will choose high or low productivity as their preferred means of expressing their degree of satisfaction with the job. Satisfaction, for example, can lead to setting higher goals or to complacency.

### See Also the Following Articles

Emotion ■ Industrial/Organizational Psychology, Overview ■ Motivation and Culture ■ Motives and Goals ■ Values and Culture

### Further Reading

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Barrick, M., & Mount, M. (2000). Select on conscientiousness and emotional stability. In E. Locke (Ed.), *Handbook of principles of organizational behavior*. Oxford, UK: Blackwell.
- Fisher, C., & Locke, E. (1992). The new look in job satisfaction research and theory. In C. Cranny, P. Smith, & E. Stone (Eds.), *Job satisfaction*. New York: Lexington-Macmillan.
- Greenberg, J. (2000). Promote procedural justice to enhance acceptance of work outcomes. In E. Locke (Ed.), *Handbook of principles of organizational behavior*. Oxford, UK: Blackwell.
- Hackman, R., & Oldham, G. (1980). *Work re-design*. Reading, MA: Addison-Wesley.
- Howard, A., & Bray, D. (1988). *Managerial lives in transition*. New York: Guilford.
- Judge, T. (2000). Promote job satisfaction through mental challenge. In E. Locke (Ed.), *Handbook of principles of organizational behavior*. Oxford, UK: Blackwell.
- Judge, T., Locke, E., & Durham, C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. In L. Cummings, & B. Staw (Eds.), *Research in organizational behavior* (Vol. 19). Greenwich, CT: JAI.
- Judge, T., Locke, E., Durham, C., & Kluger, A. (1998). Dispositional effects on life and job satisfaction: The role of core evaluations. *Journal of Applied Psychology*, 83, 17–34.
- Judge, T., Thoreson, C., Bono, J., & Patton, G. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127, 376–407.
- Lazarus, R., & Lazarus, B. (1994). *Passion and reason*. New York: Oxford University Press.
- Locke, E. (1976). The nature and causes of job satisfaction. In M. Dunnette (Ed.), *Handbook of industrial & organizational psychology*. Chicago: Rand McNally.
- Locke, E. (2000a). Motivation, cognition and action: An analysis of studies of task goals and knowledge. *Applied Psychology: An International Review*, 49, 837–853.

- Locke, E. (2000b). *The prime movers: Traits of the great wealth creators*. New York: AMCOM.
- Locke, E. (2001). Self-set goals and self-efficacy as mediators of incentives and personality. In M. Erez, U. Kleinbeck, & H. Thierry (Eds.), *Work motivation in the context of a globalizing economy*. Mahwah, NJ: Erlbaum.
- Locke, E., & Latham, G. (1990). *Goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Locke, E., McClear, K., & Knight, D. (1996). Self-esteem and work. In C. Cooper, & I. Robertson (Eds.), *International review of industrial & organizational psychology*. Chichester, UK: Wiley.
- Peikoff, L. (2001). *Objectivism: The philosophy of Ayn Rand*. New York: Dutton.
- Rand, A. (1964). *The virtue of selfishness*. New York: Signet.
- Rand, A. (1969). *The romantic manifesto*. New York: World.



# Work Role, Values Sought in the

Donald G. Zytowski

National Career Assessment Services, Inc., Ames, Iowa, USA

1. General Values
  2. Work Values
  3. Conceptual Issues
  4. Assessment Issues
  5. Validity
  6. Case Illustration
  7. Conclusion
- Further Reading

## GLOSSARY

**industrial/organizational psychology** The branch of psychology that is concerned with worker performance and satisfaction from the perspective of the employer. It emphasizes worker selection and placement and also job descriptions in the service of employers.

**job satisfaction** The hedonic tone, or the level of satisfaction or dissatisfaction, associated with one's job or occupation. It is assessed in either of two modes: a general statement in scalar response form or responses to a number of facets of work, such as income, work conditions, and prestige. These facets are typically the conceptual source of work values.

**vocational psychology** The branch of psychology that is concerned with work and careers as they apply to the individual. It emphasizes young people's career development and planning and adults' adjustment to their careers, maximizing satisfactory vocational adjustment.

Work values are personal preferences for selected rewards of work or of a job. They are one of several

individual difference variables in the affective domain that are involved in the career decision making and planning of youth as well as in the work behaviors and job satisfaction of employed persons.

## 1. GENERAL VALUES

A value may be generally defined as that which a person believes is good, true, or beautiful. There are many classes of things that may be valued. Rokeach specifies two classes: instrumental (ways of behaving) and terminal (end states of existence). Others distinguish personal values such as respect for others, belief in God, or living frugally; moral values such as honesty; and cultural values such as the importance of family. Different cultures or individuals may value opposites, as in valuing military service versus pacifism.

## 2. WORK VALUES

The concept of work values emerged from the study of job satisfaction. It was found that generalized expressions of job satisfaction were approximately equivalent to the cumulated satisfaction with specific aspects of jobs, such as income, hours worked, opportunities for advancement, and level of independence. From inquiring about how satisfied one is with, for example, the work aspect of income, it follows to inquire as to how important (how valued or how needed) income is to the individual.

Assessment of the importance of various aspects of work and jobs forms the basis for the measurement of work values. A number of work values inventories have been published for both theoretical investigations and use in employee selection and career development.

Work values reside securely in the affective domain of behavior, somewhat less specific than interests (preferences for activities, although sometimes occupations and school subjects) but more specific than personality characteristics. Super proposed an "onion model" with personality at the core and values and then interests in successive layers, with increasing responsiveness to external, environmental factors.

### 3. CONCEPTUAL ISSUES

The concept of work values suffers from several issues that are yet to be resolved. Most prominent are definitions of terms, their number, and their underlying structure.

#### 3.1. Definitions

An unfortunate conceptual confusion underlies our comprehension of work values. Dawis and Lofquist posit needs in the individual and reinforcers in work. For them, values are the underlying dimensions of needs. Super, on the other hand, specifies values as objectives that one seeks in order to satisfy more basic needs. The definition given previously attempts to avoid the need-value problem by offering preferences in their place and the rewards of work, jobs, or occupations as the objects of those preferences.

#### 3.2. The Number of Work Values

A persistent question concerns the optimal number of work values. Dawis and Lofquist list 21, which have been found to be distributed among six higher order factors. Super first issued a measure of 15 work values similar to those of Dawis and Lofquist and then a subsequent version that added 6 more. (This latter edition has been used extensively in cross-cultural studies.) Other assessments in published inventories and on-line career exploration programs contain as many as 30 work values. The potential number may be nearly infinite, limited only by the eventual triviality of the aspects preferred (e.g., cubicle or window office and reserved parking privileges).

Another complicating factor in delineating the nature and extent of work values as aspects of jobs or occupations is that some job attributes can be bipolar

in valence. Danger in an occupation (such as firefighting) can be attractive to some but repellant to others.

#### 3.3. Underlying Structure

Some investigators have attempted to develop the structure of work values. A popular ad hoc structure assumes three sources of satisfactions: the work (solving problems, recording data, etc.), concomitants or work situation (supervision, independence, and work environment), and outcomes (pay and prestige). These are not unlike the classification of the sources of job satisfaction into categories of the work and the conditions of work. A factor analysis by Dawis and Lofquist of their 21 needs revealed six defining scales that Dawis speculates are bipolar and constitute the dimensions of work values: Achievement vs Comfort, Altruism vs Status, and Autonomy vs Safety. Other schemes classify work values as Motivation and Hygiene factors; as Social, Power, Uncertainty, and Goal orientations; or Careerist, Wage Earner, and Entrepreneur.

### 4. ASSESSMENT ISSUES

A number of measures of work values are available. They are in part distinguished from one another by their method of measurement. One method is Rokeach's pair comparison; the other is the scaled, or Likert, response format. Pair comparison yields a rank order of values without a certain zero point. This is a handicap to statistical analysis but is eminently practical in vocational and career counseling. Also, as the number of variables compared increases, the number of items in a pair comparison inventory increases exponentially, which respondents are apt to view as tedious. Assessments employing scaled responses typically tap need or importance level rather than preference, and they depend for their interpretation on norms.

In general, assessment of work values has not attained the level of sophistication typical of measures of vocational interest.

### 5. VALIDITY

The stability of work values has been assumed, perhaps from the findings for interests, but there is little empirical evidence in support of this assumption. Values apparently are little related to abilities, somewhat more

related to interests, and more related to personality traits, with occasionally as much as a .40 correlation.

Occupations are far less homogeneous with respect to work values than to interests; thus, work values do not differentiate among occupations as effectively as interests. Patterns of certain work values may be more associated with occupational levels, such as professional compared with skilled trades. Differences are reliably observed in the work values of people of diverse national or cultural origins who are employed in the same job or work assignment by multinational employers.

Given the genesis of work values in the job satisfaction literature, it would be expected that they would predict job satisfaction. Dawis and Loquist demonstrate this very convincingly. Additionally, work values are better predictors of job satisfaction than interests and better predictors of job dissatisfaction than satisfaction. Note that the criterion variable is job satisfaction, not occupational satisfaction. Moreover, work values have been linked with job tenure and some performance variables (sales success), and even in one study with the incidence of coronary heart disease.

## 6. CASE ILLUSTRATION

“Jenny” is a 30-year-old woman currently employed as a legal assistant. She grew up in a rural environment, living across a gravel road from her grandparents, but now lives and works in the largest metropolitan area in her state. She says she likes her job, largely because her employers are frequently out of the office, leaving her to work independently. She is satisfied with her pay but finds the work sometimes lacks challenge. At home, she has a substantial collection of house plants, and she could imagine working in a plant nursery, if not owning one, someday.

Her profile on a revised version of Super’s Work Values Inventory reveals highest (top quartile) ranks on Security and Lifestyle (being able to do what you want after work). Independence is below the 75th percentile level but more separated from the next lowest score than from the highest two. In the bottom quartile are Mental Challenge, Creativity, Achievement, and Variety. Between these extremes are Income, Coworkers, Supervision, Work Environment, and Prestige.

It is not good practice to attempt to reconcile a work values profile exactly with an individual’s current employment. Other job factors (e.g., the employment market) and other values (e.g., family matters) may

operate. Jenny’s current position apparently satisfies a sufficient number of her work values for her to continue in it. Jenny might use her results to refine her current position, to consider a next position, or possibly to pursue the horticulture field. Although her value on Lifestyle and Independence is not inconsistent with owning a plant nursery, she might do well to consider how she might achieve this in light of her highest rank on Security and her very low score on Achievement. One would not use the results of her Work Values Inventory to validate or caution against Jenny’s hopes but, rather, to help her envision how she might attain them and function within them should she succeed.

## 7. CONCLUSION

Despite their relatively long history in differential psychology, the concept of work values appears to be poorly conceptualized. Several measures of work values have been developed and have demonstrated some utility. They appear to account for significant variance in job satisfaction and, to a lesser extent, job performance. Work values do not apparently differentiate occupations, but they have utility in assisting vocational development and career planning of individuals.

### See Also the Following Articles

- Industrial/Organizational Psychology across Cultures
- Industrial/Organizational Psychology, Overview
- Vocational Psychology, Overview

### Further Reading

- Dawis, R. V. (1991). Vocational interests, values and preferences. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 833–872). Palo Alto, CA: Consulting Psychologist Press.
- Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis, MN: University of Minnesota Press.
- Hofstede, G. H. (2001). *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage.
- Rokeach, M. (1973). *The nature of human values*. New York: Free Press.
- Sverko, B., & Super, D. E. (Eds.). (1995). *Life roles, values, and careers*. Palo Alto, CA: Davies-Black.



# Work Safety

Dov Zohar

Technion–Israel Institute of Technology, Haifa, Israel, and Institute for Work and Health, Toronto, Canada

1. Individual-Level Perspective
  2. Group-Level Perspective
  3. Organization-Level Perspective
- Further Reading

## GLOSSARY

**accident proneness** An invalidated hypothesis suggesting that some individuals are frequently involved in accidents due to accident-prone personality.

**behavior-based safety** An intervention approach using Skinner's behavior modification principles (frequent contingent reinforcement) to overcome tendencies promoting unsafe behavior.

**developmental hypothesis** A proposition suggesting that young workers suffer heightened injury due to factors associated with lack of emotional maturity, adventure seeking, underestimating risks, and carelessness at young age.

**healthy organization** An organization that actively promotes employees' physical and emotional welfare.

**human factors engineering** An interdisciplinary field of research focusing on human and machine (or job) attributes influencing the degree of fit between the two.

**inexperience hypothesis** A proposition ascribing heightened injury of young workers to limited knowledge and skills as well as little or no previous work experience; that is, age covaries with experience.

**melioration effect** A strong tendency to overweight immediate or short-term outcomes, even when this compromises expected utility in the long run; melioration promotes unsafe behavior, whose short-term benefits often include greater speed, lesser effort, and personal comfort.

**recency effect** A strong tendency to overweight recent or frequently recurring outcomes, even when this compromises expected utility in the long run; recency and melioration exert additive effects on unsafe behavior.

**safety climate** A consensual perception of employees concerning the priority of safety in the workplace, derived from enacted policies and practices indicative of managerial commitments.

**supervision-based safety** An intervention approach using readily available supervisory incentives (recognition and feedback) to change the unfavorable cost/benefit ratio associated with safe behavior.

**transformational leadership** A value-based (vs transaction-based) leadership pattern characterized by commitment to the psychological and physical welfare of group members.

Work accidents cost the American economy an estimated \$132 billion per year in addition to human suffering and loss of life. A fatal work injury occurs every 2 hours, and a disabling injury occurs every 8 seconds. Overall world estimates suggest that approximately 100 million work injuries, 100,000 of which are fatal, occur each year. Because 85 to 90% of accidents involve a combination of unsafe acts and unsafe conditions, it is evident that human behavior is at the core of occupational safety. This suggests a significant role for psychology in terms of explanatory models and alleviation practices. Theoretical explanations and corresponding interventions can be grouped by level of analysis, stemming from the fact that behavior in organizations is influenced by individual-, group-, and

organization-level factors and that a comprehensive description requires cross-level integration. This review of significant contributions of psychology to work safety is organized accordingly.

## 1. INDIVIDUAL-LEVEL PERSPECTIVE

Individual-level explanations of unsafe behavior focus on pathogenic factors, mostly personality traits, youthfulness/inexperience, and substance abuse. Recent developments offer complementary (nonpathogenic) explanations focusing on known behavior biases that are likely to promote unsafe behavior during routine work.

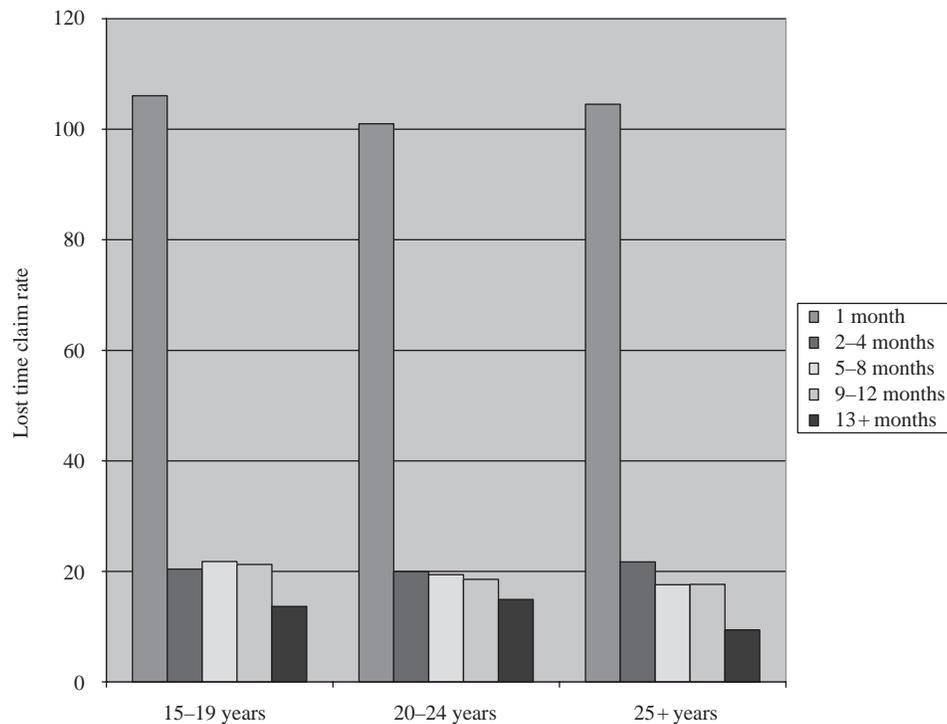
Early research on behavioral safety was concerned with the accident proneness hypothesis, stimulated by studies during the 1920s that indicated that some individuals are frequently involved in accidents. Interest in the accident-prone personality waned due to researchers' inability to establish its core characteristics and to indications that proneness was an artifact of inadequate statistical analysis. Subsequent studies focused on individual personality traits and their relationship with accident involvement. This work is largely empirical, consisting of weak arguments and post hoc explanations. Literature reviews concluded that extraversion (sensation/excitement seeking) and neuroticism were associated with increased accident likelihood. Subsequent reviews added aggression, social maladjustment, impulsivity, and external locus of control to extraversion and neuroticism as relevant personality attributes. More recently, researchers tested relationships between the five primary personality dimensions and accidents, reporting a relationship among neuroticism, low conscientiousness, low agreeableness, and openness. However, because effect sizes have been small to moderate and no coherent theoretical framework has emerged (i.e., each personality dimension requires a different post hoc explanation), the utility of personality testing as a screening device for high-risk jobs remains questionable.

The second pathogenic factor is known as "young workers." Adolescents and young adults sustain injuries at two to four times the rate of those over 25 years of age. Research focused on two rival hypotheses. The developmental hypothesis ascribes heightened injury to factors associated with lack of emotional maturity and cognitive development, greater adventure seeking, limit testing, underestimating risks, and carelessness

at young age. The inexperience hypothesis ascribes heightened injury to limited knowledge and skills as well as to little or no previous work experience (i.e., age covaries with experience). Empirical evidence favors the inexperience hypothesis, suggesting that the injury rate in all age groups is highest immediately after starting a new job (Fig. 1). Therefore, the lack of experience associated with young age and overrepresentation of young workers in short-tenure jobs involving frequent job changes are the underlying factors that drive elevated injury rates for young workers. This finding has important practical implications because inexperience, unlike developmental factors, can be addressed through better training and personal coaching. These data emphasize the importance of early safety training and supervised entry given that the first few weeks at a new job are critical.

Substance abuse (mostly alcohol) is the third pathogenic factor given the prevalence of medium to heavy drinking in blue-collar workers. Literature reviews reveal mixed or inconsistent evidence, mainly due to methodological shortcomings of available research. Most studies fail to discriminate between on-the-job drinking and off-the-job drinking because respondents were only asked to report substance use "during the past month." However, studies that included this distinction found on-the-job substance use to be a significant predictor of work injuries. This agrees with laboratory research that revealed impairment of basic cognitive and psychomotor skills, for example, impaired time estimation, divided attention, vigilance, and complex reaction time after consumption. Therefore, workplaces where alcoholic drinks are allowed/available (e.g., during lunch breaks) run a greater risk of alcohol-related accidents. These effects indicate a need for screening of alcoholic individuals, employee assistance programs, and policy implementation regarding on-site consumption. However, substance use may also be related to increased risk of injury due to a common cause such as engagement in deviant behavior, general disobedience, or impulsive behavior. Job stress may also raise the likelihood of both substance abuse and accidents.

Decision-making research provides alternative explanations of unsafe behavior (i.e., without reverting to pathogenic factors), drawing on universal behavioral biases that can also explain unsafe behavior in routine work. A cost/benefit analysis incorporating such biases suggests that appraised benefits of unsafe behavior often outweigh those of safe behavior. One relevant bias, known as the melioration effect, stipulates that decision makers tend to overweight immediate or short-term



**FIGURE 1** Graph illustrating that first-month lost time claim rates are higher at any age. Reprinted from Breslin and Smith (2003).

outcomes, even when this compromises expected utility in the long run. Consequently, the immediate cost of safe behavior, such as slower pace, extra effort, and/or personal discomfort, is often given greater weight than are low-probability and longer term risks avoidable by safe behavior. A second bias is the recency effect, which stipulates that decision makers tend to overweight recent outcomes when making decisions, even when this may compromise expected utility in the long run. This tendency results in underweighting low-probability outcomes, even when they may carry large costs (e.g., injury). Because unsafe behavior leads to immediate and frequent benefits (e.g., increased pace, greater comfort) and safe behavior results in nonevents at best (i.e., avoidance of low-probability injury), the “benefits” of unsafe behavior are likely to be overweighted. Jointly, these biases imply that choosing safe behavior in routine work is not as natural as one might assume based on the self-preservation postulate (i.e., physiological needs as the most basic needs).

### 1.1. Intervention Approaches

The best individual-level interventions are based on the complementary paradigms of human factors engineering

and behavior modification. Human factors/ergonomics interventions are based on the premise that errors and accidents arise from various types of person–job misfits where job demands exceed operators’ (mental or physical) capacity. Using job-analytic techniques, key misfits are identified, leading to job redesign and better human–machine interface where the likelihood of error or overload is reduced. Behavior modification approaches combine safety training and behavior modification. The effect size nearly doubles when training is combined with behavior modification in comparison with training alone. Safety behavior modification, also known as behavior-based safety (BBS), is an industrial/organizational application of clinical and educational behavior modification programs stemming from Skinner’s reinforcement principles. BBS interventions modify the subjective expected utility or valence of safe behavior by offering rewards whose frequency and immediacy enlist the melioration and recency effects in favor of safe behavior. BBS interventions typically employ publicly displayed charts of the frequency (percentage) of focal safety behaviors in comparison with designated targets. Data are based on daily or weekly behavior sampling observations by external observers or coworkers, and

the frequently updated discrepancy between the current level and the designated goal provides the necessary incentive or reinforcement for change (i.e., discrepancy reduction signals future rewards). Although BBS interventions have proven to be effective, maintaining change after program termination is still problematic (i.e., rapid extinction).

## 2. GROUP-LEVEL PERSPECTIVE

The main factors known to affect group safety performance are leadership and intragroup processes. Leadership is among the best predictors of safety behavior. According to the full-range leadership model, this relationship stems from the fact that effective transactional leaders organize the work better and get subordinates to perform more reliably through better monitoring and rewarding practices. Transformational leaders facilitate open communication, promote learning and development, and encourage workers to shift from passive compliance to active participation—all of which have been shown to promote safety performance. Transformational leaders are also concerned with safety issues because their interactions with group members are characterized by balanced social exchange, nurturing mutual commitment to psychological and physical welfare. This stands in contrast to transactional leaders whose safety commitments depend on external circumstances, mainly whether senior managers consider safety to be an important performance indicator.

Intragroup processes have been the subject of research since Trist and colleagues' seminal work on sociotechnical systems during the early 1960s. This work tested different work designs in coal mines, including autonomous work groups in which group members performed multiple tasks and share the responsibility for meeting production goals. Results indicated improved productivity and safety, apparently due to better coordination among group members due to better understanding of each other's work and higher motivation/alertness stemming from more decision latitude and less monotony. A recent literature review of autonomous teams highlighted inconsistent results yet suggested an overall positive effect on safety outcomes stemming from more effective group processes such as coordination of action around interdependent tasks and mutual concern for members' well-being associated with greater group cohesion.

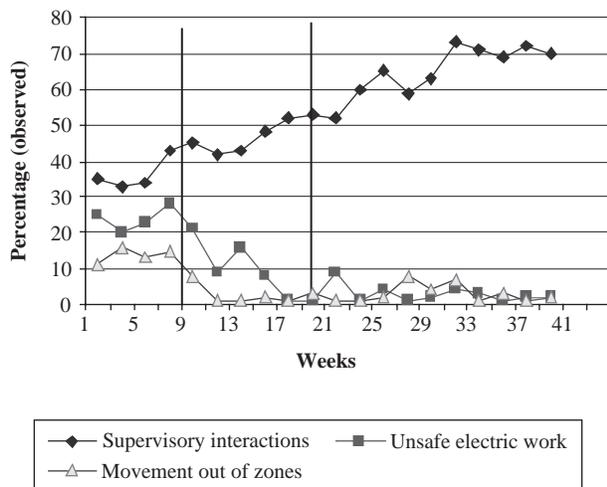
Intragroup processes have also been studied from a cognitive perspective, focusing on teams performing tasks that demand high performance reliability (e.g.,

flying aircraft, manning control rooms in chemical installations). Effective teams develop common mental models (i.e., shared understandings of their task/team system) that provide consensual expectations and anticipation of other team members' actions. Perceptual convergence allows team members to synchronize and coordinate task-related interactions in addition to monitoring and providing backup when needed. This results in fewer errors and, hence, increased safety. Although most of the work on mental models has been done with teams performing high-reliability tasks that involve periods of peak demand (e.g., cockpit crews), these results can be generalized to ordinary teams.

### 2.1. Intervention Approaches

The close connection between leadership and safety indicates that leadership training is a viable approach to improved safety performance. This can be done through reputable leadership development programs (i.e., pending availability of objective evaluation data). Safety-oriented leadership training is also effective. The most extensively tested training program employs a transactional leadership perspective designed to augment supervisory safety-oriented monitoring and rewarding practices (known as supervision-based safety). The results suggest that when supervisors actively monitor and reward safety performance, subordinates modify their behavior accordingly, resulting in effect size equaling or surpassing that of behavior-based safety interventions (Fig. 2). Furthermore, if top management is included in the intervention and supervisory roles are redefined to include safety performance as an important criterion, the effect of the intervention does not diminish after it has ended.

Intragroup processes must be assessed in terms of the distinction between action processes and interpersonal processes. Action processes involve collective monitoring, coordination, and backup, deriving from shared mental models and mutual understanding of the task/team system. It follows that process improvement must involve cross-training and continuous learning as main vehicles for shared understanding. Learning must be integrated in daily routines rather than restricted to formal training (and retraining) sessions, using routine mistakes and successes as learning opportunities (e.g., mission/work shift debriefings). Interpersonal processes, on the other hand, concern conflict management, collective confidence/efficacy, and team cohesion. The most popular intervention approach here is team building, often allied with organization development interventions.



**FIGURE 2** Synchronous change in frequency of safety-oriented supervisory interactions and workers' unsafe behaviors during a 3-month supervision-based safety intervention. Reprinted from Zohar and Luria (2003).

Team-building programs involve introducing participative goal setting, improving interpersonal relations based on openness and sharing, working through major problems in the team, and clarifying roles of individual team members. Despite the popularity of organization development, literature reviews have found only minor effects on group performance criteria, including errors and safety. This has led to the development of better balanced approaches combining elements from both processes. The most widely used approach is crew resource management (CRM), which uses detailed performance feedback to create shared mental models and social skills. Although CRM was developed for aviation crews, evaluation studies suggest its applicability to other industrial contexts.

### 3. ORGANIZATION-LEVEL PERSPECTIVE

Organization-level explanations for unsafe behavior in routine work include safety climate as the main construct. Safety climate emerges from employees' individual assessments of managerial concern for their safety and health. Such concern is expressed and operationalized by managerial practices in routine and nonroutine work conditions. When a group of employees share the same assessment and agree on the priority of safety in the workplace, individual perceptions become organizational climate. Thus, safety climate

perceptions refer to attributes of managerial actions that indicate the priority of safety in a company or the importance of acting safely while performing a job. This allows employees to assess the probable consequences of working safely or fast. Climate perceptions clarify supervisory expectations, resulting in a positive relationship between safety climate level and workers' safety behavior, as has been demonstrated repeatedly in relevant studies. Research evidence also offers consistent evidence of a positive relationship between transformational leadership and safety climate stemming from greater concern for members' welfare. Although the terms "safety climate" and "safety culture" are used interchangeably in the relevant literature, only the former is accurate considering the nature of measurement scales involved.

The other main organization-level explanation involves human resources management (HRM) practices in "healthy" organizations (i.e., organizations that actively promote employees' physical and emotional welfare). Practices in healthy organizations reflect employees' commitment (rather than managerial control) as the primary managerial orientation. As in autonomous work group designs, commitment-based orientation presupposes that workers will perform better (and more safely) if decision making is decentralized and is complemented by training, information sharing, compensation adjusted to skill and responsibility levels, and employment security. Furthermore, in the spirit of transformational leadership, HRM practices in healthy organizations are based on mutual concern for physical and emotional well-being, so that line managers are directly responsible for workers' safety, in contrast to the control orientation where injuries are attributed to workers' behavior and non-compliance with safety procedures. Workers reciprocate with proactive safety behaviors (also known as safety citizenship behavior) such as helping coworkers, communicating safety information, and initiating safety improvements. These HRM practices have also been identified as "best practices" in high-performance organizations, identifying a common denominator between organizational effectiveness and safety performance.

#### 3.1. Intervention Approaches

The climate and HRM perspectives suggest that managerial safety practices must change before safety behavior is improved. This creates a link between organization-level leadership and safety performance. Modifying HRM practices toward a commitment-based

orientation requires a transformational leadership perspective. This premise is confirmed by HRM practices in high-performance organizations. Because climate originates from enacted policies and practices (rather than from formal policy statements), improved safety climate requires greater managerial involvement throughout the organizational hierarchy.

### See Also the Following Articles

Engineering Psychology ■ Safety and Risks: Errors and Accidents in Different Occupations

### Further Reading

- Bass, B. M. (1990). *Bass & Stogdill's handbook of leadership*. New York: Free Press.
- Breslin, F. C., & Smith, P. (2003). *The relationship between job tenure and lost-time claim rates among adolescent, young adult, and adult workers*. Montreal: Canadian Association of Research on Work and Health.
- Cellier, J. M., Eyrolle, H., & Bertrand, A. (1995). Effects of age and level of work experience on occurrence of accidents. *Perceptual and Motor Skills*, 80, 931–940.
- Frone, M. R. (1998). Predictors of work injuries among employed adolescents. *Journal of Applied Psychology*, 83, 565–576.
- Frone, M. R. (1999). Work stress and alcohol use. *Alcohol Research and Health*, 23, 284–291.
- Geller, E. S. (1996). *The psychology of safety*. Randor, PA: Chilton Book.
- Hansen, C. P. (1988). Personality characteristics of the accident involved employee. *Journal of Business and Psychology*, 2, 346–365.
- Hofmann, D. A., & Morgeson, F. P. (2003). The role of leadership in safety. In J. Barling, & M. Frone (Eds.), *The psychology of workplace safety* (pp. 159–180). Washington, DC: American Psychological Association.
- Keehn, J. D. (1961). Accident tendency, avoidance learning, and perceptual defense. *Australian Journal of Psychology*, 13, 157–169.
- Macdonald, S. (1997). Workplace alcohol and drug testing: A review of the scientific evidence. *Drug and Alcohol Review*, 16, 251–259.
- National Research Council. (1998). *Protecting youth at work*. Washington, DC: National Academy Press.
- Pfeffer, J. (1998). Seven practices of successful organizations. *California Management Review*, 40, 96–124.
- Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10, 117–125.
- Stajkovic, A. D., & Luthans, F. (1997). A meta-analysis of the effects of organizational behavior modification on task performance, 1975–95. *Academy of Management Journal*, 40, 1122–1149.
- Turner, N., & Parker, S. K. (2003). The effect of teamwork on safety processes and outcomes. In J. Barling, & M. Frone (Eds.), *The psychology of workplace safety* (pp. 35–62). Washington, DC: American Psychological Association.
- U.S. Bureau of Labor Statistics. (2002). *Workplace injuries and illnesses in 2001*. Washington, DC: U.S. Department of Labor.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- Zacharatos, A., & Barling, J. (2003). High-performance work systems and occupational safety. In J. Barling, & M. Frone (Eds.), *The psychology of workplace safety* (pp. 203–222). Washington, DC: American Psychological Association.
- Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on micro-accidents in manufacturing jobs. *Journal of Applied Psychology*, 85, 587–596.
- Zohar, D. (2003). The influence of leadership and climate on occupational health and safety. In D. Hofmann, & L. Tetrick (Eds.), *Health and safety in organizations* (pp. 201–230). San Francisco: Jossey-Bass.
- Zohar, D., & Luria, G. (2003). The use of supervisory practices as leverage to improve safety behavior: A cross-level intervention model. *Journal of Safety Research*, 34, 567–577.



# Work Teams

Steve W. J. Kozlowski

Michigan State University, East Lansing, Michigan, USA

Bradford S. Bell

Cornell University, Ithaca, New York, USA

1. Introduction
  2. Team Types and Task Characteristics
  3. Team Composition
  4. Team Development
  5. Team Processes and Performance
  6. Building and Managing Effective Teams
  7. Conclusion
- Further Reading

Work teams are composed of two or more individuals; perform organizationally relevant tasks; share one or more common goals; interact socially; exhibit interdependencies in task work flows, goals, and/or outcomes; maintain and manage boundaries; and are embedded in a broader organizational context that sets boundaries, constrains the team, and influences exchanges with other units in the organization. Work team effectiveness is enabled by team processes that combine individual efforts into a collective product.

## GLOSSARY

**team composition** The collective attributes of team members, with an emphasis on how the similarities, differences, and combination of member attributes influence team processes and effectiveness.

**team development** The progression of a newly formed team from a loose collection of individuals to a well-integrated collective capable of coordinating effort to accomplish the team's task.

**team effectiveness** A multifaceted factor defined by the degree of the team's productive output, satisfaction of member needs, and the willingness of members to continue working together.

**team processes** Cognitive, affective, and behavioral mechanisms that enable team members to combine the potential of their attributes and resources to resolve constraints, coordinate effort, and achieve success.

**work team** A collective of two or more individuals that performs organizationally relevant tasks, shares one or more common goals, interacts, exhibits task interdependencies, manages boundaries, and is embedded in a broader organizational context.

## 1. INTRODUCTION

The idea of people working together in teams is certainly not new. It is quite likely that our human ancestors combined their skills and strength for the basic survival functions of hunting and defense, and human history is represented by groups working together to explore, achieve, and conquer. Yet, as the modern concept of work in large organizations developed during the 20th century, it was primarily centered on individual jobs. During the past two decades, however, strategic, economic, and technological forces have prompted organizations world-wide to substantially restructure work around interdependent teams. Teams allow more diverse skills and experience to bear on a problem, and they provide the potential for more flexible and adaptive responses to the unexpected. For example, organizations can use cross-functional teams to combine the expertise needed to solve complex problems and can

use virtual team arrangements to link these experts regardless of where they are located. Although work teams offer organizations many potential benefits, they also present challenges: Grouping people together into teams does not guarantee that they will be effective.

Most theoretical frameworks for understanding team effectiveness—either implicitly or explicitly—follow the input → process → output logic posed by McGrath in 1964. Inputs represent the characteristics of individual members (e.g., abilities, skills, personalities, and demographics), task requirements and interdependencies (i.e., work flow, goal, or outcome linkages), and organizational factors (e.g., leadership, training, and resources). From a normative perspective, these factors can be viewed as a set of resources that can contribute to team effectiveness and constraints that have to be resolved for the team to be effective. Processes represent the psychological mechanisms that enable team members to combine the potential of their characteristics and resources to resolve the constraints and achieve success. In particular, the task as a source of interdependence requirements is a key constraint that has to be resolved. Outcomes represent internal and external evaluations of team performance and the effect of the experience on team members. When team members synergistically combine the potential of their characteristics and resources to enact team processes that fit or resolve the constraints—creating synergy or process gains—their performance is effective. When team members fail to combine their potential, the resulting process loss impedes effectiveness. Indeed, the inability to capitalize on the potential of team members in cohesive decision-making groups can lead to catastrophic failures, such as those described by Janis in his 1982 book on the phenomenon he labeled groupthink.

This article discusses key factors that are responsible for effective work teams, including team types and task characteristics, member composition, team development, and processes that enable individuals to successfully combine their efforts to achieve team products.

## 2. TEAM TYPES AND TASK CHARACTERISTICS

One challenge to understanding team effectiveness is that teams have a variety of different forms, with new forms being invented to meet emerging needs (e.g., transnational and virtual teams). Team forms have

different constraints and therefore require different process mechanisms for effective performance: One size does not fit all. In an effort to better document these differences, team researchers have developed typologies to classify team forms as shown in Table I. Other researchers have argued that classification per se does not advance understanding of the constraining factors that distinguish teams, and that a more useful approach is to identify the underlying dimensions that distinguish team types. In a 2003 review, Kozlowski and Bell posited the following set of dimensions for characterizing team constraints: (i) external environment or organizational context in terms of its dynamics and degree of required coupling; (ii) team boundary permeability and spanning; (iii) member diversity and collocation/spatial distribution; (iv) internal coupling requirements; (v) work flow interdependence with its implications for goal, role, process, and performance demands; and (vi) temporal characteristics that determine the nature of performance episodes and cycles and the team life cycle.

The team task is a critical characteristic distinguishing teams both within and across types. In 1972, Steiner presented a classic typology of team tasks based on the complexity of member interdependence and resulting coordination demands required for the combination of member effort. For example, simple additive tasks sum individual work effort or products (e.g., a rope pull or a typing pool). Conjunctive tasks are limited by the performance of the weakest team member (e.g., mountaineering), whereas performance on disjunctive tasks is determined by the strongest member (e.g., problem solving). In organizations, more complex discretionary tasks allow wide latitude for how members combine their efforts, necessitating that members monitor and coordinate effort to accomplish collective outcomes. As discussed later, it is these demands for active, coordinated interdependence that place similarly complex demands on enabling team processes. Configuring team processes to fit the demands and constraints imposed by the team's task interdependence is the key to creating effective work teams.

## 3. TEAM COMPOSITION

As mentioned previously, one set of inputs that influence team effectiveness is the characteristics of individual members. Recognizing that team processes and performance will be influenced by the number and type of people who are its members, considerable research has focused on the effects of team composition. One issue that

**TABLE I**  
**Types of Work Teams**

<i>Team category</i>	<i>Key distinguishing characteristics</i>	<i>Examples</i>
Production	Core employees who cyclically produce tangible products and vary with regard to discretion from supervisor-led to semiautonomous	Automobile assembly Mining units Wood harvesting units
Service	Engage in repeated interactions with customers who have different needs, making the nature of the transactions variable	Insurance service groups Retail store teams Hospital service teams
Management	Senior managers of meaningful business units with primary responsibility for directing and coordinating lower level units under their authority	Bank management teams Health care teams Division management teams
Project	Temporary entities that execute specialized, time-constrained tasks and then disband; are generally cross-functional	New product teams R&D project groups Process improvement teams
Action/performing	Composed of interdependent experts who engage in complex, time-constrained performance events involving audiences, adversaries, or challenging environments	Military tank crews Rescue units Professional musician groups
Advisory	Temporary groups designed to solve problems and recommend solutions; work outside of, and in parallel with, production processes	Quality circles Employee involvement teams Selection committees
Crews	Characterized by the capability and necessity to form and be immediately prepared to perform together effectively; membership is dynamic and highly expert	Aircraft crews Military combat units Surgical teams
Top management	Teams composed of individuals at the executive level of an organization or corporation	Corporate executive groups Hospital top management teams
Transnational	Global or international teams; membership often spans national, cultural, and organizational boundaries	Global sales teams Multinational R&D teams
Virtual	Teams whose members are dispersed in space and interact through communication and information-sharing technologies	Distributed work teams Air traffic control teams Mission control teams Military command and control

has garnered attention concerns the best size for various types of teams. Although it is tempting to try to simplify this issue through broad-based recommendations (e.g., more is better), it appears that the “optimal” group size depends on a number of specific contingencies, such as

the level of interdependency required by the team’s task and the stability of the external environment.

The extent to which team processes and outcomes are influenced by the homogeneity or heterogeneity of member demographic characteristics is another issue

that has been the focus of considerable attention. Like team size, whether diversity is advantageous or detrimental to team functioning and performance appears to depend on a number of factors. Research suggests that diversity is particularly valuable when teams face creative and intellectual tasks and the diversity is in terms of skills and expertise rather than demographic characteristics. Research on demographic diversity has been extended to consider the team composition effects of factors such as personality and cognitive ability on team effectiveness. Both team-level conscientiousness, a personality trait, and team cognitive ability appear to be fairly potent, positive predictors of team effectiveness. Research also suggests that other personality factors, such as agreeableness, may be important and that the effects of both personality and cognitive ability depend on the task and the amount of team member interaction required for effective performance. Future research will shed more light on such issues.

Team composition has been a popular topic because of its theoretical and practical implications. Theoretically, team composition research goes to the heart of understanding how individual attributes combine to form effective interdependent groups. An even better understanding of such issues should emerge in the near future as researchers focus more attention on clearly articulating the processes by which individual actions contribute to collective team outcomes. Practically, an understanding of team composition can serve as a valuable tool for selecting and constructing effective teams. Recommendations generated from this research can help guide staffing efforts designed to produce the optimal blend of employee characteristics and can be used to effectively manage the diversity that exists in today's transnational and virtual teams.

#### 4. TEAM DEVELOPMENT

Existing teams with an ongoing life span will have outflows and inflows of new members, necessitating a process of socializing newcomers to adopt team norms, values, and goal expectations. Although there is a solid research foundation for understanding socialization to organizations, there is virtually no research pertaining to team socialization. In other instances, teams are formed anew. This is a common practice in organizations as project or virtual teams are formed to address specific problems. New teams go through a developmental process as individuals endeavor to enact norms, values, and goal expectations to guide their interaction and efforts.

Much of the conceptual work in this area takes a descriptive perspective in that it endeavors to characterize the natural process of group development. Stage models, such as the much imitated classic posed by Tuckman in 1965, describe a linear series of developmental stages to capture the team life cycle: forming, storming, norming, and performing. During the forming stage, team members first come together and begin to explore the group. They attempt to define the group task and to structure how they will accomplish it. There can be considerable ambiguity and differences of opinion as members realize that accomplishing the group task will be difficult. Lack of structure and ambiguity promote conflict, which typifies a shift to the storming stage. Disagreement and dismay ensue as members argue about group actions and form factions. With time, the group shifts to norming, as ground rules, roles, and goals are clarified and established. Conflict is replaced by cooperation, and members develop a sense of cohesion. With these normative expectations in place, the group shifts to the performing stage, with a focus on task accomplishment. Members are able to prevent or resolve intragroup conflict when it arises. They become attached to the team and derive satisfaction from progress toward collective goals.

Although some developmental process is necessary for members of new teams to develop a normative structure to guide interactions, this stage model illustrates that the process can be costly in terms of time and emotion. As a result, a variety of "team-building" interventions have been proposed in an effort to improve the normative structure of teams. Unfortunately, often such interventions are provided after teams have developed and enduring problems have been created.

The model proposed by Tuckman was based on groups working on unstructured tasks (i.e., clinical, therapy, and T-groups) and therefore heavily emphasizes interpersonal processes as individuals struggle to create some structure to guide their interactions. In a classic 1988 article, Gersick presented another descriptive model of group development based on the observation of 8 work and 8 student project teams that described the process as a two-stage punctuated equilibrium. The key factor in this model is the entrainment of group development to an external deadline that paced progress. Very early group interactions set norms that patterned activity to the midpoint of the groups' life cycle (deadline). At that point, there was a significant transformation—the punctuated equilibrium. Groups reorganized and focused on task achievement. Although some scholars regard the models as conflicting, Chang and colleagues

reported complementary processes of linear progression, indicative of Tuckman's and other stage models, and temporal entrainment, indicative of Gersick's punctuated equilibrium. Thus, both models exhibit evidence for their descriptive validity with respect to developmental processes that occur naturally. From an application perspective, however, the focus is more on prompting optimal processes rather than on what happens naturally. Thus, in contrast to the descriptive approach, prescriptive or normative approaches specify desirable conditions that should be created to foster better or faster development by targeting team processes that underlie team effectiveness. As discussed later, normative approaches provide a basis for designing interventions.

## 5. TEAM PROCESSES AND PERFORMANCE

Considerable effort has been focused on trying to understand the nature of team processes and how team processes contribute to or inhibit effective performance. There are literally thousands of research articles addressing these issues. Social psychology research on small groups has tended to focus on interpersonal processes that influence the nature and quality of member attraction and interaction, with relatively little attention to the group task. In contrast, research on work teams accepts the importance of interpersonal processes but treats the team task as central and focuses more attention on processes that reflect task-driven interactions. In their 2003 review on work teams, Kozlowski and Bell classified team processes into cognitive, affective-motivational, and behavioral mechanisms in an effort to organize this research. Descriptions of key process mechanisms that influence the performance of work teams are summarized in [Table II](#).

Cognition is a property of individuals, but team members need to act in concert. Cognitive mechanisms attempt to capture in a collective fashion team members' task-relevant perceptions, knowledge, or information. A prominent approach posits that team processes and performance are enhanced when members share a common understanding of the task environment (mental models), its goal-role-strategy requirements (coherence), and perceptions of imperatives from the broader organizational context (climate). Much of this work has been performed within the past decade, but it has been supportive and shows promise. Other approaches posit that knowledge and information do

not need to be shared *per se*; rather, it is the availability and access to the informational resources that make it useful to the team. For example, transactive memory is based on the idea that different team members will attend to different information and will know different things. As long as members know "who knows what," the information can be available to the team collectively. Similarly, team learning is a complex process that entails both individual and collective components. Work on these latter mechanisms is also recent, and also promising, although the empirical findings are less well established. It is likely that all these concepts will become more mature as researchers continue to shed light on cognitive processes in work teams.

Affective and motivational processes are also important. In addition to what team members know, the extent to which they bond to the team and its task, interact smoothly with minimal conflict, and are motivated to accomplish the team mission positively influences team effectiveness. For example, research indicates that task cohesion contributes positively to group task performance, whereas social cohesion (member attraction) appears to be less important to performance. On the other hand, social cohesion contributes to satisfaction and can help keep a good team viable over the long term. Research on conflict suggests that interpersonal conflict inhibits effectiveness, whereas task conflict has the potential to contribute to effectiveness when the group task is nonroutine. For example, differences of opinion in a group problem-solving task may surface a superior set of options. Work in this area is still in the process of better mapping relationships. Of particular importance is maturing work that gets at team motivational processes in the form of collective efficacy—a shared sense of group competence—that influences the goals that teams set, the effort they expend for achievement, and their persistence in the face of challenges. This line of work shows consistent contributions of collective efficacy to team effectiveness. Indeed, meta-analytic findings indicate that collective efficacy is an important predictor of team performance.

Cognitive and affective-motivational processes capture what team members think, how they feel, and what they are prepared to do. Behavioral processes represent what team members actually do to combine individual effort and action to accomplish team outcomes. One of the challenges of working in teams is that individuals have to allocate effort to accomplish their own responsibilities but also have to cooperatively aid others and the team as a whole. For simple tasks, the degree of cooperation is discretionary and

**TABLE II**  
**Cognitive, Behavioral, and Affective/Motivational Processes That Underline Team Effectiveness**

<i>Cognitive</i>		<i>Affective/motivational</i>		<i>Behavioral</i>	
<i>Process</i>	<i>Description</i>	<i>Process</i>	<i>Description</i>	<i>Process</i>	<i>Description</i>
Mental models	Team members' shared, organized understanding and knowledge about key elements of the team's task environment	Cohesion	Combination of shared commitment or attraction to the team task or goal and team members' attraction to or liking of the group	Coordination	Activities required to manage interdependencies with the team work flow
Team climate	Team-level shared perceptions of important contextual factors that affect team functioning	Collective mood	Team emotion or affective tone; examines how individual-level emotions combine at the team level	Cooperation	The willful contribution of personal efforts to the completion of interdependent jobs
Team coherence	Shared comprehension of the task situation and corresponding goals, strategies, and role linkages	Collective efficacy	A team's shared belief in its own collective ability to organize and execute a course of action required to produce given levels of attainment	Communication	Taskwork communication involves exchanging task-related information and developing team solutions; teamwork communication focuses on establishing patterns of interaction
Transactive memory	Team-level shared system for encoding, storing, and retrieving information	Group potency	The collective belief of a group that it can be effective		
Team learning	Relatively permanent changes in the knowledge of an interdependent set of individuals associated with experience	Conflict (divisiveness)	Can manifest as task conflict (disagreement about task content) and/or interpersonal conflict (interpersonal incompatibilities)		

largely guided by normative expectations. Complex tasks, however, have interdependence demands that place exacting requirements on whom needs to do what and when. These demands for temporal pacing, synchronicity, and entrainment distinguish coordination from discretionary cooperation. Communication helps to enable both cooperation and coordination.

## 6. BUILDING AND MANAGING EFFECTIVE TEAMS

Given the increasing use and importance of work teams, one might imagine that there is substantial interest in designing, selecting, training, and leading teams to be effective. Indeed, there is. This is a huge area for application. It is also the case, however, that much of what passes for practice in this area is not based on a solid scientific foundation. It is not possible to conduct a thorough summary of this area in the brief space available. However, those areas of practice that have merit and those that are open to question can be identified. Regarding specific interventions available in the marketplace, *caveat emptor* is appropriate advice.

Teams are often constructed and assembled with little thought about what factors need to be in place to help ensure that they have the potential to be effective. A model proposed by Hackman in 1987 takes a prescriptive perspective on team design that specifies factors necessary to support team processes that contribute to team effectiveness. The model proposes that there must be (i) a supportive organizational context that provides necessary skills via training, information, and motivating rewards; (ii) an appropriate group structure with a team composition of the correct mix of knowledge and skills, norms to guide processes, and motivation to achieve; and (iii) coaching, leadership, and support systems that provide needed resources and reduce obstacles. Teams meeting these design specifications are expected to be more effective in terms of performance, member need satisfaction, and team viability over time. This is a very useful conceptual framework for thinking about what conditions have to be created by design to promote team effectiveness.

With respect to more specific interventions, several of the areas summarized in this article have relevance. For example, team composition research has the potential to guide the combination of individual characteristics (e.g., ability, personality, knowledge, and skill) to meet team design specifications. Research on these characteristics

and their influence on team performance is still in its infancy. The area of team development also has the potential to contribute to team effectiveness by creating a normative structure to guide team processes. Indeed, a variety of interventions that fall under the rubric of team-building are purported to do just that. Team-building represents a huge area of practice, but one that is based largely on anecdotal support. Meta-analytic findings indicate that there is no empirical support for an effect of team-building on objective indicators of team effectiveness. The lack of supportive findings could be due to the infrequent evaluation of team-building interventions and the difficulty in measuring improvements in team effectiveness. It could also be due to the fact that team-building interventions are most often delivered to teams well after team development has concluded. It is quite likely that it is far more difficult to change team norms after they have been established than to influence and shape them during early team development. Thus, the timing of team-building interventions may be critical to their potential to be effective. Recent theory and research point to team leaders as playing a pivotal role in the development of effective teams by instilling a shared vision of the team's mission, creating a climate supportive of that mission, instructing coherent goal–role–strategy linkages (e.g., mental models), building task cohesion, and prompting collective efficacy during early team formation and development. This work is promising.

In addition, team training is an area that has amassed a sufficient research foundation—a science of team training—to usefully guide application. For example, a 1997 review by Salas and Cannon-Bowers identifies several training techniques and targets that have been shown to improve team effectiveness. Specific techniques that have received research support include cross-training (to allow members to understand others task requirements), coordination training (to enable better combination of effort), and team leader training (to prompt team development, shared mental models, and collective efficacy). These and other team training techniques target the attitudes (e.g., collective orientation), behaviors (e.g., compensatory behavior), and cognitions (e.g., shared knowledge) that underlie team effectiveness.

## 7. CONCLUSION

A host of strategic, economic, and technological forces have prompted a steady shift from work organized around individuals to team-based work structures. As teams pervade modern work organizations, team effectiveness

becomes an increasingly important driver of organizational success. Fortunately, there has been commensurate growth in the amount of research devoted to understanding teams and team effectiveness in work settings and, as this article highlights, much has been learned about the many different factors that impinge upon team effectiveness. Much work still needs to be done, but existing knowledge and recent advances will spur future developments in the understanding of work teams.

### See Also the Following Articles

Groups, Productivity within ■ Groupthink ■ Intergroup Relations and Culture

### Further Reading

- Bell, B. S., & Kozlowski, S. W. J. (2002). Virtual teams: Implications for leadership. *Group and Organization Management*, 27, 12–49.
- Chang, A., Bordia, P., & Duck, J. (2003). Punctuated equilibrium and linear progression: Toward a new understanding of group development. *Academy of Management Journal*, 46, 106–117.
- Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23, 239–290.
- Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, 31, 9–41.
- Hackman, J. R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). New York: Prentice Hall.
- Janis, I. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes* (2nd ed.). Boston: Houghton Mifflin.
- Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 333–375). London: Wiley.
- McGrath, J. E. (1964). *Social psychology: A brief introduction*. New York: Holt, Rinehart & Winston.
- Salas, E., & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quinones, & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 249–279). Washington, DC: American Psychological Association.
- Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.
- Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups from the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics: Theory, Research, and Practice*, 4, 44–67.
- Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, 63, 384–399.
- West, M. A. (2001). The human team: Basic motivations and innovation. In N. Anderson, D. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 2, pp. 270–288). Thousand Oaks, CA: Sage.



# Writing, Teaching of

Virginia W. Berninger

University of Washington, Seattle, Washington, USA

1. Historical Perspective
  2. Writing with and without Computers
  3. Related Processes and Developmental Stages
  4. Teaching Writing across the Grades
  5. Individual Differences in Student Writers
  6. Role of Teacher Knowledge
- Further Reading

## GLOSSARY

**alphabetic principle** Set of correspondences between a phoneme and the one- or two-letter spelling units that represent it that are applied in a different direction in spelling (from phoneme to spelling unit) than in reading (from spelling unit to phoneme).

**automaticity** Retrieval and production of a letter form or written word that is nonstrategic, effortless, direct, and rapid.

**expository (essay) writing** Writing that is informational (descriptive or compare or contrast) and/or argumentative (opinion or persuasive).

**morphemes** Word parts that, alone or attached to other word parts, convey meaning and sometimes tense, number, comparison (inflectional suffixes), or part of speech (derivational suffixes).

**morphological awareness** Conscious reflection and manipulation of morphemes in short-term and working memory.

**narrative writing** Creating a story schema with a plot or sequence of events, characters, a setting, conflict, and a resolution to the conflict or outcome.

**orthographic awareness** Consciously reflecting about and manipulating letter units in words in short-term and

working memory that correspond to phonemes and/or morphemes.

**phoneme** Smallest unit of sound that (a) has no meaning of its own but makes a difference in meaning and (b) corresponds to one or more spelling units.

**phonological awareness** Conscious reflection and manipulation of word parts in spoken words, including phonemes, syllables, and rimes in short-term and working memory.

**process writing** An instructional approach that emphasizes multiple drafts and revisions rather than a one-time product.

**scaffolding** Teacher-provided guided assistance and feedback.

**spelling unit (grapheme)** A single- or multiple-letter unit in a written word that corresponds to a phoneme.

**strategies** Reflective approaches that are nonautomatic and employ a set of procedures to reach writing goals.

**syntactic awareness** Consciously reflecting about and manipulating the grammatical structures for combining words in sentences or conversational utterances in short-term and working memory.

**think-alouds** A research or teaching tool in which the writer or writing instructor says what he or she is thinking out loud while engaged in each step of a writing process.

Effective teaching of writing initially focuses on transcription skills (handwriting and spelling) and generating short texts and subsequently focuses on specific levels of text generation (word choice, sentence syntax, and paragraph and discourse organization), specific kinds of discourse structures (narrative and expository [informational and persuasive]), specific

kinds of school writing assignments (report writing, lecture notes, study skills, and creative writing), and the cognitive processes of writing (planning, translating, and reviewing/revision). The most effective teaching approaches depend on the writer's stage of writing development. Some subprocesses of writing develop at uneven rates, creating intraindividual and interindividual differences that pose challenges for teaching writing to students with diverse writing abilities. Writing is a complex process that is learned through instruction and practice over a long period of time, beginning during the preschool years and extending into the adult years.

## 1. HISTORICAL PERSPECTIVE

Four paradigm shifts during the final two decades of the 20th century dramatically altered writing instruction at the beginning of the 21st century. First, the notion that reading should be taught during the elementary school years and writing should be taught during the middle and high school years has been replaced with the research-based understanding that, because writing emerges during early childhood, kindergarten and primary grade school children benefit from writing instruction as well as from reading instruction.

Second, the focus shifted from pedagogy alone (the teacher's instructional behavior) to learning (the student's cognitive processes during writing). "Think-alouds" used as a research tool revealed that composing draws on the following cognitive processes: idea generating; setting goals related to content, organization, and audience; making plans to reach goals; text generating; monitoring produced text and plans in progress; and revising to repair and improve text. Teachers developed methods to teach these processes by (a) thinking aloud as they modeled a writing process (e.g., planning, drafting, reviewing), (b) providing multiple-step algorithms for children to apply in their independent writing (e.g., a checklist with hints for reviewing and revising text), (c) props such as "think-plan" sheets for helping students to plan before writing and "think-revise" sheets for helping students to revise after writing a draft, and (d) scaffolding (i.e., guided assistance and feedback) as students attempted to use the various processes in their own writing.

Third, based on insights from expert professional writers, the "process writing" movement replaced the prior emphasis on the writing product that was handed in for a grade but was seldom revised based on teacher feedback to improve the writing. Professional writers

write and rewrite many drafts in which they incorporate feedback from their editors before publishing their final drafts. The process writing movement places more value on the process leading to the final product than on the initial writing product.

Fourth, the process writing movement discovered the social nature of writing and refuted the prior view that writing is an independent activity produced for an audience of one—the teacher. Students do meet individually with teachers to discuss their writing in progress, but they also receive feedback from peers and often "publish" their final versions for their peers to read and discuss.

## 2. WRITING WITH AND WITHOUT COMPUTERS

Students in the 21st century need to learn to write by hand as well as by keyboard. Writing with pencil or pen is still important because forming letters by hand engages the attention system in a way that a key press does not, and the enhanced attention to letter forms facilitates learning to read and spell words. Also, written communication is often needed in contexts where computer technology is not available. Finding letters on the keyboard is a necessary first step to learning to operate a word processing program through a keyboard and should be taught as explicitly as is handwriting. Computerized spell checks do not eliminate the need for systematic spelling instruction. Spell checks merely provide a check for typing errors; unless the writer can recognize the correct spelling among the choices offered, the spell check does not fix the spelling. Some writing goals are achieved best with pen or pencil, whereas computers have added value for writing goals such as revision and completion of long-term writing assignments. Voice recognition software for dictating written text has many technical difficulties to overcome. Students still need to be taught to compose written text using pencils, pens, and keyboards.

## 3. RELATED PROCESSES AND DEVELOPMENTAL STAGES

Writing draws on the sensory, motor, language, cognitive, memory, and executive functions of the brain, but the relative contributions of these various processes change at different developmental stages as the

brain constructs, through instructional experiences, an evolving functional writing system. Low-level skills such as handwriting and spelling are separable from high-level skills involved in composing. Low-level skills typically exert their greatest constraints early in writing development. Once low-level skills are automatic, high-level skills exert the greatest constraints on written composition. Handwriting, spelling, and composing all have their own developmental trajectories (Table I).

### 3.1. Handwriting Development

The normal developmental trajectory for manuscript letters, cursive letters, and keyboarding is from accurate to automatic letter production (Table I). These stages are first achieved for printing (manuscript) during the first two grades and then for cursive during the third and fourth grades. Cursive is more difficult because not only letter forms but also connecting loops must be mastered.

### 3.2. Spelling Development

Table I summarizes the typical developmental patterns from phonetic spelling, to phonemic spelling (alphabetic principle), to morphological spelling as well as the shift from sound-based (phonetic or alphabetic) strategies to automatic retrieval of orthographic word forms. The third and fourth stages overlap during spelling development. Also, there are tremendous individual differences in the rates at which children proceed through these stages and master them.

### 3.3. Composition Development

Table I summarizes the early stages of composition development. Paragraph structure with a true topic sentence and supporting propositions often does not develop until the upper elementary school grades. During the upper elementary and middle school grades, narrative writing becomes more elaborated with complex plots and character development, and proficiency with a variety of expository structures is still evolving. For the informational essay structure, proficiency with descriptive essays develops earlier and faster than does proficiency with compare-and-contrast essays. For the argument essay structure, proficiency with opinion essays develops earlier and faster than does proficiency with persuasive essays in which arguments and counterarguments are presented and the writer takes a position

and defends it. Composing is mastered only after many years of instruction and practice.

### 3.4. Integration with Other Language Systems

Writing development is not complete until students integrate writing with the other language systems: for listening (e.g., early spelling development, taking notes during lectures), for speaking (e.g., giving speeches with audiovisual aids), and for reading (e.g., taking notes from source materials to write reports or study for tests). All of these cross-language integration activities place large demands on the executive functions of the brain for coordinating multiple processes.

## 4. TEACHING WRITING ACROSS THE GRADES

### 4.1. Kindergarten

To develop the fine motor skills for using a pencil, children should play with pegboards and blocks, string beads, use crayons and markers for free drawing, and solve paper-and-pencil mazes. To develop the strength for using a pencil, children should play with clay. Children should be taught to write and name alphabet letters accurately. Instructional activities for accomplishing these goals include the following: (a) tracing over dotted lines to form alphabet letters, (b) studying numbered arrow cues to learn the sequential component strokes in forming alphabet letters, (c) naming letters while writing them, and (d) learning special names and stories for each letter (e.g., “Mr. K” is a kicker who kicks footballs and soccer balls). Teachers should model composing short texts and teach the following strategy: “What I Think, I Can Say, What I Can Say, I Can Write.” Children should be taught strategies for using cue cards (with pictures of key words with the sound for each alphabet letter or letter group) to generate the letters that spell the sounds in spoken words. Daily opportunities to compose freely and illustrate the writing with drawing should be provided.

### 4.2. Primary School Grades

To achieve handwriting automaticity (i.e., effortless printing of legible manuscript letters), which transfers to improved written composition, children should be asked to write letters from memory and

**TABLE I**  
**Developmental Trajectories of Component Writing Skills**

---

<i>Handwriting development</i>	
First stage	Accurate (legible) production of manuscript letters Individual letters are recognizable even if not in word context
Second stage	Automatic (e.g., fast, effortless) legible production of manuscript letters
Third stage	Accurate (legible) production of cursive letters and their connecting strokes Individual letters are recognizable even if not in word context
Fourth stage	Automatic (e.g., fast, effortless) legible production of cursive letters and their connecting strokes
<i>Keyboarding development</i>	
First stage	Finding keyboard letters by trial and error
Second stage	Finding keyboard letters accurately and quickly with visual cues
Third stage	Finding keyboard letters accurately and automatically (e.g., fast, effortlessly) without visual cues
<i>Spelling development</i>	
First stage	Words spelled as perceived in speech (e.g., phonetic cues) and in letter names (e.g., /cac/, /yd/, /mt/).
Second stage	Words spelled with the alphabetic principle (i.e., correspondences between phonemes and one- and two-letter spelling units) and both plausible nonconventional (e.g., /caik/, /whide/, /emptee/) and conventional (e.g., cake, wide, empty) spellings.
Third stage	Conventional spelling of words of increasing length; of words in which partial spelling units must be memorized (due to silent letters, alternations of the alphabetic principle for spelling the same sound, and schwas such as <u>m</u> other, <u>u</u> pon); words with morphemic units (i.e., inflectional suffixes for tense, plurality, and comparison as well as derivational suffixes that mark part of speech); words spelled with morphological rules (e.g., for adding suffixes to the root or stem)
Fourth stage	Automatic retrieval of the whole written word from long-term memory
<i>Composition development</i>	
Precomposing free association	Write a comment about a word in the previous sentence that is not the topic. [Mary laughed. Laugh don't cry.]
First stage	Topic-comment: Generate a comment about the topic in the prior sentence, as in oral discourse. [Mary laughed. Tom also laughed.]
Second stage	Wheel structure representing a topic (i.e., the central hub) with multiple comments (i.e., spokes) about the topic. [I like to play tag. I like to play hide-and-see. I like to play Monopoly.]
Third stage	Elaborated wheel structures (i.e., comments about spokes). [I like to play games. I like to play tag. I play tag with my cousins. I like to play hide-and-see. I play that with Jim and Jane next door.]
Fourth stage	Emerging discourse structure: Simple narratives with event sequences and some mention of characters and/or setting are produced. Also, simple essays (expository texts) are produced, but in the form of knowledge telling or listing that does not always adapt the message for the audience.
Fifth stage	Refining discourse structures and emerging knowledge-transforming strategies that adapt writing to the audience. [There are many kinds of games children can play. Some are played outside. Tag and hide-and-see are examples of outdoor games. Some are played inside. "Go fish" and Monopoly are examples of inside games.]

---

not merely copy them. Practicing each alphabet letter once each day—in a different random order each day—is a more efficient way in which to develop automaticity than is mindless copying of just a few letters over and over. After studying model letters with numbered arrow cues, children should cover them for a few seconds and then write them from memory. This activity should be teacher led with frequent naming of the letters because verbal names are retrieval cues that help to make the process of getting letter forms in and out of memory more automatic. Other ways in which to develop automatic letter retrieval and production include (a) writing each letter once from dictation and (b) writing letters that come before or after other letters. When cursive is introduced during third grade, the goal should initially be accurate letter form production and accurate connecting strokes.

Systematic spelling instruction for words of one and two syllables should be introduced during first grade and continued through third grade. Using high-frequency Anglo-Saxon words in the writing of primary grade school children, teachers should teach the following: (a) alphabetic principle for going from phonemes in spoken words to one- and two- letter spelling units (which contrasts with the alphabetic principle for decoding words during reading that goes from letters to phonemes), (b) alternations (the options for spelling the same phoneme), (c) predictable spelling patterns larger than alphabetic principle (e.g., word families such as /ight/), and (d) morphemic units (e.g., inflectional suffixes that have multiple sound–spelling relationships for plurals [/s/ as in cats, /s/ as in bees, and /es/ as in busses] and for tense [/ed/ as in looked, opened, and added]). The high-frequency words should be practiced alone and in sentences. Daily spelling-to-sentence dictation is more effective than weekly spelling tests with lists only. To transfer knowledge of single-word spelling to composing, children should be asked to write compositions containing a set of high-frequency words and to keep frequently misspelled words in their compositions in a personal dictionary.

Free journal writing is not sufficient for written composition. Children should be given compositional prompts (i.e., topics) to write about, compose, and share with peers daily. Teachers should give daily feedback and help children to set goals for the next day (e.g., to write one more sentence, to choose more interesting words) based on a visible record of the progress each day (e.g., a graph of how many words or sentences they have written).

### 4.3. Upper Elementary School Grades

The transition from third to fourth grade is important in writing development because the writing requirements of the curriculum increase substantially during fourth grade and beyond. Students are now expected to explain in writing their understanding of passages they read and to write reports based on source materials for the content subjects. Children whose letter production and keyboarding are not automatic and whose spelling is not grade appropriate will face special challenges in keeping up with the increasing writing requirements. Their teachers need to provide individually tailored instruction in handwriting and/or spelling (and monitor progress in development of transcription skills) along with composition instruction. Explicit instruction in using word processing programs for writing assignments should be introduced for all students.

Continued systematic spelling instruction for longer, morphologically complex words of Latinate origin (often three syllables or more and containing derivational suffixes that signal grammatical information about part of speech) and of Greek origin should be provided for all students. Morphological spelling rules should be taught for adding suffixes to roots or bases (i.e., roots already having suffixes). Strategies for memorizing word-specific spellings containing schwas or homonyms should also be taught. In general, spelling benefits from training in phonological, morphological, and orthographic awareness of words, their parts, and their interrelationships.

The instructional goals for written composition at this stage of writing development draw on reading–writing connections and include (a) learning to write well-formed paragraphs and longer discourse structures for specific genre (narrative and expository [essay]), (b) learning to express through writing the author's understanding of a text written by another author (reading comprehension), and (c) learning to take notes from source material, organize notes, and write reports based on the notes for content subjects such as science and social studies as well as for language arts assignments such as book reports. Effective instructional techniques include (a) examination of model paragraphs and model texts that contain prototypical structures that children can use in their own writing, (b) summarization activities in which children write the main idea and most important supporting details for the texts they have read, and (c) teacher modeling and teaching of explicit strategies for each of the steps in writing reports. Composing instruction should aim at each level of

language involved in producing quality text: word, sentence, and discourse. Vocabulary development and word awareness activities help children to improve in quality of word choice. Three kinds of syntactic awareness activities improve quality of sentence writing: unscrambling randomly presented words to create a grammatically acceptable sentence, crossing out unnecessary words in overly wordy sentences, and using sentence-combining activities that express in one well-formed sentence the ideas contained in two other sentences. Graphic organizers and outlining can be used to teach organizational techniques at the discourse level. Checklists for reviewing and revising at each level of language are also helpful.

#### **4.4. Middle and High School**

The most important writing goal during the middle and high school developmental stage is to acquire the cognitive ability to transition from knowledge telling to knowledge transforming so as to meet the needs of the audience more effectively. Explicit instruction in executive functions and internalization of writing plans is critical for this goal. Student writers benefit greatly from (a) teacher–student dialogues aimed at engaging students intellectually in the planning, translating, and reviewing/revising processes of composing; (b) use of computers in the writing program and continuing instruction in keyboarding and word processing; and (c) daily teacher feedback about student performance displayed visibly (e.g., with graphs) and linked to setting goals for the next writing activity. The goal(s) for writing instruction should be broader than merely teaching sentence syntax and should include genre-specific discourse writing skills and strategies for writing for multiple purposes, including (a) completing long-term writing assignments that are revised in response to feedback along the way, (b) report writing for content subjects such as social studies and science, (c) creative writing to express original ideas and experiment with genres such as poetry and plays, (d) note taking during lectures, (e) note taking during textbook reading assignments, (f) using notes from lectures and textbooks to study for tests, and (g) test taking (essay writing tests). Some middle and high school students still need help with basic transcription skills (handwriting and spelling), vocabulary use and word choice, syntactic structures, and discourse structures.

Most adolescents make the transition from learning to write to writing to learn and writing to solve problems. Writing plays an important role in the transition from

schooling to adult life. High school graduation is greatly dependent on completing writing assignments. Passing state standards in writing on high-stakes tests is becoming a requirement for high school graduation in some states. College entrance tests are increasingly requiring writing assessments. Many college-level courses also require advanced writing skills. Increasingly in the 21st century, writing is affecting the options open to adults for earning a living. Many jobs involve the use of computers that require both writing and reading to enter written language and view it on the monitors. Teaching the use of computers in writing instruction is a matter of teaching writing–reading connections.

#### **5. INDIVIDUAL DIFFERENCES IN STUDENT WRITERS**

At all grade levels, individual students show some normal variations in their relative strengths and weaknesses with specific writing skills and related processes. However, in some cases, specific writing skills fall outside the normal range and students have specific learning disabilities despite their normal intelligence or thinking ability. In such cases, the students should be referred to a multidisciplinary team that includes a psychologist, a speech and language pathologist, and an occupational or physical therapist. Common specific learning disabilities include dyslexia, which is impaired single-word reading and spelling, and dysgraphia, which is impaired handwriting and/or spelling without impaired reading. Some children have both dyslexia and dysgraphia. Other students have language learning disabilities and have trouble using oral language to learn in school and learning written language. Students with specific learning disabilities require specialized instruction.

#### **6. ROLE OF TEACHER KNOWLEDGE**

Teachers who have knowledge of research findings about the writing process and effective writing instruction are more likely to help their students to achieve developmentally appropriate writing goals than are those who do not. However, teachers who have research knowledge also need to evaluate the effectiveness of their implementation of research-supported instructional practices in their own classrooms for their own students. The application of research to

**TABLE II**  
**Instructional Design, Pedagogical, and Evaluation/Modification Principles**

- 
- A. Instructional design principles for planning instruction at specific levels
1. Incidental writing instruction—teaching skills only as individual students require special kinds of assistance—is not sufficient at any level. Lessons in handwriting, keyboarding, spelling, and composing, with and without computers, should be highly preplanned so that teachers can respond flexibly to individual students' needs without losing sight of the planned overall instructional goals in a systematic program of instruction.
  2. At all levels, plan how instructional components are packaged close in time within a lesson. Teaching to all levels of language close in time within a single lesson helps to overcome the temporal limitations of working memory and has been shown to create functional writing systems in which component writing processes are coordinated.
  3. At the middle and high school levels, plan how lessons are sequenced over the course of the week, month, semester, and year to teach the multiple steps involved in long-range writing assignments and complex skill development.
- B. Pedagogical principles for implementing instruction at all levels
1. Aim instruction at a level that is not too easy or too hard but is intellectually engaging and should lead to growth.
  2. Teach for transfer. That is, provide opportunity to practice skills in isolation and then in multiple contexts to promote generalization to novel contexts.
  3. Teach low-level skills to automaticity to free up resource-limited working memory for high-level goals.
  4. Avoid habituation (i.e., where the brain becomes unresponsive to repetitive activity) by teaching automatization of low-level skills only for brief periods of time.
  5. Also teach high-level, nonautomatic skills in an explicit way with teacher-prompted cueing of relevant dimensions, teacher modeling, teacher–student interactive dialogue, teacher-devised strategies, and teacher-led scaffolding of the application of all of these as students apply them to their writing.
  6. Teach executive functions for internally guiding and self-regulating the writing process.
- C. Evaluation/Modification principles for monitoring student progress at all levels
1. Monitor student progress daily, weekly, and monthly as well as at the beginning, middle, and end of the year with curriculum-based, criterion-referenced, and norm-referenced measures. Use results for instructional decision making (revise if necessary just as in writing).
  2. Provide visible records (e.g., through graphs) of student performance for teachers and students to review on a regular basis.
  3. Adjust teaching approach if necessary if students progress too slowly or quickly.
- 

*Source.* Adapted from Wong and Berninger (2004).

practice is not always simple and straightforward, and specific implementations need to be evaluated. Like writers who must plan, translate, and review and revise their writing, teachers must plan, translate, and review and revise their applications of research to practice. Table II summarizes research-supported general principles that are helpful in this regard.

### See Also the Following Articles

Learning ■ Literacy, Improvement of ■ Reading, Teaching of

### Further Reading

Alley, G., & Deschler, D. (1979). *Teaching the learning-disabled adolescent: Strategies and methods*. Denver, CO: Love Publishing.

Berninger, V., & Abbott, S. (2003). *PAL research-supported reading and writing lessons*. San Antonio, TX: Psychological Corporation.

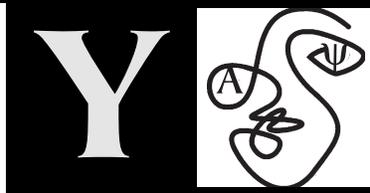
Berninger, V., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of research on learning disabilities* (pp. 345–363). New York: Guilford.

Berninger, V., & Richards, T. (2002). *Brain literacy for educators and psychologists*. San Diego: Academic Press/Elsevier.

Clay, M. (1982). Research update: Learning and teaching writing—A developmental perspective. *Language Arts*, 59, 65–70.

Englert, C. S., Raphael, T., Anderson, L., Anthony, H., & Stevens, D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal*, 28, 337–372.

- Graves, D. (1983). *Writing: Teachers and children at work*. Exeter, NH: Heinemann.
- Harris, K., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation*. Cambridge, MA: Brookline.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum.
- Hillocks, G. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. *American Journal of Education*, 93, 133–170.
- Klein, P. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11, 203–270.
- MacArthur, C., & Graham, S. (1987). Learning disabled students' composing with three methods: Handwriting, dictation, and word processing. *Journal of Special Education*, 21, 2–42.
- Scardamalia, M., & Bereiter, C. (1989). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics*, Vol. 2: *Reading, writing, and language learning* (pp. 142–175). Cambridge, UK: Cambridge University Press.
- Treiman, R. (1993). *Beginning to spell*. Cambridge, UK: Cambridge University Press.
- Venezky, R. (1999). *The American way of spelling*. New York: Guilford.
- Wong, B., & Berninger, V. (2004). Cognitive processes of teachers in implementing composition research in elementary, middle, and high school classrooms. In B. Shulman, K. Apel, B. Ehren, E. Silliman, & A. Stone (Eds.), *Handbook of language and literacy development and disorders*. New York: Guilford.
- Wong, B. Y. L., Butler, D. L., Ficzere, S. A., & Kuperis, S. (1996). Teaching adolescents with learning disabilities and low achievers to plan, write, and revise opinion essays. *Journal of Learning Disabilities*, 29, 197–212.
- Wong, B. Y. L., Butler, D. L., Ficzere, S. A., & Kuperis, S. (1997). Teaching adolescents with learning disabilities and low achievers to plan, write, and revise compare-and-contrast essays. *Learning Disabilities Research & Practice*, 12, 2–15.



# Youth, Employment of

Fred W. Vondracek

*The Pennsylvania State University, University Park, Pennsylvania, USA*

1. Youth Employment Defined
  2. The Context of Youth Employment
  3. The Developmental Significance of Youth Employment
  4. Strategies for Facilitating Youth Employment
- Further Reading

family, community, economic, and cultural contexts within which the youth's work is occurring. Thus, youth employment may result in both positive and negative consequences, although when properly structured it can be a valuable experience on the way to adulthood and independence.

## GLOSSARY

**apprenticeship model** A highly structured means for introducing youth to the world of work, for training them, and for integrating them into the labor force; it has been most successfully employed in Germany.

**historical context** Conditions of youth employment in different historical epochs.

**occupational identity** A sense of belonging and commitment to an occupational group, typically acquired after a period of exploring various career options.

**part-time employment** Paid work of less than 40 hours per week, typically performed by individuals who are also full-time students.

**youth employment** Paid work performed by individuals 16–24 years old.

Youth employment is a fact of life for the vast majority of young people in the United States. The same does not necessarily hold for youth from other countries or for youth from different historical periods. Moreover, the developmental significance of youth employment varies for each individual depending on the individual's characteristics, intensity of the work experience, and

## 1. YOUTH EMPLOYMENT DEFINED

Although there is no universal agreement on the definitions of youth or employment, for present purposes, youth employment is defined as paid work by individuals 16–24 years old. Both part-time employment (usually occurring among individuals who are also full-time students) and full-time employment are included in this definition.

## 2. THE CONTEXT OF YOUTH EMPLOYMENT

It is impossible to discuss youth employment without taking into account the various levels of context within which it takes place. Youth work takes place in relation to the families and communities to which young people belong. It occurs in relation to the cultural norms and expectations that prevail among these families and communities, and it is defined largely by the economic circumstances that occur at any given historical period. Nevertheless, because individual

circumstances vary greatly from one person to another, there can be considerable variation in how youth employment is defined and experienced, even when individuals share historical time and macrocontextual circumstances.

### **2.1. Historical Context of Youth Employment**

In preindustrial societies, children and adolescents were integral to the economic survival of families. As early as they were able to perform any useful work, they were pressed into service, working side-by-side with adults, helping to produce food and to provide shelter and safety. As industrialization progressed, parents increasingly found work away from home, largely removing them as effective socialization-to-work agents. Adolescents were increasingly faced with the necessity of discovering their own path into the world of work and acquiring their own occupational identity and with it their independence. Moreover, as industrialized nations became wealthier and their commerce became more complex, they required a more highly educated and trained workforce. As a result, education rather than labor began to occupy most children's and adolescents' time, and much of the education and training was explicitly designed to prepare youth for an ever-expanding variety of occupations.

### **2.2. Cultural and Socioeconomic Context of Youth Employment**

As noted previously, youth employment varies not only by historical time but also by cultural and socioeconomic context. For example, in some cultures employment opportunities for males and females differ greatly. Accordingly, boys and girls are differentially socialized into early work experiences and they emerge with different expectations and aspirations about their future occupational careers. Economic necessity may also impact the nature and extent of youth employment. Thus, youth from poor socioeconomic backgrounds are more likely to quit school and go to work to contribute to their own and their family's economic survival. In contrast, youth employment is typically optional for youth from more privileged backgrounds, where it may be considered a step toward "growing up" or simply a "good experience."

## **3. THE DEVELOPMENTAL SIGNIFICANCE OF YOUTH EMPLOYMENT**

Although a fair amount has been written about the detrimental impact of child labor on children's development, the impact of work on development in adolescence and early adulthood has not received adequate attention by researchers. This is surprising in view of the fact that most adolescents in the United States have a good deal of work experience by the time they graduate from high school. Data from the Monitoring the Future study at the University of Michigan indicate that approximately three-fourths of boys and girls work part-time while they are in high school, and of these, more than one-third work more than 20 hours per week. However, researchers appear to be deeply divided on the issue of whether youth employment has generally positive or negative consequences for development.

### **3.1. Negative Effects of Youth Employment**

When students are employed part-time, some energy and time that might be devoted to studying is diverted, creating the potential for compromised school performance and disrupted social relationships. Moreover, young part-time workers are more likely to be exposed to (and thus learn) negative adult attitudes toward working as well as other adult behaviors, such as smoking and drinking, which are viewed as problem behaviors when exhibited by adolescents. The causal direction of these negative attitudes and behaviors is unlikely to be the same in all cases, however. Some researchers have found that poor school performance and adult-like behaviors precipitate early entrance into the labor force, whereas others have found that these behaviors follow part-time work, especially if it involves many hours per week. Contributing to this negative scenario is the fact that most youth employment opportunities occur in low-level work settings, such as fast-food establishments, or in low skill positions, such as janitor or clerk. Adolescents from low socioeconomic status families may get "stuck" in such low-level jobs and foreclose on any real opportunity to be occupationally upwardly mobile. Others may view working in these circumstances as completely unrelated to their aspirations and ultimate career destination, resulting in disinterest or cynical attitudes toward work in general.

### 3.2. Positive Effects of Youth Employment

Although the potential negative consequences of youth employment have received what some consider an inordinate amount of attention, most youth, as well as their families, feel overwhelmingly positive about part-time work. They tend to adhere to the idea that part-time work pursued in moderation represents a beneficial experience. Jeylan Mortimer and colleagues at the University of Minnesota have contributed significant findings regarding youth employment through their Youth Development Study. They found that stable work, limited to less than 20 hours per week and including good relationships with supervisors, tended to buffer the effects of family problems and to promote positive vocational development. Other researchers have confirmed that young people who work for pay tend to learn valuable lessons in self-discipline and that they often develop important insights into aspects of work that they value highly and those that they do not value. Moreover, youth employment is often instrumental in helping young workers to know what they do not want to do—arguably at least as important as knowing what they do want to do. Equally important, youth employment facilitates the development of a vocational or occupational identity, which is a key component of one's overall identity. Thus, in important ways, youth employment assists adolescents and young adults in answering the question, "Who am I?"

### 4. STRATEGIES FOR FACILITATING YOUTH EMPLOYMENT

Youth employment, when properly structured and integrated into the lives of adolescents and young adults, can be a valuable experience in the transition to adulthood and independence. Paid part-time work may be particularly valuable when it involves moderate hours and is viewed as complementary to the completion of schooling or training. In some ways, however, paid full-time work may be even more important for those who leave school before they complete it and those who go to work immediately after high school completion. For them, the alternatives to full-time work are underemployment or unemployment, neither of which are likely to promote a smooth transition into adulthood. As a result, there has been a

dual approach to facilitating youth employment, with one branch placing a heavy emphasis on making valuable part-time learning experiences available for youth who are still in school and the other focusing on a smooth transition of youthful workers into the labor force, thereby avoiding the risks associated with youth unemployment.

Developed countries have often focused on integrating work experience and schooling, putting in place policies and programs that have been derived from the highly successful German apprenticeship model. Partnerships between public and private sectors and between schools and businesses have often been the foundation of such efforts. In less developed countries, the primary focus has been on providing employment for young people that is safe and offers them at least a reasonable chance of supporting their young families. It should be clear, therefore, that just as youth employment is quite variable in nature and highly context specific, so are methods to facilitate youth employment.

### See Also the Following Articles

Part-Time Work ■ School-to-Work Transition

### Further Reading

- Barling, J., & Kelloway, E. K. (Eds.). (1999). *Young workers: Varieties of experience*. Washington, DC: American Psychological Association.
- Hamilton, S. F. (1990). *Apprenticeship for adulthood: Preparing youth for the future*. New York: Free Press.
- Heinz, W. (Ed.). (2000). *From education to work: Cross national perspectives*. Cambridge, UK: Cambridge University Press.
- Mortimer, J. T., & Finch, M. D. (Eds.). (1996). *Adolescents, work, and family: An intergenerational developmental analysis*. Thousand Oaks, CA: Sage.
- Petersen, A. C., & Mortimer, J. T. (Eds.). (1994). *Youth unemployment and society*. Cambridge, UK: Cambridge University Press.
- Schulenberg, J., Maggs, J. L., & Hurrelmann, K. (Eds.). (1997). *Health risks and developmental transitions during adolescence*. Cambridge, UK: Cambridge University Press.
- Stern, D., & Eichorn, D. (Eds.). (1989). *Adolescence and work: Influences of social structure, labor markets, and culture*. Hillsdale, NJ: Erlbaum.
- Waddell, S. (2001). *What works in engaging business in youth employment and livelihood strategies*. Baltimore, MD: International Youth Foundation.